A randomization test for mean vector in high dimension

Rui Wang[a], Xingzhong Xu[a,b,*]

[a] *School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China*

[b] *Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China*

Abstract

The strength of randomization tests is that the size of the test is exact under certain symmetry assumption for distribution. In this paper, we study a randomization test for mean vector in high dimensional setting. In classical statistics, a major down-side to randomization tests is that they are computational intensive. Surprisingly, it is not the case in high dimensional setting. We give an implementation of the randomization procedure, the time complexity of which does not rely on data dimension. The theoretical property of randomization test is another important issue. So far, the asymptotic behaviors of randomization tests have only been studied in fixed dimension case. We investigated the asymptotic behavior of the randomization test in high dimensional setting. It turns out that even if the symmetry assumption is violated, the randomization test has correct level asymptotically. The asymptotic power function is also given. With fast implementation and good theoretical properties, the randomization test can be recommended in practice.

Keywords: Randomization test, High dimension, Symmetry assumption, Asymptotic power function

*Corresponding author

Email address: xuxz@bit.edu.cn (Xingzhong Xu)

1. Introduction

Consider i.i.d. random sample $X_1, \dots, X_n \in \mathbb{R}^p$ which has means $\mu = (\mu_1, \dots, \mu_p)^T$ and covariance matrix Σ. The one sample testing problem

$$H_0 : \mu = 0_p \quad \text{versus} \quad H_1 : \mu \neq 0_p \tag{1}$$

has been extensively studied by many researchers. A classical test statistic is Hotelling's T^2, $T_1 = n\bar{X}^T S^{-1} \bar{X}$, where $\bar{X}$ and $S = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$ are the sample mean vector and sample covariance matrix, respectively. However, the high-dimensional case $d > n$ invalidates the Hotelling's test. Bai and Saranadasa (1996) is the first paper modifying Hotelling's T^2 statistic for testing (1) in high dimensional setting. This seminal paper removed S^{-1} from T_1 and proposed a test statistic based on $T_2 = \bar{X}^T \bar{X}$. Many subsequent papers relaxed the assymptions and generalized the idea of Bai and Saranadasa (1996). For example, Srivastava (2009) proposed a test based on $T_3 = \bar{X}^T [\mathrm{diag}(S)]^{-1} \bar{X}$, where $\mathrm{diag}(S)$ is a matrix with diagonal elements equal to that of S and off-diagonal elemnts equal to 0. Chen and Qin (2010) proposed a test based on $T_4 = \sum_{i \neq j} X_i^T X_j$. Wang et al. (2015) proposed a test based on $T_5 = \sum_{i \neq j} Y_i^T Y_j$, where Y_i is defined as $Y_i = X_i / \|X_i\|$ if $X_i \neq 0$; and $Y_i = 0$ if $X_i = 0$.

Note that the test statistics T_2–T_5 all can be written as a generalized quadratic form of data, see Jong (1987), and their theoretical proves mostly rely on an application of martingale central limit theorem (MCLT).

The critical value of existing high dimensional tests are often determined by asymptotical distribution. We will call it asymptotic method. Asymptotic method can guarantee the test level asymptotically. However, in many real world problems, e.g., gene testing (see Bradley Efron (2007)), sample size n may be very small. Hence the Type I error rates of Asymptotic method may be far away from nominal levels in this case.

The idea of randomization test dates back to Fisher (1936), which is a tool to determine the critical value for a given test statistic. Romano (1990) described a general construction of the randomization test. In high dimension setting, randomization test is widely used in applied statistics, see Bradley Efron (2007).

It's strength is in that the resulting test procedure has exact level under mild condition. Although there are many papers give theoretical analysis for fixed p case, to the best of our knowledge, there's no existing theoretical work concerning high dimension setting.

Suppose the distribution of X_i is symmetric about 0 under the null hypothesis. T is certain test statistic for testing (1). Let $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. Rademacher variables ($\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = 1/2$) which are independent of the data. The conditional distribution

$$\mathcal{L}(T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)|X_1, \ldots, X_n) \tag{2}$$

is the uniform distribution on 2^n values. The critical value of the randomization test is defined as the $1-\alpha$ quantile of the above distribution. More specifically, the test function equals to 1 or 0 if $T(X_1, \ldots, X_n)$ is greater or not greater than the critical value. The resulting test is a level α test if the distribution of X_i is symmetric under null. By some refinement of the test function when $T(X_1, \ldots, X_n)$ equals to the critical value, the test is exact, see Romano (1990). Since such extreme case occurs with little probability 2^{-n}, the refinement is often dropped in practice, which only losses a little power.

Equivalently, the randomization test can be implemented by p-value. Define

$$p(X_1, \ldots, X_n) = \Pr(T(\epsilon_1 X_1, \ldots, \epsilon_n X_n) \geq T(X_1, \ldots, X_n)|X_1, \ldots, X_n). \tag{3}$$

The randomization test rejects the null hypothesis if $p(X_1, \ldots, X_n) \leq \alpha$.

It's easy to see that the randomization version of Bai and Saranadasa (1996)'s test and Chen and Qin (2010)'s test are both equivelent to the randomization test based on

$$T(X_1, \ldots, X_n) = \sum_{j<i} X_i^T X_j. \tag{4}$$

In this paper, we will study this special statistic for illustration. From now on, $T(X_1, \ldots, X_n)$ will refer to (4). Other quadratic based statistic can be studied by similar method.

Our results show that even if the null distribution is not symmetric, the randomization test is still asymptotically exact under mild assumptions. We also give the local asymptotic power.

In fixed p setting, it's well known that randomization test has much higher time complexity than asymptotic method, which historically hampered it's use. Surprisingly, in high dimension setting the randomization test can be implemented as efficiently as asymptotic method.

Maybe the most widely used randomization method is the two sample permutation test. As Romano (1990) pointed out, the asymptotic property of randomization tests depends heavily on the particular problem and, the two sample case is quite distinct from the one sample case. The method used in this paper can not be applied to permutation test.

2. Randomization test

We assume, like Chen and Qin (2010) and Bai and Saranadasa (1996), the following multivariate model:

$$X_i = \mu + \Gamma Z_i \quad \text{for } i = 1, \ldots, n, \tag{5}$$

where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i \Gamma_i^T = \Sigma_i$ and $\{Z_i\}_{i=1}^n$ are m-variate i.i.d. random vectors satisfying $\mathrm{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = I_m$, the $m \times m$ identity matrix. Write $Z_i = (z_{i1}, \ldots, z_{im})^T$, we assume $\mathrm{E}(z_{ij}^4) = 3 + \Delta < \infty$ and

$$\mathrm{E}(z_{il_1}^{\alpha_1} z_{il_2}^{\alpha_2} \cdots z_{il_q}^{\alpha_q}) = \mathrm{E}(z_{il_1}^{\alpha_1}) \mathrm{E}(z_{il_2}^{\alpha_2}) \cdots \mathrm{E}(z_{il_q}^{\alpha_q}) \tag{6}$$

for a positive integer q such that $\sum_{l=1}^q \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$.

In the following, $T(X_1, \ldots, X_n)$ will be specialized to (4).

An important assumption in Chen and Qin (2010) is $\mathrm{tr}(\Sigma^4) = o\{\mathrm{tr}^2(\Sigma^2)\}$, which is equavelent to

$$\frac{\lambda_{\max}(\Sigma)}{\sqrt{\mathrm{tr}\Sigma^2}} \to 0. \tag{7}$$

Although Chen and Qin's results is for two sample case, their results can be proved similarly for one sample case. We restate their theorems:

Theorem 1. *Under* (5), (6), (7) *and local alternatives*

$$\mu^T \Sigma \mu = o(n^{-1} \mathrm{tr}\Sigma^2), \tag{8}$$

we have

$$\frac{T(X_1, \ldots, X_n) - \frac{n(n-1)}{2}\mu^T \mu}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}} \xrightarrow{\mathcal{L}} N(0,1). \tag{9}$$

And

Theorem 2. *Under* (5), (6), (7) *and*

$$n^{-1}\mathrm{tr}(\Sigma)^2 = o(\mu^T \Sigma \mu), \tag{10}$$

we have

$$\frac{T(X_1, \ldots, X_n) - \frac{n(n-1)}{2}\mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \xrightarrow{\mathcal{L}} N(0,1). \tag{11}$$

We will call the conditional distribution

$$\mathcal{L}\Big(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n} (X_i^T X_j)^2}} \Big| X_1, \ldots, X_n\Big)$$

the randomization distribution. Let ξ_α^* be the $1-\alpha$ quantile of the randomization distribution. Then the test function $\phi(X_1, \ldots, X_n)$ of randomization test equals to 1 when

$$\frac{T(X_1, \ldots, X_n)}{\sqrt{\sum_{1\le j<i\le n} (X_i^T X_j)^2}} > \xi_\alpha^*$$

and equals to 0 otherwise. Since ξ_α^* relies on data, the rejection region is determined by not only $T(X_1, \ldots, X_n)$ but also randomization distribution.

To study the asymptotic property of ξ_α^*, we need to derive the asymptotic behavior of randomization distribution. Since the randomization distribution itself is random, we need to define in what sence the convergence is. Let F and G be two distribution function on $\mathbb{R}$ and the Levy metric ρ of F and G is defined as

$$\rho(F, G) = \inf\{\epsilon : F(x-\epsilon) - \epsilon \le G(x) \le F(x+\epsilon) + \epsilon \quad \text{for all} \quad x\}.$$

It's well known that $\rho(F_n, F) \to 0$ if and only if $F_n \xrightarrow{\mathcal{L}} F$. Our first result is:

Theorem 3. *Under* (5)*,* (6)*,* (7) *and*

$$\mu^T\mu = o\big(\sqrt{\mathrm{tr}\Sigma^2}\big), \tag{12}$$

we have that

$$\rho\Big(\mathcal{L}\Big(\frac{T_2(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}\big(X_i^T X_j\big)^2}}\Big| X_1, \ldots, X_n\Big), N(0,1)\Big) \xrightarrow{P} 0.$$

Under Theorem 3's conditions, the randomization distribution is asymptotically a normal distribution. It's also interesting to understand the behavior of randomization distribution when condition (12) is not valid. We have the following asymptotic result:

Theorem 4. *Under* (5)*,* (6)*,* (7) *and*

$$\sqrt{\mathrm{tr}\Sigma^2} = o\big(\mu^T\mu\big), \tag{13}$$

we have

$$\rho\Big(\mathcal{L}\Big(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}\big(X_i^T X_j\big)^2}}\Big| X_1, \ldots, X_n\Big), \frac{\sqrt{2}}{2}\big(\chi_1^2 - 1\big)\Big) \xrightarrow{P} 0. \tag{14}$$

Once the asymptotic distribution of the randomization distribution is obtained, the asymptotic behaivior of ξ_α^* can be derived immediately. Let $\Phi(\cdot)$ be the cumulative distribution function (CDF) of standard normal distribution, we have

Corollary 1. *Under the conditions of Theorem 3,* $\xi_\alpha^* \xrightarrow{P} \Phi(1-\alpha)$.

Corollary 2. *Under the conditions of Theorem 4,* $\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2}(\Phi^{-1}(1-\frac{\alpha}{2}) - 1)$.

Now we are ready to derive the asymptotic power of randomization test. Since the limit property of T is different under (8) and (10). The following two theorems give the power under (8) and (10), separately.

Theorem 5. *Suppose conditions* (5)*,* (6)*,* (7) *and* (8) *holds.*

If (12) *holds,*

$$\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}>\xi_\alpha^*\Big)=\Phi(-\Phi^{-1}(1-\alpha)+\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}})+o(1).$$

If (13) *holds,*

$$\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}>\xi_\alpha^*\Big)\to 1.$$

Theorem 6. *Under* (5), (6), (7) (10) *and either* (12) *or* (13),

$$\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}>\xi_\alpha^*\Big)=\Phi(\frac{\sqrt{n}\mu^T\mu}{2\sqrt{\mu^T\Sigma\mu}})+o(1),$$

Remark 1. The theorem doesn't assume that the distribution of X_i is symmetric under null. Hence Theorem 5 shows that the level of randomization test is robust agianst asymmetry.

Remark 2. Neither (8) or (12) implies the other one. For example, suppose $\Sigma=I_p$, then (8) is equivalent to $\mu^T\mu=o(p/n)$ and (12) is equivalent to $\mu^T\mu=o(\sqrt{p})$. In this case, if $\sqrt{p}/n\to 0$, then (8) implies (12); conversely, if $\sqrt{p}/n\to\infty$, then (12) implies (8).

The randomization test rejects the null when

$$T(X_1,\ldots,X_n)>\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}\xi_\alpha^*.$$

The asymptotic method rejects the null when

$$T(X_1,\ldots,X_n)>\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}\Phi^{-1}(1-\alpha).$$

Note that the larger the reject region, the more powerful the test is. Thus we compare $\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}\xi_\alpha^*$ and $\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\hat{\Sigma}^2}\Phi^{-1}(1-\alpha)$. Suppose $\mathrm{tr}\hat{\Sigma}^2$ is a ratio consistent estimator of $\mathrm{tr}\Sigma^2$. By Lemma 5 in appendix,

$$\frac{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}\xi_\alpha^*}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\hat{\Sigma}^2}\Phi^{-1}(1-\alpha)}=(1+o_P(1))\frac{\sqrt{\mathrm{tr}(\Sigma+\mu\mu^T)^2}\xi_\alpha^*}{\sqrt{\mathrm{tr}\Sigma^2}\Phi^{-1}(1-\alpha)},$$

which tends to 1 when (12) holds, and tends to ∞ when (13) holds. Hence randomization may loss some power.

3. Algorithm

Because the randomized test statistic is (conditionally) uniformly distributed on 2^n different values, the quantile ξ^*_α is not computationally feasible even when n is moderate. In practice, randomization test is often realized through an approximation of p-value (3). More specifically, we sample $\epsilon^*_1, \ldots, \epsilon^*_n$ and compute $T^* = T(\epsilon^*_1 X_1, \ldots, \epsilon^*_n X_n)$. Repeat B times for a large B and we obtain T^*_i, $i = 1, \ldots, B$. Let

$$\tilde{p}(X_1, \ldots, X_n) = \frac{1}{B+1}\big(1 + \sum_{i=1}^{B} \mathbf{1}_{\{T^*_i \geq T_0\}}\big).$$

It can be shown

$$\Pr(\tilde{p}(X_1, \ldots, X_n) \leq u) \leq u \quad \text{for all } 0 \leq u \leq 1,$$

and as $B \to \infty$, $\tilde{p}(X_1, \ldots, X_n) - p(X_1, \ldots, X_n) \xrightarrow{P} 0$ where $p(X_1, \ldots, X_n)$ is defined by (3) (see E. L. Lehmann (2005) page 636).

Note that

$$T^* = T(\epsilon^*_1 X_1, \ldots, \epsilon^*_n X_n) = \sum_{1 \leq j < i \leq n} X_i^T X_j \epsilon^*_i \epsilon^*_j.$$

We need to compute T^* for B times, which is seemingly time consuming. Nevertheless, $X_i^T X_j$ $(1 \leq j < i \leq n)$ can be computed in advance, which has complexity $O(n^2 p)$ and may be the most time consuming part in large p small n case. Once we have obtained $X_i^T X_j$, the computation of T^*_i has only complexity $O(n^2)$. The total complexity is thus $O(n^2 p + n^2 B)$. The randomization procedure only add minor complexity when p is large. The randomization method doesn't need an estimator of variance, which is a must need in asymptotic method. A good variance estimator is complicated in form, see Chen and Qin (2010), which add much time complexity to asymptotic method. Hence the randomization method is very competitive compared to asymptotic method even in the sense of computation efficiency. This is different from low dimension setting where randomization is a lot slower than asymptotic method.

If we only care about the decision (reject or accept) and the p-value is not needed, the computation efficiency of randomization method can be further improved. In fact, the rejection region is $\tilde{p}(X_1, \ldots, X_n) \leq \alpha$ or

$$\sum_{i=1}^{B} (1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B - (B+1)\alpha + 1.$$

Note that the left hand side is a sum of non-negative vaues. We can reject the null once $\sum_{i=1}^{B_0} (1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B - (B+1)\alpha + 1$ for some B_0. Similarly, the acception region is

$$\sum_{i=1}^{B} \mathbf{1}_{\{T_i^* \geq T_0\}} > (B+1)\alpha - 1.$$

we can accept the null once the sum of left hand side exceeds the right hand side. The algorithm 1 summarizes our previous argument.

4. Simulation studies

In this section, we report the simulation performance of the randomization test in various setting. We compare randomization method with asymptotic method. Asymptotic method needs a ratio consistent estimator of $\mathrm{tr}\Sigma^2$. However, a good estimator is often complicated and thus time-consuming. For simplicity, throughout the simulation the asymptotic method is implemented by using true $\mathrm{tr}\Sigma^2$, which will in principle perform better. The empirical power and size are computed based on 1000 simulations.

Let $c = \sqrt{n(n-1)}\mu^T\mu/\sqrt{2\mathrm{tr}\Sigma^2}$ be the signal to noise ratio (SNR). The theoretic asymptotic power is an increasing function of SNR. Throughout the simulation, we will scale μ to reach different level of SNR. Our simulation consider two mean structure: dense mean and sparse mean. In the dense mean setting, each coordinate of μ is independently generated from $U(2,3)$ and then μ is scaled to reach a given SNR. In the sparse mean setting, we randomly select 5% of μ p coordinates to be non-zero. Each non-zero coordinate is again independently generated from $U(2,3)$ and then scaled to reach a given SNR.

We consider two innovation structure: the moving average model and the factor model.

Algorithm 1 Randomization Algorithm

Require: α, B

  for $1 \leq j < i \leq n$ **do**

    $D_{ij} \leftarrow X_i^T X_j$

  end for

  $T_0 \leftarrow \sum_{1 \leq j < i \leq n} D_{ij}$

  Set $A \leftarrow 0$.

  for $i = 1$ to B **do**

    Generate $\epsilon_1, \ldots, \epsilon_n$ according to $\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = \frac{1}{2}$.

    if $\sum_{1 \leq j < i \leq n} D_{ij} \epsilon_i \epsilon_j \geq T_0$ **then**

      $A \leftarrow A + 1$

      if $A > (B+1)\alpha - 1$ **then**

        return Accept

      end if

    else

      if $i - A \geq B - (B+1)\alpha + 1$ **then**

        return Reject

      end if

    end if

  end for

We first conduct the moving average model. In model (5), we set $m = p + k$ and $(\Gamma)_{ij} = \rho_{j-i}$ for $j = i, \ldots, i + k$ and $(\Gamma)_{ij} = 0$ otherwise. More precisely,

$$X_{ij} = \sum_{l=0}^{k} \rho_l Z_{i,j+l} \tag{15}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Here Z_{ij}'s are i.i.d. random variables with distribution F for $i = 1, \ldots, n$ and $j = 1, \ldots, p + k$. Like Chen and Qin (2010), we consider two different F. One is $N(0, 1)$, and the other is $(\text{Gamma}(4, 1) - 4)/2$. We set $k = 5$ and $k = p$. The ρ_i's are generated independently from $U(2, 3)$ and kept fixed throughout the simulation. Table 1 and Table 2 list the empirical power and size for the moving average model. It's not surprising that randomization test can control level well in normal case since it can be proved in theory. The results also show that the randomization method can control level well even in Gamma case, which is not symmetric under null. It justify the robustness of randomization method. On the other hand, asymptotic method has small size when dependence is weak and has inflated size when dependence is strong. Randomization method also shares similar power with asymptotic method.

In the simulation study of Fan et al. (2007), data are generated from a factor model to reflect aspects of gene expression data. The model involves three group factor and one common factor among all p variables. Their data generation mechanism is adopted in our next simulation study. We denote by $\{\epsilon_{ij}\}_{1 \le i \le n, 1 \le j \le p}$ a sequence of independent $N(0, 1)$ and by $\{\chi_{ij}\}_{1 \le i \le n, 1 \le j \le 4}$ a sequence of independent random variables with distribution $(\chi_6^2 - 6)/\sqrt{12}$. Note that χ_{ij} has mean 0, variance 1 and skewness $\sqrt{12}/3$. The data is generated by model

$$X_{i,j} = \frac{a_{j1}\chi_{i1} + a_{j2}\chi_{i2} + a_{j3}\chi_{i3} + b_j\chi_{i4} + \epsilon_{ij}}{(1 + a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + b_j^2)^{1/2}} \quad i = 1, \ldots, n \text{ and } j = 1, \ldots, p,$$

where $a_{jk} = 0$ except that $a_{j1} = a_j$ for $j = 1, \ldots, \frac{1}{3}p$, $a_{j2} = a_j$ for $\frac{1}{3}p + 1, \ldots, \frac{2}{3}p$ and $a_{j3} = a_j$ for $\frac{2}{3}p + 1, \ldots, p$. As in Fan et al. (2007), we consider two configurations of factor loadings. In case I we set $a_j = 0.25$ and $b_j = 0.1$ for

Table 1: Empirical power and size of moving average model with normal innovation. $p = 600$, $n = 100$. TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

		Dense means				Sparse means			
		$k = 3$		$k = 500$		$k = 3$		$k = 500$	
SNR	TP	RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.040	0.006	0.042	0.065	0.048	0.015	0.047	0.059
0.5	0.126	0.190	0.069	0.125	0.169	0.146	0.063	0.081	0.111
1.0	0.260	0.429	0.235	0.218	0.262	0.361	0.175	0.145	0.193
1.5	0.442	0.629	0.425	0.314	0.384	0.633	0.413	0.229	0.313
2.0	0.639	0.795	0.632	0.399	0.458	0.817	0.652	0.415	0.524
2.5	0.804	0.923	0.828	0.469	0.529	0.960	0.861	0.645	0.839
3.0	0.912	0.966	0.919	0.582	0.647	0.998	0.975	0.819	0.999

$j = 1, \ldots, p$ and in case II a_i and b_i are generated independently from $U(0, 0.4)$ and $U(0, 0.2)$. Table 3 list the simulation results. Although the distribution is not symmetric, the results show that the level of randomization method is robust while asymptotic method suffers from level inflation.

5. Conclusion remark

6. Appendix

CLT for quadratic form of Rademacher varaibles. The proof of the Theorem 3 is based on a CLT of the quadratic form of Rademacher variables. Such a CLT can be also used to study the asymptotic behavior of many other randomization test. Let $\epsilon_1, \ldots, \epsilon_n$ be indepent Rademacher variables. Consider quadratic form $W_n = \sum_{1\leq j<i\leq n} a_{ij}\epsilon_i\epsilon_j$, where $\{a_{ij}\}$ are nonrandom numbers. Here $\{\epsilon_i\}$ and $\{a_{ij}\}$ may depend on n, a parameter we suppress. Obviously, $\mathrm{E}(W_n) = 0$ and $\mathrm{Var}(W_n) = \sum_{1\leq j<i\leq n} a_{ij}^2$.

Table 2: Empirical power and size of moving average model with Gamma innovation. $p = 600$, $n = 100$. TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

		Dense means				Sparse means			
		$k = 3$		$k = 500$		$k = 3$		$k = 500$	
SNR	TP	RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.050	0.016	0.043	0.060	0.033	0.005	0.050	0.069
0.5	0.126	0.211	0.077	0.128	0.160	0.159	0.062	0.065	0.098
1.0	0.260	0.410	0.228	0.255	0.308	0.386	0.159	0.151	0.188
1.5	0.442	0.609	0.418	0.318	0.375	0.643	0.410	0.231	0.321
2.0	0.639	0.799	0.634	0.392	0.449	0.875	0.685	0.373	0.501
2.5	0.804	0.912	0.809	0.496	0.550	0.961	0.858	0.589	0.828
3.0	0.912	0.963	0.908	0.543	0.596	0.985	0.967	0.843	1.000

Proposition 1. *A sufficient condition for*

$$\frac{W_n}{\sqrt{\sum_{1\le j<i\le n} a_{ij}^2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{16}$$

is that

$$\sum_{j<k}\Big(\sum_{i:i>k} a_{ij}a_{ik}\Big)^2 + \sum_{j<i} a_{ij}^4 + \sum_{j<k<i} a_{ij}^2 a_{ik}^2 = o\Big(\big(\sum_{j<i} a_{ij}^2\big)^2\Big). \tag{17}$$

Proof. Define $U_{in} = \epsilon_i \sum_{j=1}^{i-1} a_{ij}\epsilon_j$, $i = 2, \ldots, n$, and $\mathcal{F}_{in} = \sigma\{\epsilon_1, \ldots, \epsilon_i\}$, $i = 1, \ldots, n$. Now $W_n = \sum_{i=2}^n U_{in}$, $\{U_{in}\}$ is a martingale difference array with respect to $\{\mathcal{F}_{in}\}$. To prove the proposition, we shall verify two conditions (See Pollard (1984)):

$$\frac{\sum_{i=2}^n \mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n})}{\sum_{1\le j<i\le n} a_{ij}^2} \xrightarrow{P} 1, \tag{18}$$

and

$$\frac{\sum_{i=2}^n \mathrm{E}\big(U_{in}^2\{U_{in}^2 > \epsilon \sum_{1\le j<i\le n} a_{ij}^2\}|\mathcal{F}_{i-1,n}\big)}{\sum_{1\le j<i\le n} a_{ij}^2} \xrightarrow{P} 0, \tag{19}$$

for every $\epsilon > 0$.

Table 3: Empirical power and size of factor model innovation. $p = 600$, $n = 100$. TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

		Dense means				Sparse means			
		Case I		Case II		Case I		Case II	
SNR	TP	RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.040	0.050	0.046	0.057	0.052	0.061	0.047	0.060
0.5	0.126	0.141	0.166	0.136	0.154	0.104	0.126	0.105	0.129
1.0	0.260	0.222	0.256	0.233	0.266	0.207	0.239	0.215	0.240
1.5	0.442	0.363	0.411	0.369	0.392	0.367	0.411	0.373	0.415
2.0	0.639	0.460	0.501	0.470	0.509	0.517	0.581	0.532	0.589
2.5	0.804	0.582	0.613	0.565	0.607	0.718	0.757	0.728	0.780
3.0	0.912	0.646	0.678	0.673	0.701	0.834	0.889	0.870	0.900

Proof of (18). Since $\mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n}) = \big(\sum_{j=1}^{i-1} a_{ij}\epsilon_j\big)^2$, we have

$$\sum_{i=2}^{n} \mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n}) = \sum_{i=2}^{n} \big(\sum_{j=1}^{i-1} a_{ij}\epsilon_j\big)^2 = \sum_{i=2}^{n} \big(\sum_{j=1}^{i-1} a_{ij}^2 + 2 \sum_{j,k:j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k\big)$$
$$= \sum_{i=2}^{n}\sum_{j=1}^{i-1} a_{ij}^2 + 2 \sum_{j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k.$$

But

$$\mathrm{E}\big(\sum_{j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k\big)^2 = \mathrm{E}\big(\sum_{j<k}\big(\sum_{i:i>k} a_{ij}a_{ik}\big)\epsilon_j\epsilon_k\big)^2$$
$$= \sum_{j<k}\big(\sum_{i:i>k} a_{ij}a_{ik}\big)^2 = o\big(\big(\sum_{j<i} a_{ij}^2\big)^2\big),$$

where the last equality holds by assumption. Hence (18) holds.

Proof of (19). By Markov's inequality, we only need to prove

$$\frac{\sum_{i=2}^{n} \mathrm{E}\big(U_{in}^4|\mathcal{F}_{i-1,n}\big)}{\big(\sum_{1\le j<i\le n} a_{ij}^2\big)^2} \xrightarrow{P} 0. \tag{20}$$

Since the relavant random variables are all positive, we only need to prove (20) converges to 0 in mean. But

$$\begin{aligned}
\sum_{i=2}^{n} \mathrm{E} U_{in}^4 &= \sum_{i=2}^{n} \mathrm{E}(\sum_{j:j<i} a_{ij}\epsilon_j)^4 = \sum_{i=2}^{n} \mathrm{E}(\sum_{j:j<i} a_{ij}^2 + 2 \sum_{j,k:j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k)^2 \\
&= \sum_{i=2}^{n} \big((\sum_{j:j<i} a_{ij}^2)^2 + 4\mathrm{E}(\sum_{j,k:j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k)^2\big) \\
&= \sum_{i=2}^{n}(\sum_{j:j<i} a_{ij}^4 + 6 \sum_{j,k:j<k<i} a_{ij}^2 a_{ik}^2) \\
&= \sum_{j<i} a_{ij}^4 + 6 \sum_{j<k<i} a_{ij}^2 a_{ik}^2 = o\big((\sum_{j<i} a_{ij}^2)^2\big),
\end{aligned}$$

where the last equality holds by assumption. Hence (19) holds. □

Asymptotic normality.

Lemma 1. *Suppose $\{\eta_n\}$ is a sequence of 1-dimensional random variables, weakly converges to η, a random variable with continuous distribution function. Then we have*

$$\sup_x |\Pr(\eta_n \le x) - \Pr(\eta \le x)| \to 0.$$

Lemma 2. *Suppose $A = (a_{ij})$ is an $m \times m$ positive semi-definite matrix. Under (6), we have*

$$\mathrm{E}(Z_i^T A Z_i)^2 \asymp (\mathrm{tr} A)^2 \tag{21}$$

Proof. Notice that

$$\begin{aligned}
(Z_i^T A Z_i)^2 =&(\sum_{j=1}^{m} a_{jj} z_{ij}^2 + 2\sum_{k<j} a_{jk} z_{ij} z_{ik})^2 \\
=&(\sum_{j=1}^{m} a_{jj} z_{ij}^2)^2 + 4(\sum_{j=1}^{m} a_{jj} z_{ij}^2)(\sum_{k<j} a_{jk} z_{ij} z_{ik}) + 4(\sum_{k<j} a_{jk} z_{ij} z_{ik})^2 \\
=&\sum_{j=1}^{m} a_{jj}^2 z_{ij}^4 + 2\sum_{k<j} a_{jj} a_{kk} z_{ij}^2 z_{ik}^2 + 4(\sum_{j=1}^{m} a_{jj} z_{ij}^2)(\sum_{k<j} a_{jk} z_{ij} z_{ik}) \\
&+ 4(\sum_{k<j} a_{jk}^2 z_{ij}^2 z_{ik}^2 + \sum_{k<j, l<\alpha:\mathrm{card}(\{k,j\}\cap\{l,\alpha\})<2} a_{jk} a_{\alpha l} z_{ij} z_{ik} z_{i\alpha} z_{il})
\end{aligned} \tag{22}$$

Hence

$$\begin{aligned}\mathrm{E}(Z_i^T A Z_i)^2 =& \sum_{j=1}^n a_{jj}^2 \mathrm{E} z_{ij}^4 + 2\sum_{k<j} a_{jj}a_{kk}\mathrm{E}(z_{ij}^2 z_{ik}^2) + 4\sum_{k<j} a_{jk}^2 \mathrm{E}(z_{ij}^2 z_{ik}^2) \\ \asymp& \sum_{j=1}^n\sum_{k=1}^n a_{jj}a_{kk} + \sum_{j=1}^n\sum_{k=1}^n a_{jk}^2 = (\mathrm{tr}(A))^2 + \mathrm{tr}A^2.\end{aligned} \tag{23}$$

By Cauchy inequality, $0 \le \mathrm{tr}A^2 \le (\mathrm{tr}A)^2$. The conclusion holds.

□

Lemma 3. *Under* (5), (6), *for* $i \neq j$ *we have*

$$\mathrm{E}(X_i^T X_j)^4 = O(1)\Big(\mathrm{tr}\big(\Sigma + \mu\mu^T\big)^2\Big)^2. \tag{24}$$

Proof.

$$\begin{aligned}(X_i^T X_j)^4 =& (Z_i^T\Gamma^T\Gamma Z_j + \mu^T\Gamma Z_i + \mu^T\Gamma Z_j + \mu^T\mu)^4 \\ \le& 64\big((Z_i^T\Gamma^T\Gamma Z_j)^4 + (\mu^T\Gamma Z_i)^4 + (\mu^T\Gamma Z_j)^4 + (\mu^T\mu)^4\big)\end{aligned} \tag{25}$$

We deal with the first term by applying Lemma 2 twice.

$$\begin{aligned}\mathrm{E}(Z_i^T\Gamma^T\Gamma Z_j)^4 =& \mathrm{E}(Z_i^T\Gamma^T\Gamma Z_j Z_j^T\Gamma^T\Gamma Z_i)^2 = \mathrm{E}\mathrm{E}\big((Z_i^T\Gamma^T\Gamma Z_j Z_j^T\Gamma^T\Gamma Z_i)^2 | Z_j\big) \\ \asymp& \mathrm{E}(Z_j^T\Gamma^T\Sigma\Gamma Z_j)^2 \asymp (\mathrm{tr}\Sigma^2)^2,\end{aligned} \tag{26}$$

Similarly, we have

$$\begin{aligned}&\mathrm{E}(\mu^T\Gamma Z_i)^4 = \mathrm{E}(Z_i^T\Gamma^T\mu\mu^T\Gamma Z_i)^2 \asymp (\mu^T\Sigma\mu)^2 \\ &\le \lambda_{\max}^2(\Sigma)(\mu^T\mu)^2 \le \mathrm{tr}(\Sigma^2)(\mu^T\mu)^2 \le \big(\mathrm{tr}(\Sigma^2)\big)^2 + (\mu^T\mu)^4.\end{aligned} \tag{27}$$

Thus, the conclusion holds. □

Lemma 4. *Under* (5), (6), *suppose* $i \neq j$, $i \neq k$, $j \neq k$, *we have*

$$\mathrm{E}(X_i^T X_j)^2 (X_k^T X_i)^2 = O(1)\Big(\mathrm{tr}\big(\Sigma + \mu\mu^T\big)^2\Big)^2. \tag{28}$$

Proof. Note that

$$\begin{aligned}&\mathrm{E}(X_i^T X_j)^2 (X_k^T X_i)^2 = \mathrm{E}\mathrm{E}\big((X_i^T X_j)^2 (X_k^T X_i)^2 | X_i\big) = \mathrm{E}(X_i^T(\Sigma + \mu\mu^T)X_i)^2 \\ =& \mathrm{E}\big(Z_i^T\Gamma^T(\Sigma + \mu\mu^T)\Gamma Z_i + 2\mu^T(\Sigma + \mu\mu^T)\Gamma Z_i + \mu\Sigma\mu + (\mu^T\mu)^2\big)^2 \\ \le& 4\mathrm{E}(Z_i^T\Gamma^T(\Sigma + \mu\mu^T)\Gamma Z_i)^2 + 16\mathrm{E}(\mu^T(\Sigma + \mu\mu^T)\Gamma Z_i)^2 + 4(\mu\Sigma\mu)^2 + 4(\mu^T\mu)^4.\end{aligned}$$

By Lemma (2),

$$\begin{aligned}\mathrm{E}(Z_i^T\Gamma^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2 &\asymp \big(\mathrm{tr}(\Gamma^T(\Sigma+\mu\mu^T)\Gamma)\big)^2\\ &= \big(\mathrm{tr}\Sigma^2+\mu^T\Sigma\mu\big)^2 \le 2\big(\mathrm{tr}\Sigma^2\big)^2+2\big(\mu^T\Sigma\mu\big)^2.\end{aligned}$$

And

$$\mathrm{E}(\mu^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2 = \mu^T(\Sigma+\mu\mu^T)\Sigma(\Sigma+\mu\mu^T)\mu$$

$$=\mu^T\Sigma^3\mu+2(\mu^T\mu)(\mu^T\Sigma^2\mu)+(\mu^T\mu)^2(\mu^T\Sigma\mu).$$

But for $i=1,2,\ldots$, we have

$$\mu^T\Sigma^i\mu \le \lambda_{\max}^i(\Sigma)\mu^T\mu \le (\mathrm{tr}(\Sigma^2))^{i/2}\mu^T\mu.$$

Thus,

$$\begin{aligned}&\mathrm{E}(X_i^TX_j)^2(X_k^TX_i)^2\\ =&O(1)\big((\mathrm{tr}\Sigma^2)^2+(\mathrm{tr}\Sigma^2)^{3/2}(\mu^T\mu)+(\mathrm{tr}\Sigma^2)(\mu^T\mu)^2+(\mathrm{tr}\Sigma^2)^{1/2}(\mu^T\mu)^3+(\mu^T\mu)^4\big)\\ =&O(1)\Big(\mathrm{tr}\big(\Sigma+\mu\mu^T\big)^2\Big)^2.\end{aligned} \tag{29}$$

□

Lemma 5. *Under* (5) *and* (6)*, we have*

$$\frac{\sum_{j<i}(X_i^TX_j)^2}{\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2} \xrightarrow{P} 1. \tag{30}$$

Proof. Since

$$\begin{aligned}\mathrm{E}(X_i^TX_j)^2 =&\mathrm{E}(X_i^TX_jX_j^TX_i) = \mathrm{E}(X_i^T(\Sigma+\mu\mu^T)X_i)\\ =&\mathrm{E}\mathrm{tr}((\Sigma+\mu\mu^T)X_iX_i^T) = \mathrm{tr}(\Sigma+\mu\mu^T)^2,\end{aligned} \tag{31}$$

we have

$$\mathrm{E}\sum_{j<i}(X_i^TX_j)^2 = \frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2 \tag{32}$$

So we only need to consider the variance. According to $\mathrm{card}(\{i,j\}\cap\{k,l\}) = 0,1,2$, we have

$$\begin{aligned}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2 =& \sum_{j<i}(X_i^TX_j)^4 + \sum_{j<i,k<l:\{i,j\}\cap\{k,l\}=\phi}(X_i^TX_j)^2(X_k^TX_l)^2\\ &+2\sum_{j<i<k}\big((X_i^TX_j)^2(X_k^TX_i)^2+(X_i^TX_j)^2(X_k^TX_j)^2+(X_k^TX_j)^2(X_k^TX_i)^2\big).\end{aligned} \tag{33}$$

In (33), there are $n(n-1)/2$, $n(n-1)(n-2)(n-3)/4$ and $n(n-1)(n-2)/6$ terms in each summation. By Lemma 3 and Lemma 4, we have

$$\begin{aligned}\mathrm{E}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2 =&\frac{n(n-1)(n-2)(n-3)}{4}\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\\ &+O(1)\big(\frac{n(n-1)}{2}+n(n-1)(n-2)\big)\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2.\end{aligned} \tag{34}$$

Hence

$$\frac{\mathrm{Var}(\sum_{j<i}(X_i^TX_j)^2)}{\big(\mathrm{E}\sum_{j<i}(X_i^TX_j)^2\big)^2}=\frac{\mathrm{E}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2-\big(\mathrm{E}\sum_{j<i}(X_i^TX_j)^2\big)^2}{\big(\mathrm{E}\sum_{j<i}(X_i^TX_j)^2\big)^2}=O(\frac{1}{n}).$$

Thus the conclusion holds. $\square$

Lemma 6. *Under* (5), (6), (7) *and*

$$\mu^T\mu=o(\sqrt{\mathrm{tr}\Sigma^2}), \tag{35}$$

we have

$$\sum_{j<k}\big(\sum_{i:i>k}X_i^TX_jX_i^TX_k\big)^2=o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big). \tag{36}$$

$$\sum_{j<k}(X_i^TX_j)^4=o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{37}$$

$$\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2=o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{38}$$

Proof.

$$\begin{aligned}&\mathrm{E}\sum_{j<k}\big(\sum_{i:i>k}X_i^TX_jX_i^TX_k\big)^2\\ =&\mathrm{E}\sum_{j<k}\Big(\sum_{i:i>k}(X_i^TX_j)^2(X_i^TX_k)^2+2\sum_{i_1,i_2:i_1>i_2>k}X_{i_1}^TX_jX_{i_1}^TX_kX_{i_2}^TX_jX_{i_2}^TX_k\Big).\end{aligned}$$

By Lemma 3, we have

$$\mathrm{E}\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2=O(n^3)\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2.$$

And

$$
\begin{aligned}
&\mathrm{E}\sum_{j<k<i_2<i_1} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \\
=&\frac{n(n-1)(n-2)(n-3)}{6}\mathrm{tr}(\Sigma+\mu\mu^T)^4 \\
\le&\frac{n(n-1)(n-2)(n-3)}{6}8(\mathrm{tr}(\Sigma)^4+(\mu^T\mu)^4) \\
\le&O(n^4)(\lambda_{\max}^2(\Sigma)\mathrm{tr}(\Sigma)^2+(\mu^T\mu)^4) \\
=&o\Big(n^4\big(\mathrm{tr}\Sigma^2\big)^2\Big),
\end{aligned}
$$

where the last line follows by assumption (7) and (35). Thus (36) holds. (37) and (38) follow by Lemma 3 and Lemma 4.

□

Proof of Theorem 3. By a standard subsequence argument, we only need to prove

$$
\rho\Big(\mathcal{L}\Big(\frac{T_2(\epsilon_1X_1,\ldots,\epsilon_iX_i,\ldots,\epsilon_nX_n)}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}\Big|X_1,\ldots,X_n\Big),N(0,1)\Big)\xrightarrow{a.s.}0 \tag{39}
$$

along a subsequence. But there is a subsequence $\{n(k)\}$ along which (36), (37) and (38) holds almost surely. By Proposition 1, we have

$$
\mathcal{L}\Big(\frac{T_2(\epsilon_1X_1,\ldots,\epsilon_iX_i,\ldots,\epsilon_nX_n)}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}\Big|X_1,\ldots,X_n\Big)\xrightarrow{\mathcal{L}}N(0,1)
$$

almost surely along $\{n(k)\}$, which means (39) holds along $\{n(k)\}$.

□

Proof of Theorem 5. Note that

$$
\begin{aligned}
&\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}>\xi_\alpha^*\Big) \\
=&\Pr\Big(\frac{T(X_1,\ldots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}>\xi_\alpha^*-\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}\Big)
\end{aligned} \tag{40}
$$

If (12) holds, by Lemma 5, we have

$$
\frac{\sum_{1\le j<i\le n}(X_i^TX_j)^2}{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}\xrightarrow{P}1.
$$

By Corollary 1, we have $\xi^*_\alpha \xrightarrow{P} \Phi(1-\alpha)$. Thus,

$$
\begin{aligned}
(40) =& \Pr\Big(\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}} - \frac{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}}\xi^*_\alpha > -\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}}\Big)\\
=& \Pr\Big(N(0,1) - \Phi(1-\alpha) > -\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}}\Big) + o(1)\\
=&\Phi(-\Phi(1-\alpha) + \frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}}) + o(1),
\end{aligned}
$$

where the last two equality holds by Theorem 1, Slutsky's theorem and Lemma 1.

If (13) holds, by Lemma 5, we have

$$
\frac{\sum_{1\le j<i\le n}(X_i^TX_j)^2}{\frac{n(n-1)}{2}\left(\mu^T\mu\right)^2} \xrightarrow{P} 1.
$$

Thus

$$
\begin{aligned}
&\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}\\
=&\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}}\frac{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}}{\sqrt{\frac{n(n-1)}{2}\left(\mu^T\mu\right)^2}}\frac{\sqrt{\frac{n(n-1)}{2}\left(\mu^T\mu\right)^2}}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}} \xrightarrow{P} 0.
\end{aligned}
$$

Since $\xi^*_\alpha \xrightarrow{P} \frac{\sqrt{2}}{2}(\Phi^{-1}(1-\frac{\alpha}{2})-1)$ by Corollary 2, we have that $(40) \to 1$.

□

Proof of Theorem 6. Note that

$$
\begin{aligned}
&\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}} > \xi^*_\alpha\Big)\\
=&\Pr\Big(\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}} > \frac{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\xi^*_\alpha - \frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\Big).
\end{aligned}
\tag{41}
$$

If (12) holds, the theorem follows by Theorem 2 and the fact that if (12) holds, the coefficient of ξ^*_α in (41) tends to 0.

If (13) holds, the theorem follows by noting that

$$(41) = \Pr\Big(\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2 n\mu^T\Sigma\mu}} > -(1+o_P(1))\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2 n\mu^T\Sigma\mu}}\Big).$$

□

Proof of Theorem 4.

$$\begin{aligned}\sum_{j<i} X_i^T X_j \epsilon_i\epsilon_j =& \sum_{j<i} Z_i^T\Gamma^T\Gamma Z_j\epsilon_i\epsilon_j \\ &+ \sum_{j<i}\mu^T\Gamma Z_i\epsilon_i\epsilon_j + \sum_{j<i}\mu^T\Gamma Z_j\epsilon_i\epsilon_j + \mu^T\mu\sum_{j<i}\epsilon_i\epsilon_j \\ \overset{def}{=}& C_1 + C_2 + C_3 + C_4.\end{aligned}$$

Term C_4 plays a major role. Note that

$$C_4 = \frac{n}{2}\mu^T\mu\Big(\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i\Big)^2 - 1\Big).$$

By central limit theorem, we have

$$\rho\Big(\mathcal{L}\Big(\frac{C_4}{\frac{n}{2}\mu^T\mu}\Big| X_1,\ldots,X_n\Big), \chi_1^2 - 1\Big) \xrightarrow{a.s.} 0.$$

Next we show that C_1, C_2 and C_3 are negligible under the assumptions of the theorem. By a standard subsequence argument and Slutsky's theorem, we can obtain

$$\rho\Big(\mathcal{L}\Big(\frac{T(\epsilon_1 X_1,\ldots,\epsilon_i X_i,\ldots,\epsilon_n X_n)}{\frac{n}{2}\mu^T\mu}\Big| X_1,\ldots,X_n\Big), \chi_1^2 - 1\Big) \xrightarrow{P} 0 \qquad (42)$$

by showing that

$$\mathrm{E}\Big(\Big(\frac{C_i}{\frac{n}{2}\mu^T\mu}\Big)^2\Big| X_1,\ldots,X_n\Big) \xrightarrow{P} 0, \quad i = 1,2,3.$$

It in turn suffices to show

$$\mathrm{E}\Big(\frac{C_i}{\frac{n}{2}\mu^T\mu}\Big)^2 \to 0, \quad i = 1,2,3. \qquad (43)$$

By direct calculation, we have

$$\mathrm{E}(C_1^2) = \mathrm{E}\mathrm{E}(C_1^2|X_1,\ldots,X_n) = \sum_{j<i}\mathrm{E}(Z_i^T\Gamma^T\Gamma Z_j)^2 = \frac{n(n-1)}{2}\mathrm{tr}\Sigma^2,$$

and

$$\mathrm{E}(C_2^2) = \mathrm{E}(C_3^2) = \frac{n(n-1)}{2}\mu^T\Sigma\mu \leq \frac{n(n-1)}{2}\sqrt{\mathrm{tr}\Sigma^2}\mu^T\mu.$$

Thus (43) follows by Assumption (13). Having (42) holds, the theorem follows by Slutsky's theorem, Lemma 5 and Assumption (13). □

References

Bai, Z., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statistica Sinica 6, 311–329.

Bradley Efron, R.J., 2007. On testing the significance of sets of genes. The Annals of Applied Statistics 1, 107–129.

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

E. L. Lehmann, J.P.R., 2005. Testing Statistical Hypotheses. Springer New York. doi:10.1007/0-387-27605-X.

Fan, J., Hall, P., Yao, Q., 2007. To how many simultaneous hypothesis tests can normal, student's t or bootstrap calibration be applied? Journal of the American Statistical Association 102, 1282–1288.

Fisher, R.A., 1936. The design of experiments. .

Jong, P.D., 1987. A central limit theorem for generalized quadratic forms. Probability Theory & Related Fields 75, 261–277.

Pollard, D., 1984. Convergence of stochastic processes.

Romano, J.P., 1990. On the behavior of randomization tests without a group invariance assumption. Journal of the American Statistical Association 85, 686–692.

Srivastava, M.S., 2009. A test for the mean vector with fewer observations than the dimension under non-normality. Journal of Multivariate Analysis

100, 518–532As the access to this document is restricted, you may want to look for a different version under ”Related research” (further below) orfor a different version of it.

Wang, L., Peng, B., Li, R., 2015. A high-dimensional nonparametric multivariate test for mean vector. Journal of the American Statistical Association 110, 00–00.