Limiting behavior of largest entry of random tensor constructed by high-dimensional data

Tiefeng Jiang[1] and Junshan Xie[2]
University of Minnesota and Henan University

Abstract

Let $X_k = (x_{k1}, \cdots, x_{kp})', k = 1, \cdots, n$, be a random sample of size n coming from a p-dimensional population. For fixed integer $m \geq 2$, consider a hypercubic random tensor $\mathbf{T}$ of m-th order and rank n with

$$\mathbf{T} = \sum_{k=1}^{n} \underbrace{X_k \otimes \cdots \otimes X_k}_{m \ multiple} = \Big(\sum_{k=1}^{n} x_{ki_1} x_{ki_2} \cdots x_{ki_m} \Big)_{1 \leq i_1, \cdots, i_m \leq p}.$$

Let W_n be the largest off-diagonal entry of $\mathbf{T}$. We derive the asymptotic distribution of W_n under a suitable normalization for two cases. They are the ultra-high dimension case with $p \to \infty$ and $\log p = o(n^{\beta})$ and the high-dimension case with $p \to \infty$ and $p = O(n^{\alpha})$. The normalizing constant of W_n depends on m and the liming distribution of W_n is a Gumbel-type distribution involved with parameter m.

Keywords: Tensor; extreme-value distribution; high-dimensional data; Stein-Chen Poisson approximation method.
AMS 2000 Subject Classification: 60F05, 60F15, 62H10.

[1]School of Statistics, University of Minnesota, 224 Church Street, S. E., MN 55455, USA, jiang040@umn.edu. The research of Tiefeng Jiang was supported in part by NSF Grant DMS-1406279 and DMS-1916014.

[2]School of Mathematics and Statistics, Henan University, Kaifeng, 475000, China, xjsnwpu@163.com. The research of Junshan Xie was supported in part by NNSF 11401169.

1 Introduction

In this paper we will study the limiting properties of a hypercubic random tensor constructed by high-dimensional random data. Let $p \geq 2$ be an integer and $\mathbf{X} \in \mathbb{R}^p$ be a random vector. The distribution of $\mathbf{X}$ serves as a population distribution. Let $\mathbf{X}_k = (x_{k1}, \cdots, x_{kp})', 1 \leq k \leq n$, be a random sample of size n from the population distribution generated by $\mathbf{X}$, that is, $\mathbf{X}, \mathbf{X}_1, \cdots, \mathbf{X}_n$ are independent random vectors with a common distribution. The data matrix $\mathbf{X} = (x_{ki})_{1\leq k\leq n, 1\leq i\leq p}$ produces a hypercubic random tensor $\mathbf{T} \in \mathbb{R}^{p\times\cdots\times p}$ with order m and rank n defined by

$$\mathbf{T} = \sum_{k=1}^{n} \underbrace{\mathbf{X}_k \otimes \cdots \otimes \mathbf{X}_k}_{m\ multiple} = \Big(\sum_{k=1}^{n} x_{ki_1} x_{ki_2} \cdots x_{ki_m}\Big)_{1\leq i_1,\cdots,i_m\leq p}. \qquad (1.1)\ \{?\}$$

Researchers obtain some limiting properties of tensor data defined in (1.1). By using similar techniques to those in the random matrix theory, Ambainis and Harrow (2012) obtain a limiting property of the largest eigenvalue and the limiting spectral distribution of random tensors. Tieplova (2016) studies the limiting spectral distribution of the sample covariance matrices constructed by the random tensor data. Lytova (2017) further considers the central limit theorem for linear spectral statistics of the sample covariance matrices constructed by the random tensor data. Shi *et al.* (2018) apply limiting properties of the random tensors to a anomaly detection problem in the distribution networks.

In this paper, we will study the behavior of the largest off-diagonal entry of the random tensor $\mathbf{T}$ when both n and p tend to infinity. Precisely, we will work on the asymptotic distribution of

$$W_n := \max_{1\leq i_1<\cdots<i_m\leq p} \frac{1}{\sqrt{n}} \Big| \sum_{k=1}^{n} x_{ki_1} x_{ki_2} \cdots x_{ki_m} \Big| \qquad (1.2)\ \{?\}$$

as $n \to \infty$ and $p \to \infty$.

For fixed $m \geq 2$, when the entries of the data matrix $\mathbf{X} = (x_{ki})_{1\leq k\leq n, 1\leq i\leq p}$ are i.i.d. random variables, we will show that the limiting distribution of W_n with a suitable normalization is the Gumbel-type distribution involved with parameter m. Two typical high-dimensional cases are considered: the ultra-high dimension with $p \to \infty$ and $\log p = o(n^{\beta})$ and the high-dimension with $p \to \infty$ and $p = O(n^{\alpha})$. In both cases we obtain the limiting distributions of W_n, which is different from the case that $m = 2$.

When $m = 2$, the tensor $\mathbf{T} = \mathbf{X}'\mathbf{X}$ turns out to be the sample covariance matrix, which is a very popular statistic in the multivariate statistical analysis. The largest entry of the sample covariance matrix has been studied actively. In particular, assuming $n/p \to \gamma > 0$ and $E|x_{11}|^{30+\epsilon} < \infty$ for some $\epsilon > 0$, Jiang (2004) proves that

$$W_n^2 - 4\log p + \log\log p \xrightarrow{d} W_\infty \qquad (1.3)\ \{?\}$$

where random variable W_∞ has distribution function $F(z) = e^{-\frac{1}{\sqrt{8\pi}}e^{-z/2}}$, $z \in \mathbb{R}$. Here and later the notation "$\xrightarrow{d}$" means "converges in distribution to".

A sequence of results are then obtained to relax the moment condition that $E|x_{11}|^{30+\epsilon} < \infty$. For example, Zhou (2007) shows that (1.3) holds if

$$x^6 P(|x_{11}x_{12}| > x) \to 0. \tag{1.4}$$ {?}

Liu *et al.* (2008) proves that (1.3) holds provided a weaker condition is valid, that is,

$$n^3 P(|x_{11}x_{12}| > \sqrt{n \log n}) \to 0. \tag{1.5}$$ {?}

Besides the above two results, Li and Rosalsky (2006) and Li *et al.* (2010, 2012) further study the moment condition for which (1.3) is true. In a different direction, Liu *et al.* (2008) obtains (1.3) for the polynomial rate such that $p = O(n^\alpha)$; Cai and Jiang (2011) derive (1.3) for the ultra-high dimensional case with $\log p = o(n^\alpha)$ for some $\alpha > 0$. For the compressed sensing problems and testing problems related to W_n, one is referred to the papers by, for instance, Cai and Jiang (2011), Cai *et al.* (2013), Xiao and Wu (2013) and Shao and Zhou (2014).

In this paper, for all $m \geq 2$ we study W_n from (1.2). We prove that W_n with a suitable normalization converges to the Gumbel-type distribution. The normalizing constant and the liming distribution all depend on m. These results will be stated in the next section and discussions will be made afterwards.

Throughout the paper, the symbols $\xrightarrow{p}$ and $\xrightarrow{d}$ means convergence in probability and convergence in distribution, respectively. We will also denote $b_n = o(a_n)$ if $\lim_{n\to\infty} b_n/a_n = 0$; the notation $b_n = O(a_n)$ stands for that $\{|b_n/a_n|;\, n \geq 1\}$ is a bounded sequence; $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.

The rest of the paper is organized as follows. The main results of the paper as well as discussions are stated in Section 2. In Section 3, we will first present and prove some technical lemmas, and then prove the main results.

2 Main Results

We assume that p depends on n and simply write p for brevity of notation. In case there is a possible confusion, we will write $p = p_n$. Review $\mathbf{X} = (x_{ki})_{1\leq k\leq n, 1\leq i\leq p}$ and assume

$$\begin{aligned}&\{x_{ki};\, 1 \leq k \leq n, 1 \leq i \leq p\} \text{ are i.i.d. random variables with}\\ &Ex_{11} = 0 \text{ and } Ex_{11}^2 = 1.\end{aligned} \tag{2.1}$$ {?}

The quantity W_n is defined as in (1.2) with $m \geq 2$. In the following theorems the limiting distribution is the Gumbel distribution with distribution function

$$F_\theta(z) = \exp\Big\{-\frac{1}{m!\sqrt{m\pi}}e^{-z/2}\Big\}, \quad z \in \mathbb{R}. \tag{2.2}$$ {?}

THEOREM 1 *Assume (2.1) with* $Ee^{t_0|x_{11}|^\alpha} < \infty$ *for some* $0 < \alpha \le 1$ *and* $t_0 > 0$*. Let* $m \ge 2$ *be a fixed integer and* $\beta = \frac{\alpha}{2m-\alpha}$*. If* $p \to \infty$ *and* $\log p = o(n^\beta)$ *as* $n \to \infty$*, then* $W_n^2 - 2m\log p + \log\log p \xrightarrow{d} \theta$*, where* θ *has distribution function* $F_\theta(z)$ *as in (2.2).*

The above theorem studies the ultra-high dimensional case, that is, the dimension p can be at an exponential order of the size n, and the assumption of $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $0 < \alpha \le 1$ and $t_0 > 0$ is needed to derive the limiting distribution. Next we will consider a popular high-dimensional case in the literature such that $p = p_n$ is no larger than a polynomial rate of n. We then get the same limiting distribution of W_n under a much weaker moment condition.

THEOREM 2 *Let* $\alpha > 0$ *and* $m \ge 2$ *be constants such that* $E\big[|x_{11}|^{\tau_1}\log^{\tau_2}(1+|x_{11}|)\big] < \infty$ *with* $\tau_1 = 4m\alpha + 2$ *and* $\tau_2 = 2m\alpha + \frac{3}{2}$*. If* $p \to \infty$ *and* $p = O(n^\alpha)$*, then* $W_n^2 - 2m\log p + \log\log p \xrightarrow{d} \theta$*, where* θ *has distribution function* $F_\theta(z)$ *as in (2.2).*

By the Slutsky lemma, the above theorems imply the following.

COROLLARY 1 *Assume the conditions from either Theorem 1 or Theorem 2 holds. Then,*

$$\frac{W_n}{\sqrt{\log p}} \xrightarrow{p} \sqrt{2m}.$$

As discussed earlier, the largest entry of a sample covariance matrix have been studied with the limiting distribution stated in (1.3). In this paper we study the same problem for m-order random tensors, in which the setting is a more general. We find that the normalizing constant of W_n^2 is $2m\log p - \log\log p$ and the corresponding limiting distribution is given in (2.2). Both quantities indeed depend on m. We now make some further comments below.

1. Take $m = 2$, both Theorems 1 and 2 say

$$W_n^2 - 4\log p + \log\log p \xrightarrow{d} \theta \sim F_\theta(z) = e^{-\frac{1}{\sqrt{8\pi}}e^{-z/2}},$$

which is consistent with (1.3).

2. Now, instead of studying W_n from (1.2), we consider

$$\tilde{W}_n := \max_{1\le i_1<\cdots<i_m\le p} \frac{1}{\sqrt{n}}\sum_{k=1}^n x_{ki_1}x_{ki_2}\cdots x_{ki_m}. \tag{2.3}$$ {?}

Then, by using the same proofs except changing "$|N(0,1)|$" to "$N(0,1)$" in (3.26) and (3.39), Theorems 1 and 2 still hold with the limiting distribution "$F_\theta(z)$" from (2.2) is replaced by "$F(z)$", where

$$F(z) = \exp\Big\{-\frac{1}{2m!\sqrt{m\pi}}e^{-z/2}\Big\}, \quad z \in \mathbb{R}.$$

Corollary 1 still holds without change if "W_n" is replaced by "$\tilde{W}_n$".

3. Recently Fan and Jiang (2018) study the liming behavior of $\tilde{W}_n$ from (2.3) with $m = 2$ and with $(x_{11}, \cdots, x_{1p})' \sim N(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma}_{ii} = 1$ for each i and $\mathbf{\Sigma}_{ij} \equiv \rho > 0$ for all $i \neq j$. The limiting distribution of $\tilde{W}_n$ is the Gumbel distribution if ρ is very small; that is Gaussian if ρ is large; that is the convolution of the Gumbel and the Gaussian distributions if ρ is in between. The proof is very involved. Such setting can also be extended to $\tilde{W}_n$ from (2.3) for any $m \geq 3$ with a lengthy argument. We leave it as a future work.

4. Assume that $m = 2$ and that $(x_{11}, \cdots, x_{1p})' \sim N(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a banded matrix. Cai and Jiang (2011) study W_n from (1.2) and apply their results to compressed sensing problems and tests of covariance structures. It will be interesting to see if similar dependent structures can be carried out for W_n with $m \geq 3$.

5. The proofs of Theorems 1 and 2 rely on the Chen-Stein Poisson approximation method and the moderate deviations. The major technicality comes from computing λ and bounding b_2 appeared in Lemma 3.1. The major difference between our proofs here and those in the literature is that the evaluation of λ is more involved. Furthermore, we need a significant effort to investigate b_2. Due to the assumption $m \geq 3$ the dependent structure appearing in b_2 becomes more subtle; see Lemmas 3.4 and 3.6 for details.

6. Taking $m = 2$ and $\alpha = 1$ in Theorem 2, the required moment condition in the theorem becomes $E\big[|x_{11}|^{10} \log^{5.5}(1 + |x_{11}|)\big] < \infty$. This is stronger than (1.4) and (1.5). In fact it is Lemma 3.6 that requires the above condition. It is possible that the moment assumption in Theorem 2 can be relaxed. We leave it as a future project.

7. In the paper, the random tensor $\mathbf{T}$ is constructed by the sample of a single multivariate population. In fact, the results of Theorems 1 and 2 can also be extended to the tensor constructed by the samples of several populations with the same dimension p. Let $X^{(l)} \in \mathbb{R}^p$; $l = 1, 2, \cdots, m$ be m random vectors, and the p entries of $X^{(l)}$ be i.i.d. random variables for each l. The probability distribution of each vector generates a population distribution. For each $1 \leq l \leq m$, let $(x_{k1}^{(l)}, \cdots, x_{kp}^{(l)})'$, $k = 1, \cdots, n$, be a random sample of size n from the population $X^{(l)}$. We then have a data matrix $\mathbf{X}^{(l)} = (x_{ki}^{(l)})_{1 \leq k \leq n, 1 \leq i \leq p}$ and we define a special hypercubic random tensor $\mathbf{T}' \in \mathbb{R}^{p \times \cdots \times p}$ with order m and rank n by

$$\mathbf{T}' = \Big(\sum_{k=1}^{n} x_{ki_1}^{(1)} x_{ki_2}^{(2)} \cdots x_{ki_m}^{(m)} \Big)_{1 \leq i_1, \cdots, i_m \leq p}.$$

Denote the largest element of $\mathbf{T}'$ by

$$W_n' = \max_{1 \leq i_1 < \cdots < i_m \leq p} \frac{1}{\sqrt{n}} \Big| \sum_{k=1}^{n} x_{ki_1}^{(1)} x_{ki_2}^{(2)} \cdots x_{ki_m}^{(m)} \Big|.$$

By the same argument as those in the proofs of Theorems 1 and 2, the two theorems still hold if "W_n" is replaced by "W_n'" and some uniform moment conditions on $x_{11}^{(l)}$, $1 \leq l \leq m$ are assumed.

3 Proofs

3.1 Some technical lemmas

We will start with listing some technical lemmas in our proofs. The first one is a classical Stein-Chen Poisson approximation lemma, which is frequently used in studying behaviors of maximum of almost mutual independent random variables. The following result is a special case of Theorem 1 of Arratia *et al.* (1989).

Lemma 3.1 *Let $\{\eta_\alpha, \alpha \in I\}$ be random variables on an index set I and $\{B_\alpha, \alpha \in I\}$ be a set of subsets of I, that is, for each $\alpha \in I$, $B_\alpha \subset I$. For any $t \in \mathbb{R}$, set $\lambda = \sum_{\alpha \in I} P(\eta_\alpha > t)$, Then we have*

$$\Big| P(\max_{\alpha \in I} \eta_\alpha \le t) - e^{-\lambda} \Big| \le (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3),$$

where

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t) P(\eta_\beta > t), \\ b_2 &= \sum_{\alpha \in I} \sum_{\alpha \ne \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t), \\ b_3 &= \sum_{\alpha \in I} \Big| p(\eta_\alpha > t | \sigma(\eta_\beta, \beta \notin B_\alpha)) - P(\eta_\alpha > t) \Big|, \end{aligned}$$

and $\sigma(\eta_\beta, \beta \notin B_\alpha)$ is the σ-algebra generated by $\{\eta_\beta, \beta \notin B_\alpha\}$. In particular, if η_α is independent of $\{\eta_\beta, \beta \notin B_\alpha\}$ for each α, then b_3 vanishes.

The following conclusion is about the moderate deviation of the partial sum of i.i.d. random variables. It can be seen from Linnik (1961).

Lemma 3.2 *Suppose $\{\zeta, \zeta_1, \zeta_2, \cdots\}$ is a sequence of i.i.d. random variables with zero mean and $E\zeta_i^2 = 1$. Define $S_n = \sum_{i=1}^n \zeta_i$.*

(1) If $Ee^{t_0|\zeta|^\alpha} < \infty$ for some $0 < \alpha \le 1$ and $t_0 > 0$, then

$$\lim_{n \to \infty} \frac{1}{x_n^2} \log P\Big(\frac{S_n}{\sqrt{n}} \ge x_n \Big) = -\frac{1}{2}$$

for any $x_n \to \infty, x_n = o(n^{\frac{\alpha}{2(2-\alpha)}})$.

(2) If $Ee^{t_0|\zeta|^\alpha} < \infty$ for some $0 < \alpha \le \frac{1}{2}$ and $t_0 > 0$, then

$$\frac{P\big(\frac{S_n}{\sqrt{n}} \ge x\big)}{1 - \Phi(x)} \to 1$$

holds uniformly for $0 \le x \le o(n^{\frac{\alpha}{2(2-\alpha)}})$.

Let $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$ for $x \in \mathbb{R}$. The following result is Proposition 4.5 from Chen *et al.* (2013).

Lemma 3.3 *Let η_i, $1 \le i \le n$, be independent random variables with $E\eta_i = 0$ and $Ee^{h_n|\eta_i|} < \infty$ for some $h_n > 0$ and $1 \le i \le n$. Assume that $\sum_{i=1}^n E\eta_i^2 = 1$. Then*

$$\frac{P(\sum_{i=1}^n \eta_i \ge x)}{1-\Phi(x)} = 1 + C_n(1+x^3)\gamma e^{4x^3\gamma}$$

for all $0 \le x \le h_n$ and $\gamma = \sum_{i=1}^n E\big(|\eta_i|^3 e^{x|\eta_i|}\big)$, where $\sup_{n\ge1}|C_n| \le C$ and C is an absolute constant.

Proposition 1 *Let $\{\xi_i; i \ge 1\}$ be i.i.d. random variables with $E\xi_1 = 0$, $E(\xi_1^2) = 1$ and $E(|\xi_1|^r) < \infty$ for some $r > 2$. Let $\{c_n > 0;\, n \ge 1\}$ be constants with $\sup_{n\ge1} c_n < \infty$. Assume*

$$P\big(|\xi_1| > \sqrt{n/\log n}\,\big) = o\Big(\frac{1}{n^{1+(c_n^2/2)}\sqrt{\log n}}\Big) \tag{3.1 {?}}$$

as $n \to \infty$. Then, $P(S_n \ge c_n\sqrt{n\log n}\,) \sim 1 - \Phi(c_n\sqrt{\log n}\,)$.

Amosova (1972) derives a similar result to Proposition 1 for independent but not necessarily identically distributed random variables. If $\{\xi_i; i \ge 1\}$ are i.i.d. random variables and $\lim_{n\to\infty} c_n = c$, then Amosova concludes that $P(S_n \ge c_n\sqrt{n\log n}\,) \sim 1 - \Phi(c_n\sqrt{\log n}\,)$ under the condition $E(|\xi_1|^{c^2+2+\epsilon}) < \infty$ for some $\epsilon > 0$. This moment condition implies (3.1) by the Markov inequality and hence our proposition holds. In particular, taking $c_n \equiv c > 0$, then (3.1) holds if $E\big[|\xi_1|^{c^2+2}\log^{(c^2+3)/2}(1+|\xi_1|)\big] < \infty$. In conclusion, for the i.i.d. case Proposition 1 relaxes the condition required by Amosova.

Proof of Proposition 1. By the standard central limit theorem, as $n \to \infty$,

$$\sup_{a\le x\le b}\Big|\frac{P(S_n \ge \sqrt{n}x)}{1-\Phi(x)} - 1\Big| \to 0$$

for any real numbers $b > a$. So, without loss of generality, we will prove the conclusion under the extra assumption

$$c_n\sqrt{\log n} \to \infty \tag{3.2 {?}}$$

as $n \to \infty$. The proof is divided into a few steps.

Step 1: truncation. Define $a_1 = 1$ and $a_n = \sqrt{n/\log n}$ for $n \ge 2$. Denote $K = E(|\xi_1|^r)$. Set

$$\begin{aligned}
\xi_i' &= \xi_i I(|\xi_i| \le a_n) - E\big[\xi_i I(|\xi_i| \le a_n)\big];\\
\xi_i'' &= \xi_i I(|\xi_i| > a_n) - E\big[\xi_i I(|\xi_i| > a_n)\big]
\end{aligned}$$

for $1 \leq i \leq n$. Trivially, $\xi_i = \xi_i' + \xi_i''$, $|\xi_i'| \leq 2a_n$ and

$$\big|E\big[\xi_i I(|\xi_i| > a_n)\big]\big| \leq \frac{K}{a_n^{r-1}}.$$

Furthermore,

$$\mathrm{Var}(\xi_i'') \leq E[\xi_i^2 I(|\xi_i| > a_n)] \leq \frac{K}{a_n^{r-2}}. \qquad (3.3)\ \{?\}$$

Now

$$\mathrm{Var}(\xi_i) = \mathrm{Var}(\xi_i') + \mathrm{Var}(\xi_i'') + 2\mathrm{Cov}(\xi_i', \xi_i''). \qquad (3.4)\ \{?\}$$

Use the formula that $\mathrm{Cov}(U - EU, V - EV) = E(UV) - (EU)EV$ for any random variables U and V to see

$$\begin{aligned} \mathrm{Cov}(\xi_i', \xi_i'') &= -E\big[\xi_i I(|\xi_i| \leq a_n)\big] \cdot E\big[\xi_i I(|\xi_i| > a_n)\big] \\ &= \big(E\big[\xi_i I(|\xi_i| > a_n)\big]\big)^2 \\ &\leq E\big[\xi_i^2 I(|\xi_i| > a_n)\big] \\ &\leq \frac{K}{a_n^{r-2}} \end{aligned}$$

by the assumption $E\xi_i = 0$, the Cauchy-Schwartz inequality and (3.3). This together with (3.3) and (3.4) implies that

$$\mathrm{Var}(\xi_1) \geq \mathrm{Var}(\xi_i') \geq \mathrm{Var}(\xi_1) - \frac{3K}{a_n^{r-2}}. \qquad (3.5)\ \{?\}$$

Set $S_n = \sum_{i=1}^n \xi_i$, $S_n' = \sum_{i=1}^n \xi_i'$ and $S_n'' = \sum_{i=1}^n \xi_i''$. Then $S_n = S_n' + S_n''$. Thus,

$$P(S_n > u) \leq P(S_n' > u - v) + P(|S_n''| > v). \qquad (3.6)\ \{?\}$$

for any $u > v > 0$. Moreover, $S_n' \leq S_n + |S_n''|$, we see

$$P(S_n' > u + v) \leq P(S_n > u) + P(|S_n''| > v)$$

for any $u > 0$ and $v > 0$. This leads to

$$P(S_n > u) \geq P(S_n' > u + v) - P(|S_n''| > v). \qquad (3.7)\ \{?\}$$

From the definition of ξ_i', it is easy to see that $\sup_{n\geq 1} E|\xi_1'|^r \leq 2^r K$. Note that

$$nE\big[|\xi_1| I(|\xi_1| > a_n)\big] \leq nE(|\xi_1|^r) \cdot \frac{1}{a_n^{r-1}} = \frac{nK}{a_n^{r-1}}.$$

Hence

$$\begin{aligned} P(|S_n''| > v) &\leq P\Big(|\sum_{i=1}^n \xi_i I(|\xi_i| > a_n)| > v - \frac{nK}{a_n^{r-1}}\Big) \\ &\leq nP(|\xi_1| > a_n) \end{aligned} \qquad (3.8)\ \{?\}$$

provided $v > \frac{nK}{a_n^{r-1}}$.

Step 2: the tail for S_n'. Set $\sigma'^2 = \mathrm{Var}(\xi_1')$. Trivially, $\sigma' \to 1$ as $n \to \infty$. Take $\eta_i = \frac{\xi_i'}{\sqrt{n}\sigma'}$. Then $|\eta_i| \le \frac{2}{\sigma'\sqrt{\log n}}$. Therefore we see from Lemma 3.3 that

$$\frac{P(S_n' \ge \sqrt{n}\sigma' x)}{1-\Phi(x)} = 1 + C_n(1+x^3)\gamma e^{4x^3\gamma} \qquad (3.9)\ \{?\}$$

where $\sup_{n\ge1}|C_n| \le C$ and C is an absolute constant, and

$$\begin{aligned}\gamma &\le \frac{n}{n^{3/2}\sigma'^3}\cdot E\big(|\xi_1'|^3 e^{2a_n x/(\sqrt{n}\sigma')}\big)\\ &\le \frac{1}{\sqrt{n}\sigma'^3}\cdot e^{2a_n x/(\sqrt{n}\sigma')}\cdot E(|\xi_1'|^3)\end{aligned}$$

Use the fact $\sup_{n\ge1} E|\xi_1'|^r \le 2^r K$ to see that $\sup_{n\ge1} E(|\xi_1'|^3) \le 2^r K$ if $r \ge 3$ by the Hölder inequality. If $2 < r < 3$, then write

$$\begin{aligned}E(|\xi_1'|^3) &= E(|\xi_1'|^r\cdot|\xi_1'|^{3-r})\\ &\le 2^{3-r}\Big(\frac{n}{\log n}\Big)^{(3-r)/2}\cdot E(|\xi_1'|^r)\\ &\le 8K\cdot\Big(\frac{n}{\log n}\Big)^{(3-r)/2}\end{aligned}$$

by the facts that $|\xi_1'| \le 2a_n = 2\sqrt{n/\log n}$ and that $\sup_{n\ge1} E|\xi_1'|^r \le 2^r K$. In summary, if $x = O(\sqrt{n}/a_n)$ then

$$\gamma \le \begin{cases} O\Big(\frac{(\log n)^{(r-3)/2}}{n^{(r/2)-1}}\Big), & \text{if } 2<r<3;\\ O(n^{-1/2}), & \text{if } r \ge 3\end{cases}$$

as $n \to \infty$. In particular, noting $a_n = \sqrt{n/\log n}$, we know that $\gamma \to 0$ and $x^3\gamma \to 0$ since $x = O(\sqrt{n}/a_n)$. Consequently, we have from (3.9) that

$$P(S_n' \ge \sqrt{n}\sigma' x) \sim 1 - \Phi(x) \qquad (3.10)\ \{?\}$$

under the assumption $x = O(\sqrt{n}/a_n)$.

Step 3: the tail for S_n. Take $u = c_n\sqrt{n\log n}$ and $v = \frac{2nK}{a_n^{r-1}}$. Then $v/u \to 0$ as $n \to \infty$. We still write u and v next sometimes for short notation. By (3.6), (3.7) and (3.8),

$$\begin{aligned}&P(S_n > c_n\sqrt{n\log n})\\ \le\ &P(S_n' > u - v) + nP(|\xi_1| > \sqrt{n/\log n})\end{aligned} \qquad (3.11)\ \{?\}$$

and

$$\begin{aligned}&P(S_n > c_n\sqrt{n\log n})\\ \ge\ &P(S_n' > u + v) - nP(|\xi_1| > \sqrt{n/\log n})\end{aligned} \qquad (3.12)\ \{?\}$$

In what follows, we will show both $P(S'_n > u+v)$ and $P(S'_n > u-v)$ are close to $P(S'_n > u)$. Since the two arguments have no difference, we will consider them simultaneously and write $u \pm v$ for the case $u+v$ and $u-v$, respectively. Noticing $u \pm v \sim c_n\sqrt{n \log n}$. Take $x = (u \pm v)/(\sqrt{n}\sigma')$ in (3.10). Then $x = O(\sqrt{n}/a_n)$ by the assumption $\sup_{n\ge 1} c_n < \infty$. From (3.2), $x \sim c_n\sqrt{\log n} \to \infty$. It follows that

$$\begin{aligned} P(S'_n > u \pm v) &\sim P\Big(N(0,1) > \frac{u \pm v}{\sqrt{n}\sigma'}\Big) \\ &\sim \frac{1}{c_n\sqrt{2\pi \log n}} \exp\big\{ -\frac{(u\pm v)^2}{2n\sigma'^2}\big\} \end{aligned}$$

as $n \to \infty$, where we use the fact $P(N(0,1) > x) \sim \frac{1}{\sqrt{2\pi}\,x} e^{-x^2/2}$ as $x \to \infty$. We claim

$$\exp\Big\{ -\frac{(u\pm v)^2}{2n\sigma'^2}\Big\} \cdot \exp\Big\{\frac{c_n^2 \log n}{2}\Big\} \to 1 \qquad (3.13)\ \{?\}$$

as $n \to \infty$. In fact, write $(u \pm v)^2 = u^2 + v^2 \pm 2uv$. Then

$$\begin{aligned} & -\frac{(u\pm v)^2}{2n\sigma'^2} + \frac{c_n^2 \log n}{2} \\ = & -\frac{u^2}{2n\sigma'^2} + \frac{c_n^2 \log n}{2} - \frac{v^2 \pm 2uv}{2n\sigma'^2} \\ = & \frac{c_n^2(\log n)}{2} \cdot \frac{\sigma'^2 - 1}{\sigma'^2} + O\Big(\frac{uv}{n}\Big). \end{aligned}$$

The assertion (3.5) says that $\sigma'^2 \to 1$ and $0 \le 1 - \sigma'^2 \le \frac{3K}{a_n^{r-2}} = O\big((\frac{\log n}{n})^{(r/2)-1}\big)$. Also,

$$\frac{uv}{n} = O\Big(\frac{\sqrt{n \log n}}{a_n^{r-1}}\Big) = O\Big(\frac{(\log n)^{r/2}}{n^{(r/2)-1}}\Big).$$

It follows that

$$-\frac{(u\pm v)^2}{2n\sigma'^2} + \frac{c_n^2 \log n}{2} = O\Big(\frac{(\log n)^{r/2}}{n^{(r/2)-1}}\Big).$$

We then confirms (3.13). Therefore,

$$P(S'_n > u \pm v) \sim \frac{1}{c_n\sqrt{2\pi \log n}} \cdot \frac{1}{n^{c_n^2/2}}.$$

By the given condition,

$$nP(|\xi_1| > \sqrt{n/\log n}) = o\Big(\frac{1}{n^{c_n^2/2}\sqrt{\log n}}\Big).$$

Comparing these with (3.11) and (3.12), we arrive at

$$P(S_n > c_n\sqrt{n \log n}) \sim 1 - \Phi(c_n\sqrt{\log n})$$

as $n \to \infty$. □

3.2 Main Proofs

For $1 \le s \le m-1$, define

$$\xi_k^{(s)} = \prod_{t=1}^{s} x_{kt}, \ \eta_k^{(s)} = \prod_{t=s+1}^{m} x_{kt}, \ \zeta_k^{(s)} = \prod_{t=m+1}^{2m-s} x_{kt}. \tag{3.14 {?}}$$

For a number $a > 0$ and a sequence of positive numbers $\{a_n\}$ with $\lim_{n\to\infty} a_n = a$, we define

$$\Psi_n^{(s)}(a_n) = P\Big(\big|\sum_{k=1}^{n} \xi_k^{(s)} \eta_k^{(s)}\big| \ge a_n\sqrt{n \log p}, \ \big|\sum_{k=1}^{n} \xi_k^{(s)} \zeta_k^{(s)}\big| \ge a_n\sqrt{n \log p}\Big) \tag{3.15 {?}}$$

for any $1 \le s \le m-1$. The next is a result on $\Psi_n^{(s)}(a_n)$, which is a key step in the application of the Chen-Stein Poisson approximation to prove Theorem 1.

LEMMA 3.4 *Let $\{a_n; n \ge 1\}$ be a sequence of positive numbers with $\lim_{n\to\infty} a_n = a > 0$. Under the assumptions of Theorem 1, we have that $\max_{1\le s\le m-1} \Psi_n^{(s)}(a_n) = o(p^{-a^2+\epsilon})$ for any $\epsilon > 0$.*

Proof. Let u, v and $w > 0$ be three numbers. It is easy to check that either $|u+v| > 2w$ or $|u-v| > 2w$ if $|u| \ge w$ and $|v| \ge w$. It then follows from (3.15) that

$$\begin{aligned}
\Psi_n^{(s)}(a_n) &\le P\Big(\big|\sum_{k=1}^{n} \xi_k^{(s)}\big(\eta_k^{(s)} + \zeta_k^{(s)}\big)\big| \ge 2a_n\sqrt{n \log p}\Big) \\
&\quad + P\Big(\big|\sum_{k=1}^{n} \xi_k^{(s)}\big(\eta_k^{(s)} - \zeta_k^{(s)}\big)\big| \ge 2a_n\sqrt{n \log p}\Big) \\
&:= A_n + B_n.
\end{aligned} \tag{3.16 {?}}$$

For the term A_n, trivially,

$$E[\xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)})] = 0, \ E[\xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)})]^2 = 2. \tag{3.17 {?}}$$

It is elementary that

$$\prod_{t=1}^{m} |a_t|^{\alpha/m} \le \frac{1}{m}\sum_{t=1}^{m} |a_t|^{\alpha}$$

for all $a_t \ge 0\,(t = 1, \cdots, m)$. Thus, we get

$$\begin{aligned}
E\exp\Big\{t_0|\xi_1^{(s)}\eta_1^{(s)}|^{\alpha/m}\Big\} &= E\exp\Big\{t_0\prod_{t=1}^{m}|x_{1t}|^{\alpha/m}\Big\} \\
&\le E\exp\Big\{\frac{t_0}{m}\sum_{t=1}^{m}|x_{1t}|^{\alpha}\Big\} \\
&= \prod_{t=1}^{m} E\exp\Big\{\frac{t_0}{m}|x_{11}|^{\alpha}\Big\}.
\end{aligned}$$

By assumption, $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $0 < \alpha \le 1$ and $t_0 > 0$, we see that

$$E\exp\Big\{t_0|\xi_1^{(s)}\eta_1^{(s)}|^{\alpha/m}\Big\} < \infty. \tag{3.18 {?}}$$

Noticing $0 < \alpha/m < 1$, we have

$$\begin{aligned}
& E\exp\Big\{\frac{1}{2}t_0\big|\xi_1^{(s)}\big(\eta_1^{(s)}+\zeta_1^{(s)}\big)\big|^{\alpha/m}\Big\} \\
\le\ & E\Big[\exp\Big\{\frac{1}{2}t_0\big|\xi_1^{(s)}\eta_1^{(s)}|^{\alpha/m}\Big\}\cdot\exp\Big\{\frac{1}{2}t_0\big|\xi_1^{(s)}\zeta_1^{(s)}|^{\alpha/m}\Big\}\Big] \\
\le\ & \Big[E\exp\Big\{t_0\big|\xi_1^{(s)}\eta_1^{(s)}|^{\alpha/m}\Big\}\Big]^{1/2}\cdot\Big[E\exp\Big\{t_0\big|\xi_1^{(s)}\zeta_1^{(s)}|^{\alpha/m}\Big\}\Big]^{1/2} \\
<\ & \infty
\end{aligned}$$

by the Cauchy-Schwartz inequality. From the notation $\beta = \frac{\alpha}{2m-\alpha}$ in statement of Theorem 1, we see $\frac{1}{2}\cdot\frac{\alpha/m}{2-(\alpha/m)} = \frac{\beta}{2}$. An assumption implies that $a_n\sqrt{2\log p} = o(n^{\beta/2})$. It is easy to see that $\{\xi_k^{(s)}\big(\eta_k^{(s)}+\zeta_k^{(s)}\big);\ 1 \le k \le n\}$ are i.i.d. random variables. By Lemma 3.2 (1) and (3.17), we get that, for any sufficient small $\delta > 0$,

$$\begin{aligned}
A_n &\le P\Big(\Big|\frac{1}{\sqrt{2n}}\sum_{k=1}^n \xi_k^{(s)}\big(\eta_k^{(s)}+\zeta_k^{(s)}\big)\Big| \ge a_n\sqrt{2\log p}\Big) \\
&\le 2\exp\{-(1-\delta)a_n^2\log p\} \\
&= 2p^{(\delta-1)a_n^2}.
\end{aligned}$$

Since $a_n \to a$, the above implies that, for any $\epsilon > 0$, we have

$$A_n = o(p^{-a^2+\epsilon}) \tag{3.19 {?}}$$

as $n \to \infty$. Similarly,

$$B_n = o(p^{-a^2+\epsilon}). \tag{3.20 {?}}$$

Combining (3.16), (3.19) and (3.20), we complete the proof. □

Proof of Theorem 1. The asymptotic distribution of W_n will be derived by the Chen-Stein Poisson approximation method introduced in Lemma 3.1. To do so, set $\mathbb{Z}$ be the set of integers and $I = \{(i_1, \cdots, i_m) \in \mathbb{Z}^p : 1 \le i_1 < \cdots < i_m \le p\}$. For each $\alpha = (i_1, \cdots, i_m) \in I$, define

$$X_\alpha = \frac{1}{\sqrt{n}}\Big|\sum_{k=1}^n x_{ki_1}x_{ki_2}\cdots x_{ki_m}\Big| \tag{3.21 {?}}$$

and

$$B_\alpha = \big\{(j_1, \cdots, j_m) \in I;\ \{j_1, \cdots, j_m\} \cap \{i_1, \cdots, i_m\} \ne \emptyset \text{ but } (j_1, \cdots, j_m) \ne \alpha\big\}.$$

Obviously, X_α is independent of $\{X_\beta;\ \beta \in I\backslash X_\alpha\}$. It is easy to verify that

$$|I| = \binom{p}{m} \ \text{ and } \ |B_\alpha| \le m^2 p^{m-1} \tag{3.22 {?}}$$

for each $\alpha \in I$. For any $z \in \mathbb{R}$, write

$$\nu_p = \Big[\log p - \frac{1}{2m}\big(\log\log p + 2\log(m!\sqrt{m\pi}) - z\big)\Big]^{1/2}. \tag{3.23 {?}}$$

Notice v_p may not make sense for small values of p. Since $p = p_n \to \infty$ as $n \to \infty$, without loss of generality, assume $v_p > 0$ for all $n \ge 1$. Set $\alpha_0 = \{1, 2, \cdots, m\} \in I$. By Lemma 3.1,

$$\Big| P\big(\max_{\alpha \in I} X_\alpha \le \sqrt{2m}\nu_p\big) - e^{-\lambda_p}\Big| \le b_1 + b_2, \tag{3.24 {?}}$$

where b_1 and b_2 are as in Lemma 3.1 and

$$\begin{aligned} \lambda_p &= \binom{p}{m} P\big(X_{\alpha_0} > \sqrt{2m}\nu_p\big) \\ &\sim \frac{p^m}{m!} P\Big(\big|\sum_{k=1}^n x_{k1}x_{k2}\cdots x_{km}\big| > \sqrt{2mn}\nu_p\Big). \end{aligned} \tag{3.25 {?}}$$

First, write $\psi_k = x_{k1}x_{k2}\cdots x_{km}$, $1 \le k \le n$. Then $E\psi_k = 0$ and $E\psi_k^2 = 1$. The assertion (3.18) says that $Ee^{t_0|\psi_1|^{\alpha/m}} < \infty$. Note that $\frac{\alpha}{m} \le \frac{1}{2}$ since $0 < \alpha \le 1$ and $m \ge 2$. Moreover, $\sqrt{2m}\nu_p = O(\sqrt{\log p}) = o(n^{\beta/2})$. By the definition of β, we know $\frac{\beta}{2} = \frac{1}{2} \cdot \frac{\alpha/m}{2-(\alpha/m)}$. Therefore it follows from Lemma 3.2(2) that

$$\begin{aligned} P\Big(\big|\sum_{k=1}^n x_{k1}x_{k2}\cdots x_{km}\big| > \sqrt{2mn}\nu_p\Big) &\sim P\big(|N(0,1)| > \sqrt{2m}\nu_p\big) \\ &\sim \frac{2}{\sqrt{4\pi m}\, v_p} \cdot e^{-mv_p^2}, \end{aligned} \tag{3.26 {?}}$$

where the fact $P(N(0,1) > x) \sim \frac{1}{\sqrt{2\pi}\, x} \cdot e^{-x^2/2}$ as $x \to \infty$ is used. It is easy to check that

$$\begin{aligned} &v_p \sim \sqrt{\log p}; \\ &-mv_p^2 = -\log(p^m) + \frac{1}{2}\big(\log\log p + 2\log(m!\sqrt{m\pi})\big) - \frac{1}{2}z \end{aligned}$$

as $n \to \infty$. Therefore,

$$\lambda_p \sim \frac{p^m}{m!} \frac{2}{\sqrt{4m\pi \log p}} \cdot e^{-mv_p^2} = e^{-z/2}. \tag{3.27 {?}}$$

In particular, this implies that

$$P\big(X_\alpha > \sqrt{2m}\nu_p\big) \sim \frac{m!}{p^m} e^{-z/2} \tag{3.28 {?}}$$

as $n\to\infty$ for any $\alpha\in I$. Consequently, we have from (3.22) that

$$b_1 \;\; \leq \;\; |I|\cdot|B_\alpha|\cdot P\big(X_\alpha>\sqrt{2m}\nu_p\big)^2 = O\Big(\frac{1}{p}\Big). \tag{3.29}$$ {?}

Now we estimate b_2. First,

$$\begin{aligned} b_2 \;\; &= \;\; \sum_{\alpha\in I}\sum_{\beta\in B_\alpha} P\big(X_\alpha>\sqrt{2m}\nu_p,\, X_\beta>\sqrt{2m}\nu_p\big) \\ &= \;\; \sum_{\alpha\in I}\sum_{s=1}^{m-1}\sum_{\beta\in I:|\beta\cap\alpha|=s} P\big(X_\alpha>\sqrt{2m}\nu_p,\, X_\beta>\sqrt{2m}\nu_p\big). \end{aligned} \tag{3.30}$$ {?}

For $1\leq s\leq m-1$,

$$\begin{aligned} & \{\beta\in I: |\beta\cap\{1,2,\cdots,m\}|=s\} \\ = \;\; & \{(i_1,\cdots,i_m)\in\mathbb{Z}^p: 1\leq i_1<\cdots<i_m\leq p,\, |(i_1,\cdots,i_m)\cap\{1,2,\cdots,m\}|=s\}. \end{aligned}$$

Hence, $|\{\beta\in I: |\beta\cap\{1,2,\cdots,m\}|=s\}| = \binom{m}{s}\cdot\binom{p-m}{m-s}\leq m^s p^{m-s}$. Review the notation in (3.14) and (3.15). In particular,

$$\Psi_n^{(s)}(a_n) = P\Big(\big|\sum_{k=1}^n \xi_k^{(s)}\eta_k^{(s)}\big|\geq a_n\sqrt{n\log p},\, \big|\sum_{k=1}^n \xi_k^{(s)}\zeta_k^{(s)}\big|\geq a_n\sqrt{n\log p}\Big).$$

Since x_{ij}'s are i.i.d. random variables, we see that

$$b_2\leq |I|\cdot\sum_{s=1}^{m-1} m^s p^{m-s}\cdot\Psi_n^{(s)}(a_n) \tag{3.31}$$ {?}

where

$$a_n := \frac{\sqrt{2m}\nu_p}{\sqrt{\log p}} \to \sqrt{2m}$$

as $n\to\infty$. By Lemma 3.4, for any $\epsilon>0$, we have

$$\Psi_n^{(s)}(a_n)\leq p^{-2m+\epsilon}$$

as n is large enough. This implies that

$$\begin{aligned} b_2 \;\; &\leq \;\; p^m\cdot m^m p^{m-1}\cdot p^{-2m+\epsilon} \\ &= \;\; m^m p^{-1+\epsilon}\to 0 \end{aligned} \tag{3.32}$$ {?}

as $n\to\infty$ for all $\epsilon\in(0,1)$. Combining (3.24), (3.27), (3.29) and (3.32), we complete the proof. □

The following two lemmas are prepared for the proof of Theorem 2.

Lemma 3.5 *Let x_{ij}'s be as in Theorem 2 and ν_p be as in (3.23). Define $c_n = \sqrt{2m/\log n}\nu_p$ and $\xi_1 = x_{11}x_{12}\cdots x_{1m}$. Then*

$$n^{1+(c_n^2/2)}\sqrt{\log n}\cdot P\big(|\xi_1| > \sqrt{n/\log n}\,\big) \to 0 \qquad (3.33)\ \{?\}$$

as $n\to\infty$.

Proof. Recall $\tau_2 = 2m\alpha + \frac{3}{2}$ and $\tau_1 = 4m\alpha + 2$ and $g(x) = x^{\tau_1}\log^{\tau_2}(1+x)$ for $x \ge 0$. Observe that

$$Eg(|\xi_1|) = E\big[|x_{11}|^{\tau_1}\cdots|x_{1m}|^{\tau_1}\log^{\tau_2}(1+|x_{11}|\cdots|x_{1m}|)\big].$$

Use the inequality $1+|x_{11}|\cdots|x_{1m}| \le (1+|x_{11}|)\cdots(1+|x_{1m}|)$ to see that

$$\begin{aligned}\log^{\tau_2}(1+|x_{11}|\cdots|x_{1m}|) &\le \Big[\sum_{j=1}^m \log(1+|x_{1j}|)\Big]^{\tau_2}\\ &\le m^{\tau_2-1}\sum_{j=1}^m \log^{\tau_2}(1+|x_{1j}|)\end{aligned}$$

by the convex inequality. Obviously, the given condition $E\big[|x_{11}|^{\tau_1}\log^{\tau_2}(1+|x_{11}|)\big] < \infty$ implies that $E(|x_{11}|^{\tau_1}) < \infty$. It follows that

$$\begin{aligned}Eg(|\xi_1|) &\le m^{\tau_2-1}\sum_{j=1}^m E\big[|x_{11}|^{\tau_1}\cdots|x_{1m}|^{\tau_1}\log^{\tau_2}(1+|x_{1j}|)\big]\\ &= m^{\tau_2}E\big[|x_{11}|^{\tau_1}\log^{\tau_2}(1+|x_{11}|)\big]\cdot\big(E|x_{11}|^{\tau_1}\big)^{m-1}\\ &< \infty. \qquad (3.34)\ \{?\}\end{aligned}$$

Therefore,

$$P\big(|\xi_1| > \sqrt{n/\log n}\,\big) \le \frac{Eg(|\xi_1|)}{g\big(\sqrt{n/\log n}\,\big)}.$$

Trivially, $\log(1+\sqrt{\frac{n}{\log n}}) \ge \frac{1}{3}\log n$ as n is sufficiently large. We then see that

$$g\big(\sqrt{n/\log n}\,\big) \ge 3^{-\tau_2}n^{\tau_1/2}(\log n)^{\tau_2-(\tau_1/2)} = 3^{-\tau_2}n^{2m\alpha+1}\sqrt{\log n}. \qquad (3.35)\ \{?\}$$

In summary,

$$n^{1+(c_n^2/2)}\sqrt{\log n}\cdot P\big(|\xi_1| > \sqrt{n/\log n}\,\big) = O\big(n^{(c_n^2/2)-2m\alpha}\big)$$

The condition $p = O(n^\alpha)$ implies that $\log p \le \alpha\log n + O(1)$. Then we have from (3.23) that

$$\begin{aligned}\frac{c_n^2}{2} = \frac{mv_p^2}{\log n} &\le \frac{1}{\log n}\cdot\Big(m\log p - \frac{1}{3}\log\log p\Big)\\ &\le m\alpha + \frac{1}{\log n}\cdot\Big(O(1) - \frac{1}{3}\log\log p\Big)\end{aligned}$$

as n is sufficiently large. Hence

$$n^{(c_n^2/2)-2m\alpha} \le n^{-m\alpha} \cdot \exp\Big(O(1) - \frac{1}{3}\log\log p\Big) = O\Big(\frac{1}{n^{m\alpha}(\log p)^{1/3}}\Big)$$

as $n \to \infty$. The assertion (3.33) is yielded. □

Lemma 3.6 *Let the assumptions of Theorem 2 hold. Recall ν_p as in (3.23). Set $a_n = \sqrt{2m/\log p}\,\nu_p$. Let $\Psi_n^{(s)}(a_n)$ be as in (3.15). Then $\max_{1\le s\le m-1}\Psi_n^{(s)}(a_n) = O(p^{-2m+\delta})$ for any $\delta > 0$.*

Proof. It is enough to show $\Psi_n^{(s)}(a_n) = O(p^{-2m+\delta})$ for each $1 \le s \le m-1$, where $\delta > 0$ is given. Similar to (3.16) we have that

$$\begin{aligned}
\Psi_n^{(s)}(a_n) &\le P\Big(\big|\sum_{k=1}^n \xi_k^{(s)}\big(\eta_k^{(s)} + \zeta_k^{(s)}\big)\big| \ge 2a_n\sqrt{n\log p}\Big) \\
&\quad + P\Big(\big|\sum_{k=1}^n \xi_k^{(s)}\big(\eta_k^{(s)} - \zeta_k^{(s)}\big)\big| \ge 2a_n\sqrt{n\log p}\Big) \\
&:= A_n + B_n, \qquad (3.36)\ \{?\}
\end{aligned}$$

where $\xi_k^{(s)}, \eta_k^{(s)}$ and $\zeta_k^{(s)}$ are as in (3.14). Define $V_k = \xi_k^{(s)}\big(\eta_k^{(s)} + \zeta_k^{(s)}\big)/\sqrt{2}$ for $1 \le k \le n$. Then $\sum_{k=1}^n \xi_k^{(s)}\big(\eta_k^{(s)} + \zeta_k^{(s)}\big) = \sqrt{2}\sum_{k=1}^n V_k$. Observe that V_k's are i.i.d. random variables with

$$EV_1 = 0 \ \text{ and } \ EV_1^2 = 1. \qquad (3.37)\ \{?\}$$

Review $\tau_2 = 2m\alpha + \frac{3}{2}$, $\tau_1 = 4m\alpha + 2$ and $g(x) = x^{\tau_1}\log^{\tau_2}(1+x)$ for $x \ge 0$ as in Lemma 3.5. We claim $g(x)$ is a convex function on $[0,\infty)$. In fact,

$$g'(x) = \tau_1 x^{\tau_1-1}\log^{\tau_2}(1+x) + \frac{\tau_2 x^{\tau_1}\log^{\tau_2-1}(1+x)}{1+x}$$

Since $\tau_2 > \frac{3}{2}$ and $\tau_1 > 2$, the function $\tau_1 x^{\tau_1-1}\log^{\tau_2}(1+x)$ is increasing in $x \in [0,\infty)$, and hence its derivative is non-negative. Therefore, the convexity of $g(x)$ hinges on whether $h(x) := \frac{x^{\tau_1}\log^{\tau_2-1}(1+x)}{1+x}$ is increasing on $[0,\infty)$. Trivially,

$$\begin{aligned}
h'(x) &= \frac{1}{(1+x)^2}\Big[\underbrace{(1+x)}_{J_1}\Big(\tau_1 x^{\tau_1-1}\log^{\tau_2-1}(1+x) + \underbrace{(\tau_2-1)\frac{x^{\tau_1}\log^{\tau_2-2}(1+x)}{1+x}}_{J_2}\Big) \\
&\qquad - x^{\tau_1}\log^{\tau_2-1}(1+x)\Big] \\
&\ge \frac{(\tau_1-1)x^{\tau_1}\log^{\tau_2-1}(1+x)}{(1+x)^2} \\
&\ge 0
\end{aligned}$$

by using the fact $J_1 > x$ and $J_2 \geq 0$. Thus, $g(x)$ is convex on $[0, \infty)$. Now, by the convex property,

$$\begin{aligned} Eg(|V_1|) &\leq Eg\Big(\frac{2|\xi_1^{(s)}\eta_1^{(s)}| + 2|\xi_1^{(s)}\zeta_1^{(s)}|}{2}\Big) \\ &\leq \frac{1}{2}\big[Eg\big(2|\xi_1^{(s)}\eta_1^{(s)}|\big) + Eg\big(2|\xi_1^{(s)}\zeta_1^{(s)}|\big)\big] \\ &= Eg\big(2|\xi_1^{(s)}\eta_1^{(s)}|\big). \end{aligned}$$

Since $\log(1+2x) \leq 2\log(1+x)$ for $x \geq 0$, we have $g(2x) \leq 2^{\tau_2+\tau_1}g(x)$ for $x \geq 0$. By (3.14), $\xi_1^{(s)}\eta_1^{(s)} = x_{11}x_{12}\cdots x_{1m}$. Consequently,

$$Eg(|V_1|) \leq 2^{\tau_2+\tau_1}Eg(|x_{11}x_{12}\cdots x_{1m}|) < \infty$$

by (3.34). This particularly implies $E\big[g(|V_1|)I(|V_1| > \sqrt{n/\log n})\big] \to 0$. Now,

$$A_n = P\Big(\big|\sum_{k=1}^{n} V_i\big| \geq c_n\sqrt{n\log n}\Big)$$

where $c_n := a_n\sqrt{2(\log p)/\log n}$. By the Markov inequality,

$$\begin{aligned} P\big(|V_1| > \sqrt{n/\log n}\,\big) &\leq \frac{E\big[g(|V_1|)I(|V_1| > \sqrt{n/\log n})\big]}{g\big(\sqrt{n/\log n}\,\big)} \\ &= o\Big(g\big(\sqrt{n/\log n}\,\big)^{-1}\Big). \end{aligned}$$

From (3.35),

$$g\big(\sqrt{n/\log n}\,\big) \geq 3^{-\tau_2}n^{2m\alpha+1}\sqrt{\log n}.$$

By definition, $\lim_{n\to\infty} a_n = \sqrt{2m}$ and $a_n \leq \sqrt{2m}$ as n is sufficiently large. Then

$$c_n \leq 2\sqrt{m}\cdot\Big(\frac{\log p}{\log n}\Big)^{1/2} = 2\sqrt{m\alpha}\Big[1 + O\Big(\frac{1}{\log n}\Big)\Big]$$

by the assumption $p = O(n^\alpha)$. This implies that

$$\begin{aligned} n^{1+(c_n^2/2)}\sqrt{\log n}\cdot P\big(|V_1| > \sqrt{n/\log n}\,\big) &= o\big(n^{(c_n^2/2)-2m\alpha}\big) \\ &= o\big(n^{O(1/\log n)}\big) \to 0 \end{aligned}$$

as $n \to \infty$. Therefore, it is seen from (3.37) and Proposition 1 that

$$P\Big(\big|\sum_{k=1}^{n} V_i\big| \geq c_n\sqrt{n\log n}\Big) \sim 1 - \Phi(c_n\sqrt{\log n}).$$

Noting $c_n\sqrt{\log n} = a_n\sqrt{2\log p} \sim \sqrt{4m\log p}$ and $1-\Phi(x) \sim \frac{1}{\sqrt{2\pi}\,x}\cdot e^{-x^2/2}$ as $x\to\infty$. It follows that

$$P\Big(\big|\sum_{k=1}^{n} V_i\big| \ge c_n\sqrt{n\log n}\Big) = O\Big(e^{-c_n^2(\log n)/2}\Big) = O(p^{-2m+\delta})$$

as $n\to\infty$ for any $\delta>0$. Therefore, $A_n = O(p^{-2m+\delta})$. Similarly, $B_n = O(p^{-2m+\delta})$. The proof follows from (3.36). □

Proof of Theorem 2. Set $I = \{(i_1,\cdots,i_m)\in\mathbb{Z}^p : 1\le i_1<\cdots<i_m\le p\}$. For each $\alpha=(i_1,\cdots,i_m)\in I$, define

$$X_\alpha = \frac{1}{\sqrt{n}}\big|\sum_{k=1}^{n} x_{ki_1}x_{ki_2}\cdots x_{ki_m}\big|$$

and

$$B_\alpha = \big\{(j_1,\cdots,j_m)\in I;\ \{j_1,\cdots,j_m\}\cap\{i_1,\cdots,i_m\}\neq\emptyset \text{ but } (j_1,\cdots,j_m)\neq\alpha\big\}.$$

Obviously, X_α is independent of $\{X_\beta;\ \beta\in I\backslash X_\alpha\}$. Review (3.21) - (3.25) in the proof of Theorem 1. Set $\alpha_0=\{1,2,\cdots,m\}\in I$. It is seen from Lemma 3.1 that,

$$\Big|P(\max_{\alpha\in I} X_\alpha \le \sqrt{2m}\nu_p) - e^{-\lambda_p}\Big| \le b_1+b_2, \tag{3.38} \{?\}$$

where b_1 and b_2 are as in Lemma 3.1 and

$$\lambda_p \;\sim\; \frac{p^m}{m!}P\Big(\big|\sum_{k=1}^{n} x_{k1}x_{k2}\cdots x_{km}\big| > \sqrt{2mn}\nu_p\Big).$$

Write $\sqrt{2mn}\nu_p = c_n\cdot\sqrt{n\log n}$. Immediately $c_n\to\sqrt{2m\alpha}$ as $n\to\infty$ by (3.23). Set $\xi_i = x_{i1}x_{i2}\cdots x_{im}$ for $1\le i\le n$. Then $E\xi_1=0$, $\mathrm{Var}(\xi_1)=1$ and

$$n^{1+(c_n^2/2)}\sqrt{\log n}\cdot P\big(|\xi_1|>\sqrt{n/\log n}\,\big)\to 0$$

as $n\to\infty$ by Lemma 3.5. The assumption $E\big[|x_{11}|^{\tau_1}\log^{\tau_2}(1+|x_{11}|)\big]<\infty$ implies that $E|x_{11}|^{\tau_1}<\infty$, and hence $E|\xi_1|^{\tau_1}<\infty$ with $\tau_1 = 4m\alpha+2>2$. We then have from Proposition 1 that

$$\begin{aligned} P\Big(\big|\sum_{k=1}^{n} x_{k1}x_{k2}\cdots x_{km}\big| > \sqrt{2mn}\nu_p\Big) \;&\sim\; P\big(|N(0,1)|>\sqrt{2m}\nu_p\big) \\ &\sim\; \frac{2}{\sqrt{4\pi m}\,v_p}\cdot e^{-mv_p^2} \end{aligned} \tag{3.39} \{?\}$$

as in (3.26). Hence,

$$\lambda_p \;\sim\; \frac{p^m}{m!}\frac{2}{\sqrt{4m\pi\log p}}\cdot e^{-mv_p^2} = e^{-z/2}.$$

Immediately,

$$P\big(X_\alpha > \sqrt{2m}\nu_p\big) \sim \frac{m!}{p^m} e^{-z/2}$$

as $n \to \infty$ for any $\alpha \in I$. Similar to (3.28) and (3.29), we get

$$b_1 = O\Big(\frac{1}{p}\Big). \qquad (3.40)\ \{?\}$$

Now we work on b_2. Recalling (3.30) and (3.31) we have

$$b_2 \quad \le \quad |I| \cdot \sum_{s=1}^{m-1} m^s p^{m-s} \cdot \Psi_n^{(s)}(a_n)$$

where $a_n := \sqrt{2m/\log p}\,\nu_p$ for $n \ge 1$ and

$$\Psi_n^{(s)}(a_n) = P\Big(\big|\sum_{k=1}^n \xi_k^{(s)}\eta_k^{(s)}\big| \ge a_n\sqrt{n\log p},\ \big|\sum_{k=1}^n \xi_k^{(s)}\zeta_k^{(s)}\big| \ge a_n\sqrt{n\log p}\Big)$$

and $\xi_k^{(s)}$, $\eta_k^{(s)}$ and $\zeta_k^{(s)}$ are as in (3.14). By Lemma 3.6,

$$\max_{1\le s\le m-1} \Psi_n^{(s)}(a_n) \le O(p^{-2m+\delta})$$

as $n \to \infty$ for any $\delta > 0$. By using the fact $|I| \le p^m$, we have

$$\begin{aligned} b_2 \quad &\le \quad p^m \cdot m^m p^{m-1} \cdot p^{-2m+\delta} \\ &= \quad m^m p^{-1+\delta} \to 0 \end{aligned}$$

as $n \to \infty$ for any $\delta \in (0,1)$. This joining with (3.38)-(3.40) completes the proof. □

References

?s19? [1] Ambainis, A. and Harrow, A. W. (2012). Random tensor theory: extending random matrix theory to mixtures of random product states. *Commun. Math. Phys.*, **310(1)**, 25-74.

?s1? [2] Amosova, N. N. (1972). On limit theorem for the probabilities of moderate deviations. Vestnik Leningrad. Univ., 13, 5-14.

?s2? [3] Arratia, R., Goldstein, L. and Gorden, L. (1989). Two moments suffice for Poisson approximation: The Chen-Stein method. *Ann. Probab.* **17**, 9-25.

?s3? [4] Cai, T., Fan, J. and Jiang, T. (2013). Distribution of angles in random packing on spheres. *J. Mach. Learn. Res.* **14**, 1837-1864.

?s4? [5] Cai, T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39**, 1496-1525.

?s5? [6] Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse setting. *J. Amer. Stat. Assoc.* **108**, 265-277.

?s6? [7] Chen, L. H. Y., Fang, X. Shao, Q. (2013). From Stein identities to moderate deviations. *Ann. Probab.* **41**, 262-293.

?s7? [8] Fan, J. and Jiang, T. (2018). Largest entry of sample correlation matrix from normal populations with compound symmetry structure. *Preprint*

?s8? [9] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14**, 865-880.

?s9? [10] Li, D., Liu, W. and Rosalsky, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148**, 5-35.

?s10? [11] Li, D., Qi, Y. and Rosalsky, A. (2012). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.* **111**, 256-270.

?s11? [12] Li, D. and Rosalsky, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **16**, 423-447.

?s12? [13] Linnik, J. V. (1961). On the probability of large deviations for sums of independent variabels. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.* Vol II, 289-306. Univ. California Press, Berkeley.

?s13? [14] Liu, W., Lin, Z. and Shao, Q. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18**, 2337-2366.

?s20? [15] Lytova, A. (2017). Central limit theorem for linear eigenvalue statistics for a tensor product version of sample covariance matrices. *J. Theor. Probab.*, **31**, 1024-1057.

?s14? [16] Shao, Q. and Zhou, W. (2014). Necessary and sufficient conditions for the asymptotic distribution of coherence of ultral-high dimensional random matrices. *Ann. Probab.* **42**, 623-648.

?s21? [17] Shi, X., Qiu, R., He, X., Chu, L., Ling, Z. and Yang, H. (2018). Anomaly detection and location in distribution networks: a data-driven approach. *arXiv:1801.01669.*

?s20? [18] Tieplova, D. (2016). Distribution of eigenvalues of sample covariance matrices with tensor product samples. *J. Math. Phys. Anal. Geom.*, **13(1)**, 1024-1057.

?s15? [19] Xiao, H. and Wu, W. (2013). Asymptotic thery for maximum deviations of sample covariance matrix estimates. *Stoch. Process Their Appl.* **123**, 2899-2920.

?s16? [20] Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359**, 5345-5363.