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. 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-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:

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

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

[☆]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.

∞ 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 $\sum_{l=1}^{q} \alpha_l \le 8$ 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 $T(X_1, \ldots, X_n)$. Suppose $\epsilon_1, \ldots, \epsilon_n$ 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

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

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

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

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

We will study the quadratic form of Rademacher variables. Let $\epsilon_1, \ldots, \epsilon_n$ 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*

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

2

is that

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

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 $W_n = \sum_{i=2}^{n} U_{in}$ 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):

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

and

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

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

$$\sum_{i=2}^{n} \mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n}) = \sum_{i=2}^{n}\Big(\sum_{j=1}^{i-1} a_{ij}\epsilon_j\Big)^2$$

$$= \sum_{i=2}^{n}\Big(\sum_{j=1}^{i-1} a_{ij}^2 + 2\sum_{j,k:j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k\Big)$$

$$= \sum_{i=2}^{n}\sum_{j=1}^{i-1} a_{ij}^2 + 2\sum_{j<k<i} a_{ij}a_{ik}\epsilon_j\epsilon_k.$$

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 (9) holds.

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

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

3

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

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

where the last equality holds by assumption. Hence (10) holds. $\square$

References

4