Elsevier LaTeX template[☆]

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], Global Customer Service[b,]*

[a]1600 John F Kennedy Boulevard, Philadelphia
[b]360 Park Avenue South, New York

Abstract

This template helps you to create a properly formatted LaTeX manuscript.

Keywords: `elsarticle.cls`, LaTeX, Elsevier, template

2010 MSC: 00-01, 99-00

1. Introduction

If the distribution of X_i is symmetric, then the randomization test is exact. We prove even if X_i is not symmetric, the randomization test is asymptotically exact.

We give a fast algorithm. Surprisingly, the randomization test can be implemented efficiently than asymptotically method since we don't need to estimate the variance of the test statistic.

2. Model

Consider i.i.d. random sample $X_1, \ldots, X_n \in \mathbb{R}^p$ which has means $\mu = (\mu_1, \ldots, \mu_p)^T$ and covariance matrix Σ. We consider testing the following high-

[☆]Fully documented templates are available in the elsarticle package on CTAN.

[*]Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)

[1]Since 1880.

dimensional hypothesis:

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

We assume, like chen qin and bai, the following multivariate model:

$$X_i = \mu + \Gamma Z_i \quad \text{for } i = 1, \dots, n, \tag{2}$$

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 $Z_i = (z_{i1}, \dots, z_{im})^T$, we assume $\mathrm{E}(z_{ij}^4) = 3 + \Delta < \infty$ and

$$\mathrm{E}(z_{il_1}^{\alpha_1} z_{il_2}^{\alpha_2} \cdots z_{il_q}^{\alpha_q}) = \mathrm{E}(z_{il_1}^{\alpha_1}) \mathrm{E}(z_{il_2}^{\alpha_2}) \cdots \mathrm{E}(z_{il_q}^{\alpha_q}) \tag{3}$$

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

Consider the test statistic

$$T = \|\bar{X}\|^2. \tag{4}$$

Statistics like this are studied in some high dimensional mean test literature. Most of existing papers determined the critical value by asymptotic distribution, which is not exact. Randomization method have the advantages that the test is exact. However, as far as we know, there's no existing work give a theoretical justification of the power behavior of randomization method.

Randomization. For a test statistic $T(X_1, \dots, X_n)$. Suppose $\epsilon_1, \dots, \epsilon_n$ are i.i.d. Rademacher variables ($\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = 1/2$) which are independent of the data. The critical value is defined as the $1 - \alpha$ quantile of the conditional distribution

$$\mathcal{L}(T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n) | X_1, \dots, X_n). \tag{5}$$

It's easy to see that the randomization test based on $\|\bar{X}\|^2$, $\|\bar{X}\|^2 - \frac{1}{n} S$ (S is the sample covariance matrix) and $\sum_{j<i} X_i^T X_j$ are equivalent. In what follows, we will consider

$$T_2(X_1, \dots, X_n) = \sum_{j<i} X_i^T X_j. \tag{6}$$

Then $T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n) = \sum_{j<i} X_i^T X_j \epsilon_i \epsilon_j$.

CLT for quadratic form of Rademacher varaibles. We will study the quadratic form of Rademacher variables. Let $\epsilon_1, \dots, \epsilon_n$ be indepent Rademacher variables. Consider quadratic form $W_n = \sum_{1\le j<i\le n} a_{ij}\epsilon_i\epsilon_j$, 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 $\mathrm{Var}(W_n) = \sum_{1\le j<i\le n} a_{ij}^2$.

Proposition 1. *A sufficient condition for*

$$\frac{W_n}{\sqrt{\sum_{1\le j<i\le n} a_{ij}^2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{7}$$

is that

$$\sum_{j<k}\Big(\sum_{i:i>k} a_{ij}a_{ik}\Big)^2 + \sum_{j<i} a_{ij}^4 + \sum_{j<k<i} a_{ij}^2 a_{ik}^2 = o\Big(\big(\sum_{j<i} a_{ij}^2\big)^2\Big). \tag{8}$$

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

$$\frac{\sum_{i=2}^n \mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n})}{\sum_{1\le j<i\le n} a_{ij}^2} \xrightarrow{P} 1, \tag{9}$$

and

$$\frac{\sum_{i=2}^n \mathrm{E}\big(U_{in}^2\{U_{in}^2 > \epsilon\sum_{1\le j<i\le n} a_{ij}^2\}|\mathcal{F}_{i-1,n}\big)}{\sum_{1\le j<i\le n} a_{ij}^2} \xrightarrow{P} 0, \tag{10}$$

for every $\epsilon > 0$.

Proof of (9). 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

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

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 (9) holds.

Proof of (10). By Markov's inequality, we only need to prove

$$\frac{\sum_{i=2}^n \mathrm{E}\big(U_{in}^4\big|\mathcal{F}_{i-1,n}\big)}{\big(\sum_{1\le j<i\le n} a_{ij}^2\big)^2} \xrightarrow{P} 0. \tag{11}$$

Since the relavant random variables are all positive, we only need to prove (11) converges to 0 in mean. But

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

where the last equality holds by assumption. Hence (10) holds. □

3. Asymptotic normality

Lemma 1. *Suppose $A = (a_{ij})$ is an $m\times m$ positive semi-definite matrix, then*

$$\mathrm{E}\big(Z_i^T A Z_i\big)^2 \asymp (\mathrm{tr}(A))^2 + \mathrm{tr}A^2 \tag{12}$$

Proof. Notice that

$$
\begin{aligned}
\left(Z_i^T A Z_i\right)^2 =& \left(\sum_{j=1}^{m} a_{jj} z_{ij}^2 + 2\sum_{k<j} a_{jk} z_{ij} z_{ik}\right)^2 \\
=& \left(\sum_{j=1}^{m} a_{jj} z_{ij}^2\right)^2 + 4\left(\sum_{j=1}^{m} a_{jj} z_{ij}^2\right)\left(\sum_{k<j} a_{jk} z_{ij} z_{ik}\right) + 4\left(\sum_{k<j} a_{jk} z_{ij} z_{ik}\right)^2 \\
=& \sum_{j=1}^{m} a_{jj}^2 z_{ij}^4 + 2\sum_{k<j} a_{jj} a_{kk} z_{ij}^2 z_{ik}^2 + 4\left(\sum_{j=1}^{m} a_{jj} z_{ij}^2\right)\left(\sum_{k<j} a_{jk} z_{ij} z_{ik}\right) \\
& + 4\left(\sum_{k<j} a_{jk}^2 z_{ij}^2 z_{ik}^2 + \sum_{k<j, l<\alpha:\mathrm{card}(\{i,j\}\cap\{l,\alpha\})<2} a_{jk} a_{\alpha l} z_{ij} z_{ik} z_{i\alpha} z_{il}\right)
\end{aligned}
\tag{13}
$$

Hence

$$
\begin{aligned}
\mathrm{E}\left(Z_i^T A Z_i\right)^2 =& \sum_{j=1}^{n} a_{jj}^2 \mathrm{E} z_{ij}^4 + 2\sum_{k<j} a_{jj} a_{kk} \mathrm{E}(z_{ij}^2 z_{ik}^2) + 4\sum_{k<j} a_{jk}^2 \mathrm{E}(z_{ij}^2 z_{ik}^2) \\
\asymp & \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jj} a_{kk} + \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}^2 = \left(\mathrm{tr}(A)\right)^2 + \mathrm{tr} A^2
\end{aligned}
\tag{14}
$$

□

Lemma 2. *Suppose condition* (32) *holds, then*

$$
\frac{\sum_{j<i}\left(X_i^T X_j\right)^2}{\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2} \xrightarrow{P} 1. \tag{15}
$$

Proof.

$$
\begin{aligned}
\mathrm{E}\left(X_i^T X_j\right)^2 =& \mathrm{E}(X_i^T X_j X_j^T X_i) = \mathrm{E}(X_i^T(\Sigma+\mu\mu^T)X_i) \\
=& \mathrm{Etr}((\Sigma+\mu\mu^T)X_i X_i^T) = \mathrm{tr}\left(\Sigma+\mu\mu^T\right)^2
\end{aligned}
\tag{16}
$$

Hence

$$
\mathrm{E}\sum_{j<i}\left(X_i^T X_j\right)^2 = \frac{n(n-1)}{2}\mathrm{tr}\left(\Sigma+\mu\mu^T\right)^2 \tag{17}
$$

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

$$
\begin{aligned}
\Big(\sum_{j<i}\left(X_i^T X_j\right)^2\Big)^2 =& \sum_{j<i}\left(X_i^T X_j\right)^4 + \sum_{j<i,k<l:\{i,j\}\cap\{k,l\}=\phi}\left(X_i^T X_j\right)^2\left(X_k^T X_l\right)^2 \\
& + 2\sum_{j<i<k}\Big(\left(X_i^T X_j\right)^2\left(X_k^T X_i\right)^2 + \left(X_i^T X_j\right)^2\left(X_k^T X_j\right)^2 + \left(X_k^T X_j\right)^2\left(X_k^T X_i\right)^2\Big)
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
{(X_i^TX_j)}^4 =&(Z_i^T\Gamma^T\Gamma Z_j+\mu^T\Gamma Z_i+\mu^T\Gamma Z_j+\mu^T\mu)^4\\
\leq& 64\big((Z_i^T\Gamma^T\Gamma Z_j)^4+(\mu^T\Gamma Z_i)^4+(\mu^T\Gamma Z_j)^4+(\mu^T\mu)^4\big)
\end{aligned}
\tag{19}
$$

Note that by lemma 1, we have

$$
\mathrm{E}(\mu^T\Gamma Z_i)^4=\mathrm{E}(Z_i^T\Gamma^T\mu\mu^T\Gamma Z_i)^2\asymp{(\mu^T\Sigma\mu)}^2\leq\lambda_{\max}^2(\Sigma){(\mu^T\mu)}^2. \tag{20}
$$

and apply lemma 1, we get

$$
\begin{aligned}
\mathrm{E}(Z_i^T\Gamma^T\Gamma Z_j)^4 =&\mathrm{E}(Z_i^T\Gamma^T\Gamma Z_jZ_j^T\Gamma^T\Gamma Z_i)^2\\
=&\mathrm{E}\mathrm{E}\big((Z_i^T\Gamma^T\Gamma Z_jZ_j^T\Gamma^T\Gamma Z_i)^2|Z_j\big)\\
\asymp&\mathrm{E}{(Z_j^T\Gamma^T\Sigma\Gamma Z_j)}^2\\
\asymp&(\mathrm{tr}\Sigma^2)^2+\mathrm{tr}\Sigma^4\\
\leq&(\mathrm{tr}\Sigma^2)^2+\lambda_{\max}^2(\Sigma)\mathrm{tr}\Sigma^2.
\end{aligned}
\tag{21}
$$

By condition (32), we have

$$
\mathrm{E}{(X_i^TX_j)}^4=O\Big({\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)}^2\Big) \tag{22}
$$

Similarly, we have

$$
\begin{aligned}
&\mathrm{E}{(X_i^TX_j)}^2{(X_k^TX_i)}^2\\
=&\mathrm{E}\mathrm{E}\big({(X_i^TX_j)}^2{(X_k^TX_i)}^2|X_i\big)\\
=&\mathrm{E}{(X_i^T(\Sigma+\mu\mu^T)X_i)}^2\\
=&\mathrm{E}{(Z_i^T\Gamma^T(\Sigma+\mu\mu^T)\Gamma Z_i+2\mu^T(\Sigma+\mu\mu^T)\Gamma Z_i+\mu\Sigma\mu+(\mu^T\mu)^2)}^2\\
\leq&4\mathrm{E}(Z_i^T\Gamma^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2+16\mathrm{E}(\mu^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2+4(\mu\Sigma\mu)^2+4(\mu^T\mu)^4\\
\asymp&(\mathrm{tr}(\Gamma^T(\Sigma+\mu\mu^T)\Gamma))^2+\mathrm{tr}(\Gamma^T(\Sigma+\mu\mu^T)\Sigma(\Sigma+\mu\mu^T)\Gamma)\\
&+\mu^T(\Sigma+\mu\mu^T)\Sigma(\Sigma+\mu\mu^T)\mu+(\mu\Sigma\mu)^2+(\mu^T\mu)^4\\
=&\mathrm{tr}\Sigma^4+{(\mu^T\mu)}^4+{(\mathrm{tr}\Sigma^2)}^2+2\mathrm{tr}\Sigma^2(\mu^T\Sigma\mu)+3{(\mu^T\Sigma\mu)}^2\\
&+3\mu^T\Sigma^3\mu+2(\mu^T\Sigma^2\mu)(\mu^T\mu)+(\mu^T\Sigma\mu){(\mu^T\mu)}^2\\
=&O(1)({(\mathrm{tr}\Sigma^2)}^2+(\mu^T\mu)^4)\\
=&O\Big({\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)}^2\Big)
\end{aligned}
\tag{23}
$$

The last 2 equality follows by the fact $\mu^T\Sigma^i\mu \le \lambda^i_{\max}(\Sigma)\mu^T\mu$ $(i = 1, 2, \ldots)$, the condition (32) and the inequality $ab \le a^p/p + b^q/p$ for $1/p + 1/q = 1$ and $p, q > 0$.

In (18), 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. Hence

$$\begin{aligned}\mathrm{E}\Big(\sum_{j<i}(X_i^TX_j)^2\Big)^2 =&\frac{n(n-1)(n-2)(n-3)}{4}\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\\ &+O(1)(\frac{n(n-1)}{2}+n(n-1)(n-2))\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\end{aligned} \tag{24}$$

Hence $\mathrm{Var}(\sum_{j<i}(X_i^TX_j)^2) = o\Big(\big(\mathrm{E}(\sum_{j<i}(X_i^TX_j)^2)\big)^2\Big)$, which prove the lemma. □

Lemma 3. *Suppose* (32) *holds. Assume*

$$\mu^T\mu = o(\sqrt{\mathrm{tr}\Sigma^2}), \tag{25}$$

then

$$\sum_{j<k}\Big(\sum_{i:i>k}X_i^TX_jX_i^TX_k\Big)^2 = o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{26}$$

$$\sum_{j<k}(X_i^TX_j)^4 = o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{27}$$

$$\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2 = o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{28}$$

Proof.

$$\begin{aligned}&\mathrm{E}\sum_{j<k}\Big(\sum_{i:i>k}X_i^TX_jX_i^TX_k\Big)^2\\ =&\mathrm{E}\sum_{j<k}\Big(\sum_{i:i>k}(X_i^TX_j)^2(X_i^TX_k)^2+2\sum_{i_1,i_2:i_1>i_2>k}X_{i_1}^TX_jX_{i_1}^TX_kX_{i_2}^TX_jX_{i_2}^TX_k\Big).\end{aligned} \tag{29}$$

By (23), we have

$$\mathrm{E}\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2 = O(n^3)\big(\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2 \tag{30}$$

But

$$
\begin{aligned}
&\mathrm{E}\sum_{j<k<i_2<i_1} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \\
=&\frac{n(n-1)(n-2)(n-3)}{6}\mathrm{tr}(\Sigma+\mu\mu^T)^4 \\
\leq&\frac{n(n-1)(n-2)(n-3)}{6}8(\mathrm{tr}(\Sigma)^4+(\mu^T\mu)^4) \\
\leq&O(n^4)(\lambda_{\max}^2(\Sigma)\mathrm{tr}(\Sigma)^2+(\mu^T\mu)^4) \\
=&o\Big(n^4\big(\mathrm{tr}\Sigma^2\big)^2\Big).
\end{aligned}
\tag{31}
$$

It follows that (26) holds. (27) and (28) can be proved similarly by (23) and (22).

□

As Janssen's paper, let d denote any metric on the set of probability measures $\mathcal{M}_1(\mathbb{R})$ on $\mathbb{R}$ such that convergence in $(\mathcal{M}_1(\mathbb{R}), d)$ is equivalent to weak convergence.

Theorem 1. *Suppose*

$$\frac{\lambda_{\max}^2(\Sigma)}{\mathrm{tr}\Sigma^2}\to 0, \tag{32}$$

and

$$\mu^T\mu = o(\sqrt{\mathrm{tr}\Sigma^2}), \tag{33}$$

then

$$d\Big(\mathcal{L}\Big(\frac{T_2(\epsilon_1 X_1,\dots,\epsilon_i X_i,\dots,\epsilon_n X_n)}{\sqrt{\sum_{1\leq j<i\leq n}\left(X_i^T X_j\right)^2}}\Big| X_1,\dots,X_n\Big), N(0,1)\Big)\xrightarrow{P} 0. \tag{34}$$

Proof. Similar to the proof of Chen's theorem and by Lemma 2, we have

$$\frac{T_2(X_1,\dots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\leq j<i\leq n}(X_i^T X_j)^2}}\xrightarrow{\mathcal{L}} N(0,1). \tag{35}$$

For every subsequence of $\{n\}$, there is a further subsequence $\{n(k)\}$ along which (26), (27) and (28) holds almost surely. By Proposition 1, we have

$$\mathcal{L}\Big(\frac{T_2(\epsilon_1 X_1,\dots,\epsilon_i X_i,\dots,\epsilon_n X_n)}{\sqrt{\sum_{1\leq j<i\leq n}\left(X_i^T X_j\right)^2}}\Big| X_1,\dots,X_n\Big)\xrightarrow{\mathcal{L}} N(0,1) \tag{36}$$

almost surely. Denote by ξ_α^* the $1-\alpha$ quantile of the above conditional distribution, then along $\{n(m)\}$ we have $\xi_\alpha^*\to\Phi(1-\alpha)$ almost surely, where $\Phi(\cdot)$

is the CDF of standard normal distribution. Then $\xi_\alpha^* \xrightarrow{P} \Phi(1-\alpha)$. Now the asymptotic power function can be derived by Slutsky's theorem

$$
\begin{aligned}
&\Pr\Big(\frac{T_2(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^T X_j\right)^2}} > \xi_\alpha^*\Big)\\
=&\Pr\Big(\frac{T_2(X_1,\ldots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^T X_j\right)^2}} > \xi_\alpha^* - \frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^T X_j\right)^2}}\Big)\\
=&\Pr\Big(\frac{T_2(X_1,\ldots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}} - \frac{\sqrt{\sum_{1\le j<i\le n}\left(X_i^T X_j\right)^2}}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}}\xi_\alpha^* > -\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}}\Big)\\
=&\Pr\Big(N(0,1)-\Phi(1-\alpha) > -\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}}\Big)+o(1)\\
=&\Phi(-\Phi(1-\alpha)+\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\mathrm{tr}\Sigma^2}})+o(1),
\end{aligned}
\tag{37}
$$

where the last two equality holds because an exersize of durrett.

□

4. Algorithm

The quantile might not be computationally feasible, since the randomized test statistic is (conditionally) uniformly distributed on 2^n different values. In practice, randomization test is often realized through an approximation of p-value. More specifically, first choose a large integer M. We can sample from conditional distribution of randomized statistic by generate $\epsilon_1,\ldots,\epsilon_n$ and compute $T(\epsilon_1 X_1,\ldots,\epsilon_n X_n)$. Repeat M times and we obtain T_i^*, $i=1,\ldots,M$. Denote $\xi_i = \mathbf{1}_{\{T_i^*\ge T_0\}}$. Then $\sum_{i=1}^M \xi_i/M$ is an approximation of the p-value

$$p(X_1,\ldots,X_n) = \Pr(T_i^* \ge T_0 | X_1,\ldots,X_n).$$

Hence we reject the null if $\sum_{i=1}^M \xi_i/M \le \alpha$.

Sometimes we can stop and draw a conclusion before sampling M times. In fact, we can accept the null once the sum of ξ_i exceeds $M\alpha$ and reject the null once the sum of $1-\xi_i$ reaches $M(1-\alpha)$.

Note that once we have obtained $X_i^T X_j$, the computation of T_i^* has only complexity $O(n^2)$. The total complexity is $O(n^2(p+M))$. For example, when $M = p$, time spent by randomization is at most twice of that of asymptotic method.

Algorithm 1 Randomization Algorithm

Require: α, M

Set $A \leftarrow 0$, $T_0 \leftarrow T(X_1, \ldots, X_n)$.
Compute $X_i^T X_j$ for $1 \leq j < i \leq n$
for $i = 1$ to M **do**
  Generate $\epsilon_1, \ldots, \epsilon_n$ and compute $T(\epsilon_1 X_1, \ldots, \epsilon_n X_n) = \sum_{j<i} X_i^T X_j \epsilon_i \epsilon_j$.
  if $T(\epsilon_1 X_1, \ldots, \epsilon_n X_n) > T_0$ **then**
    $A \leftarrow A + 1$
  end if
  if $A > M\alpha$ **then**
    return Accept
  end if
  if $i - A \geq M(1-\alpha)$ **then**
    return Reject
  end if
end for

References