

# *Limiting Behavior of Largest Entry of Random Tensor Constructed by High-Dimensional Data*

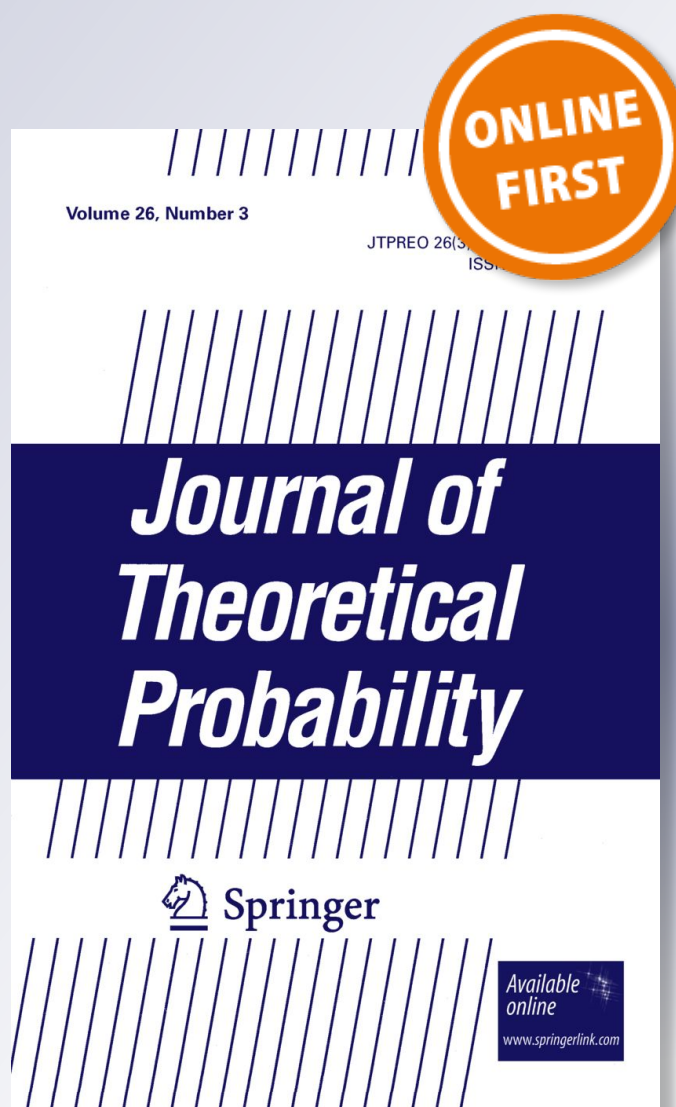
**Tiefeng Jiang & Junshan Xie**

**Journal of Theoretical Probability**

ISSN 0894-9840

J Theor Probab

DOI 10.1007/s10959-019-00958-1



 Springer

**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



# Limiting Behavior of Largest Entry of Random Tensor Constructed by High-Dimensional Data

Tiefeng Jiang<sup>1</sup> · Junshan Xie<sup>2</sup> Received: 19 August 2019 / Revised: 5 October 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

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

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

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  $p \rightarrow \infty$  and  $\log p = o(n^\beta)$  and the high-dimension case with  $p \rightarrow \infty$  and  $p = O(n^\alpha)$  where  $\alpha, \beta > 0$ . The normalizing constant of  $W_n$  depends on  $m$  and the limiting 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

**Mathematics Subject Classification (2010)** 60F05 · 62H10

---

Jiang's research was supported by NSF Grants No. DMS-1406279 and DMS-1916014, and Xie's research was supported by NNSF Grant No. 11401169 and KRPB Grant No. 20A110001.

---

✉ Junshan Xie  
junshan@henu.edu.cn

Tiefeng Jiang  
tjiang040@umn.edu

<sup>1</sup> School of Statistics, University of Minnesota, 224 Church Street, S. E., Minneapolis, MN 55455, USA

<sup>2</sup> School of Mathematics and Statistics, Henan University, Kaifeng 475000, China

### 1 Introduction

Let  $p \geq 2$  be an integer and  $X \in \mathbb{R}^p$  be a random vector. The distribution of  $X$  serves as a population distribution. Let  $X_k = (x_{k1}, \dots, x_{kp})'$ ,  $1 \leq k \leq n$ , be a random sample of size  $n$  from the population distribution generated by  $X$ , that is,  $X, X_1, \dots, X_n$  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  $\mathbf{T} \in \mathbb{R}^{p \times \dots \times p}$  with order  $m$  and rank  $n$  defined by

$$\mathbf{T} = \sum_{k=1}^n \underbrace{X_k \otimes \dots \otimes X_k}_m = \left( \sum_{k=1}^n x_{ki_1} x_{ki_2} \dots x_{ki_m} \right)_{1 \leq i_1, \dots, i_m \leq p} \tag{1}$$

Researchers have obtained some limiting properties of the tensor defined in (1). By using similar techniques to those in the random matrix theory, Ambainis and Harrow [1] obtained a limiting property of the largest eigenvalue and the limiting spectral distribution of random tensors. Tiepova [18] studied the limiting spectral distribution of the sample covariance matrices constructed by the random tensor data. Lytova [15] further considered the central limit theorem for linear spectral statistics of the sample covariance matrices constructed by the random tensor data. Shi et al. [17] applied limiting properties of the random tensors to an 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

$$W_n := \max_{1 \leq i_1 < \dots < i_m \leq p} \frac{1}{\sqrt{n}} \left| \sum_{k=1}^n x_{ki_1} x_{ki_2} \dots x_{ki_m} \right| \tag{2}$$

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

For a fixed  $m \geq 2$ , when the entries of the data matrix  $\mathbf{X} = (\mathbf{x}_{ki})_{1 \leq k \leq n, 1 \leq i \leq p}$  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  $p \rightarrow \infty$  and  $\log p = o(n^\beta)$  and the high dimension with  $p \rightarrow \infty$  and  $p = O(n^\alpha)$  where  $\alpha, \beta > 0$ . 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  $n/p \rightarrow \gamma > 0$  and  $E|x_{11}|^{30+\epsilon} < \infty$  for some  $\epsilon > 0$ , Jiang [9] proved that

$$W_n^2 - 4 \log p + \log \log p \xrightarrow{d} W_\infty \tag{3}$$

where random variable  $W_\infty$  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 [20] showed that (3) holds if

$$x^6 P(|x_{11}x_{12}| > x) \rightarrow 0 \text{ as } x \rightarrow \infty. \tag{4}$$

Liu et al. [14] proved that (3) holds provided a weaker condition is valid, that is,

$$n^3 P(|x_{11}x_{12}| > \sqrt{n \log n}) \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{5}$$

Besides the above two results, Li and Rosalsky [12] and Li et al. [10,11] further studied the moment condition for which (3) is true. In a different direction, Liu et al. [14] obtained (3) for the polynomial rate such that  $p = O(n^\alpha)$ ; Cai and Jiang [5] derived (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 Cai and Jiang [5], Cai et al. [4,6], Xiao and Wu [19] and Shao and Zhou [16].

In this paper, we prove that  $W_n$  with a suitable normalization converges to the Gumbel-type distribution for all  $m \geq 2$ . The normalizing constant and the limiting distribution all depend on  $m$ . Throughout the paper, the symbols  $\xrightarrow{p}$  and  $\xrightarrow{d}$  mean convergence in probability and convergence in distribution, respectively. We will also denote  $b_n = o(a_n)$  if  $\lim_{n \rightarrow \infty} b_n/a_n = 0$ ; 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  $\lim_{n \rightarrow \infty} a_n/b_n = 1$ .

The rest of the paper is organized as follows. The main results of the paper are stated in Sect. 2. In Sect. 3, we 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 of possible confusion, we will write  $p = p_n$ . Recall  $\mathbf{X} = (\mathbf{x}_{ki})_{1 \leq k \leq n, 1 \leq i \leq p}$  and assume

$$\{x_{ki}; 1 \leq k \leq n, 1 \leq i \leq p\} \text{ are i.i.d. random variables with } Ex_{11} = 0 \text{ and } Ex_{11}^2 = 1. \tag{6}$$

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

$$F_\theta(z) = \exp \left\{ - \frac{1}{m! \sqrt{m\pi}} e^{-z/2} \right\}, \quad z \in \mathbb{R}. \tag{7}$$

**Theorem 1** Assume (6) with  $Ee^{t_0|x_{11}|^\alpha} < \infty$  for some  $0 < \alpha \leq 1$  and  $t_0 > 0$ . Let  $m \geq 2$  be a fixed integer and  $\beta = \frac{\alpha}{2m-\alpha}$ . If  $p \rightarrow \infty$  and  $\log p = o(n^\beta)$  as  $n \rightarrow \infty$ , then  $W_n^2 - 2m \log p + \log \log p \xrightarrow{d} \theta$ , where  $\theta$  has distribution function  $F_\theta(z)$  in (7).

The above theorem studies the ultra-high-dimensional case, that is, the dimension  $p$  can be at an exponential order of the size  $n$ . Here, the assumption of  $Ee^{t_0|x_{11}|^\alpha} < \infty$

for some  $0 < \alpha \leq 1$  and  $t_0 > 0$  is needed to derive the limiting distribution. Next, we consider a popular high-dimensional case in the literature such that  $p = p_n$  is at most a polynomial power of  $n$ . We then get the same limiting distribution for  $W_n$  under a much weaker moment condition.

**Theorem 2** *Let  $\alpha > 0$  and  $m \geq 2$  be constants such that*

$$E[|x_{11}|^{\tau_1} \log^{\tau_2}(1 + |x_{11}|)] < \infty$$

*with  $\tau_1 = 4m\alpha + 2$  and  $\tau_2 = 2m\alpha + \frac{3}{2}$ . If  $p \rightarrow \infty$  and  $p = O(n^\alpha)$ , then  $W_n^2 - 2m \log p + \log \log p \xrightarrow{d} \theta$ , where  $\theta$  has distribution function  $F_\theta(z)$  in (7).*

The above two theorems obviously imply the following.

**Corollary 1** *Assume the conditions from either Theorems 1 or 2 hold. Then,*

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

As discussed earlier, the largest entry of a sample covariance matrix has been studied with the limiting distribution stated in (3). In this paper, we study the same problem for  $m$ -order random tensor, 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 (7). Both quantities indeed depend on  $m$ . We now make some further comments below.

1. Take  $m = 2$ , both Theorems 1 and 2 state that

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

which is consistent with (3).

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

$$\tilde{W}_n := \max_{1 \leq i_1 < \dots < i_m \leq p} \frac{1}{\sqrt{n}} \sum_{k=1}^n x_{ki_1} x_{ki_2} \cdots x_{ki_m}. \tag{8}$$

Then, by using the same proofs except changing “ $|N(0, 1)|$ ” to “ $N(0, 1)$ ” in (34) and (47), Theorems 1 and 2 still hold with the limiting distribution “ $F_\theta(z)$ ” from (7) is replaced by “ $F(z)$ ”, where

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

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

3. Recently Fan and Jiang [8] studied the limiting behavior of  $\tilde{W}_n$  from (8) with  $m = 2$  and with  $(x_{11}, \dots, x_{1p})' \sim N(0, \Sigma)$ , where  $\Sigma_{ii} = 1$  for each  $i$  and  $\Sigma_{ij} \equiv \rho > 0$  for all  $i \neq j$ . The limiting distribution of  $\tilde{W}_n$  is the Gumbel distribution if  $\rho$  is very

small; Gaussian if  $\rho$  is large; and the convolution of the Gumbel and the Gaussian distributions if  $\rho$  is in between. Such setting can also be extended to  $\tilde{W}_n$  in (8) for any  $m$  with a lengthy argument. We leave it as a future work.

4. Assume that  $m = 2$  and that  $(x_{11}, \dots, x_{1p})' \sim N(0, \Sigma)$ , where  $\Sigma$  is a banded matrix. Cai and Jiang [5] studied  $W_n$  from (2) and applied 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 Stein–Chen Poisson approximation method and the moderate deviations. The major technicality comes from computing  $\lambda$  and bounding  $b_2$  appeared in Lemma 1. The major difference between our proofs here and those in the literature is that the evaluation of  $\lambda$  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 4 and 6 for details.
6. Taking  $m = 2$  and  $\alpha = 1$  in Theorem 2, the required moment condition in the theorem becomes  $E[|x_{11}|^{10} \log^{5.5}(1 + |x_{11}|)] < \infty$ . This is stronger than (4) and (5). In fact, Lemma 6 requires this condition. It might be possible to relax to this moment assumption in Theorem 2. We also 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 independent populations with the same dimension  $p$ . Let  $X^{(l)} \in \mathbb{R}^p$ ;  $l = 1, 2, \dots, m$  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  $(x_{k1}^{(l)}, \dots, x_{kp}^{(l)})'$ ,  $k = 1, \dots, n$ , 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  $\mathbf{T}' \in \mathbb{R}^{p \times \dots \times p}$  with order  $m$  and rank  $n$  by

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

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

$$W'_n = \max_{1 \leq i_1 < \dots < i_m \leq p} \frac{1}{\sqrt{n}} \left| \sum_{k=1}^n x_{ki_1}^{(1)} x_{ki_2}^{(2)} \dots x_{ki_m}^{(m)} \right|.$$

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 “ $W'_n$ ” under some uniform moment conditions on  $x_{11}^{(l)}$ ,  $1 \leq l \leq m$ .

### 3 Proofs

#### 3.1 Some Technical Lemmas

We start with some technical lemmas useful for our proofs. The first one is a classical Stein–Chen Poisson approximation lemma, which is frequently used in studying

behaviors of maxima of almost mutual independent random variables. The following result is a special case of Theorem 1 of Arratia et al. [3].

**Lemma 1** *Let  $\{\eta_\alpha, \alpha \in I\}$  be random variables on an index set  $I$  and  $\{B_\alpha, \alpha \in I\}$  be a family 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*

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

where

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

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

Let  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$  for  $x \in \mathbb{R}$ . The following conclusion is about the moderate deviation of the partial sum of i.i.d. random variables (Linnik [13]).

**Lemma 2** *Suppose  $\{\zeta, \zeta_1, \zeta_2, \dots\}$  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  $0 < \alpha \leq 1$  and  $t_0 > 0$ , then*

$$\lim_{n \rightarrow \infty} \frac{1}{x_n^2} \log P\left(\frac{S_n}{\sqrt{n}} \geq x_n\right) = -\frac{1}{2}$$

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

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

$$\frac{P\left(\frac{S_n}{\sqrt{n}} \geq x\right)}{1 - \Phi(x)} \rightarrow 1$$

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

The following result is Proposition 4.5 from Chen et al. [7].

**Lemma 3** *Let  $\eta_i, 1 \leq i \leq n$ , be independent random variables with  $E\eta_i = 0$  and  $Ee^{h_n|\eta_i|} < \infty$  for some  $h_n > 0$  and  $1 \leq i \leq n$ . Assume that  $\sum_{i=1}^n E\eta_i^2 = 1$ . Then,*

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



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

**Proposition 1** Let  $\{\xi_i; i \geq 1\}$  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  $\{c_n > 0; n \geq 1\}$  be constants with  $\sup_{n \geq 1} c_n < \infty$ . Assume

$$P(|\xi_1| > \sqrt{n/\log n}) = o\left(\frac{1}{n^{1+(c_n^2/2)}\sqrt{\log n}}\right) \tag{9}$$

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

Amosova [2] derived a similar result to Proposition 1 for independent but not necessarily identically distributed random variables. If  $\{\xi_i; i \geq 1\}$  are i.i.d. random variables and  $\lim_{n \rightarrow \infty} c_n = c$ , then Amosova concluded that  $P(S_n \geq c_n \sqrt{n \log n}) \sim 1 - \Phi(c_n \sqrt{\log n})$  under the condition  $E(|\xi_1|^{c^2+2+\epsilon}) < \infty$  for some  $\epsilon > 0$ . This moment condition implies (9) by the Markov inequality and hence our proposition holds. In particular, taking  $c_n \equiv c > 0$ , then (9) holds if  $E[|\xi_1|^{c^2+2} \log^{(c^2+3)/2}(1 + |\xi_1|)] < \infty$ . 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  $n \rightarrow \infty$ ,

$$\sup_{a \leq x \leq b} \left| \frac{P(S_n \geq \sqrt{nx})}{1 - \Phi(x)} - 1 \right| \rightarrow 0$$

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

$$c_n \sqrt{\log n} \rightarrow \infty \tag{10}$$

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

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

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

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

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

Furthermore,

$$\text{Var}(\xi''_i) \leq E[\xi_i^2 I(|\xi_i| > a_n)] \leq \frac{K}{a_n^{r-2}}. \tag{11}$$

Now

$$\text{Var}(\xi_i) = \text{Var}(\xi'_i) + \text{Var}(\xi''_i) + 2\text{Cov}(\xi'_i, \xi''_i). \tag{12}$$

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

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

by the assumption  $E\xi_i = 0$ , the Cauchy–Schwarz inequality and (11). This together with (11) and (12) implies that

$$\text{Var}(\xi_1) \geq \text{Var}(\xi'_1) \geq \text{Var}(\xi_1) - \frac{3K}{a_n^{r-2}}. \tag{13}$$

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,

$$P(S_n > u) \leq P(S'_n > u - v) + P(|S''_n| > v) \tag{14}$$

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

$$P(S_n > u) \geq P(S'_n > u + v) - P(|S''_n| > v). \tag{15}$$

From the definition of  $\xi'_i$ , it is easy to see that  $\sup_{n \geq 1} E|\xi'_1|^r \leq 2^r K$ . Note that

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

Hence,

$$\begin{aligned} P(|S''_n| > v) &\leq P\left(\left|\sum_{i=1}^n \xi_i I(|\xi_i| > a_n)\right| > v - \frac{nK}{a_n^{r-1}}\right) \\ &\leq nP(|\xi_1| > a_n) \end{aligned} \tag{16}$$

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

Step 2: the tail for  $S'_n$ . Set  $\sigma'^2 = \text{Var}(\xi'_1)$ . Trivially,  $\sigma' \rightarrow 1$  as  $n \rightarrow \infty$ . Take  $\eta_i = \frac{\xi'_i}{\sqrt{n}\sigma'}$ . Then,  $|\eta_i| \leq \frac{2}{\sigma'\sqrt{\log n}}$ . Therefore, we see from Lemma 3 that

$$\frac{P(S'_n \geq \sqrt{n}\sigma'x)}{1 - \Phi(x)} = 1 + C_n(1 + x^3)\gamma e^{4x^3\gamma} \tag{17}$$

where  $\sup_{n \geq 1} |C_n| \leq C$  and  $C$  is an absolute constant, and

$$\begin{aligned} \gamma &\leq \frac{n}{n^{3/2}\sigma'^3} \cdot E(|\xi'_1|^3 e^{2a_n x / (\sqrt{n}\sigma')}) \\ &\leq \frac{1}{\sqrt{n}\sigma'^3} \cdot e^{2a_n x / (\sqrt{n}\sigma')} \cdot E(|\xi'_1|^3). \end{aligned}$$

Use the fact  $\sup_{n \geq 1} E|\xi'_1|^r \leq 2^r K$  to see that  $\sup_{n \geq 1} E(|\xi'_1|^3) \leq 2^r K$  if  $r \geq 3$  by the Hölder inequality. If  $2 < r < 3$ , then write

$$\begin{aligned} E(|\xi'_1|^3) &= E(|\xi'_1|^r \cdot |\xi'_1|^{3-r}) \\ &\leq 2^{3-r} \left(\frac{n}{\log n}\right)^{(3-r)/2} \cdot E(|\xi'_1|^r) \\ &\leq 8K \cdot \left(\frac{n}{\log n}\right)^{(3-r)/2} \end{aligned}$$

by the facts that  $|\xi'_1| \leq 2a_n = 2\sqrt{n/\log n}$  and that  $\sup_{n \geq 1} E|\xi'_1|^r \leq 2^r K$ . In summary, if  $x = O(\sqrt{n}/a_n)$  then

$$\gamma \leq \begin{cases} O\left(\frac{(\log n)^{(r-3)/2}}{n^{(r/2)-1}}\right), & \text{if } 2 < r < 3; \\ O(n^{-1/2}), & \text{if } r \geq 3 \end{cases}$$

as  $n \rightarrow \infty$ . In particular, noting  $a_n = \sqrt{n/\log n}$ , we know that  $\gamma \rightarrow 0$  and  $x^3\gamma \rightarrow 0$  since  $x = O(\sqrt{n}/a_n)$ . Consequently, we have from (17) that

$$P(S'_n \geq \sqrt{n}\sigma'x) \sim 1 - \Phi(x) \tag{18}$$

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,  $v/u \rightarrow 0$  as  $n \rightarrow \infty$ . We still write  $u$  and  $v$  next sometimes for short notation. By (14), (15) and (16), we have

$$P(S_n > c_n\sqrt{n\log n}) \leq P(S'_n > u - v) + nP(|\xi_1| > \sqrt{n/\log n}) \tag{19}$$

and

$$P(S_n > c_n\sqrt{n\log n}) \geq P(S'_n > u + v) - nP(|\xi_1| > \sqrt{n/\log n}). \tag{20}$$

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 (18). Then,  $x = O(\sqrt{n}/a_n)$  by the assumption  $\sup_{n \geq 1} c_n < \infty$ . From (10),  $x \sim c_n \sqrt{\log n} \rightarrow \infty$ . It follows that

$$P(S'_n > u \pm v) \sim P\left(N(0, 1) > \frac{u \pm v}{\sqrt{n} \sigma'}\right) \sim \frac{1}{c_n \sqrt{2\pi \log n}} \exp\left\{-\frac{(u \pm v)^2}{2n\sigma'^2}\right\}$$

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

$$\exp\left\{-\frac{(u \pm v)^2}{2n\sigma'^2}\right\} \cdot \exp\left\{\frac{c_n^2 \log n}{2}\right\} \rightarrow 1 \tag{21}$$

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

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

The assertion (13) says that  $\sigma'^2 \rightarrow 1$  and  $0 \leq 1 - \sigma'^2 \leq \frac{3K}{a_n^{r-2}} = O\left(\left(\frac{\log n}{n}\right)^{(r/2)-1}\right)$ . Also,

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

It follows that

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

We then confirm (21). 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,

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

Comparing these with (19) and (20), we arrive at

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

as  $n \rightarrow \infty$ . □

### 3.2 Main Proofs

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

$$\xi_k^{(s)} = \prod_{t=1}^s x_{kt}, \quad \eta_k^{(s)} = \prod_{t=s+1}^m x_{kt}, \quad \zeta_k^{(s)} = \prod_{t=m+1}^{2m-s} x_{kt}. \tag{22}$$

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

$$\Psi_n^{(s)}(a_n) = P\left(\left| \sum_{k=1}^n \xi_k^{(s)} \eta_k^{(s)} \right| \geq a_n \sqrt{n \log p}, \left| \sum_{k=1}^n \xi_k^{(s)} \zeta_k^{(s)} \right| \geq a_n \sqrt{n \log p}\right) \tag{23}$$

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

**Lemma 4** *Let  $\{a_n; n \geq 1\}$  be a sequence of positive numbers with  $\lim_{n \rightarrow \infty} a_n = a > 0$ . Under the assumptions of Theorem 1, we have that  $\max_{1 \leq s \leq m-1} \Psi_n^{(s)}(a_n) = o(p^{-a^2+\epsilon})$  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  $|u| \geq w$  and  $|v| \geq w$ . It then follows from (23) that

$$\begin{aligned} \Psi_n^{(s)}(a_n) &\leq P\left(\left| \sum_{k=1}^n \xi_k^{(s)} (\eta_k^{(s)} + \zeta_k^{(s)}) \right| \geq 2a_n \sqrt{n \log p}\right) \\ &\quad + P\left(\left| \sum_{k=1}^n \xi_k^{(s)} (\eta_k^{(s)} - \zeta_k^{(s)}) \right| \geq 2a_n \sqrt{n \log p}\right) \\ &:= A_n + B_n. \end{aligned} \tag{24}$$

For the term  $A_n$ , trivially,

$$\begin{aligned} E[\xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)})] &= 0, \\ E[\xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)})^2] &= 2. \end{aligned} \tag{25}$$

It is elementary that

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

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

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

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

$$E \exp \left\{ t_0 |\xi_1^{(s)} \eta_1^{(s)}|^{\alpha/m} \right\} < \infty. \tag{26}$$

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

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

by the Cauchy–Schwarz inequality again. 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  $\{\xi_k^{(s)} (\eta_k^{(s)} + \zeta_k^{(s)}); 1 \leq k \leq n\}$  are i.i.d. random variables. By Lemma 2 (1) and (25), we get that, for any sufficient small  $\delta > 0$ ,

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

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

$$A_n = o(p^{-a^2+\epsilon}) \tag{27}$$

as  $n \rightarrow \infty$ . Similarly,

$$B_n = o(p^{-a^2+\epsilon}). \tag{28}$$

Combining (24), (27) and (28), we complete the proof. □

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

$$X_\alpha = \frac{1}{\sqrt{n}} \left| \sum_{k=1}^n x_{ki_1} x_{ki_2} \cdots x_{ki_m} \right| \tag{29}$$

and

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

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

$$|I| = \binom{p}{m} \text{ and } |B_\alpha| \leq m^2 p^{m-1} \tag{30}$$

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

$$v_p = \left[ \log p - \frac{1}{2m} (\log \log p + 2 \log(m! \sqrt{m\pi}) - z) \right]^{1/2}. \tag{31}$$

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

$$\left| P(\max_{\alpha \in I} X_\alpha \leq \sqrt{2m} v_p) - e^{-\lambda_p} \right| \leq b_1 + b_2, \tag{32}$$

where  $b_1$  and  $b_2$  are as in Lemma 1 and

$$\begin{aligned} \lambda_p &= \binom{p}{m} P(X_{\alpha_0} > \sqrt{2m} v_p) \\ &\sim \frac{p^m}{m!} P\left(\left| \sum_{k=1}^n x_{k1} x_{k2} \cdots x_{km} \right| > \sqrt{2mn} v_p\right). \end{aligned} \tag{33}$$

First, write  $\psi_k = x_{k1} x_{k2} \cdots x_{km}$ ,  $1 \leq k \leq n$ . Then,  $E\psi_k = 0$  and  $E\psi_k^2 = 1$ . The assertion (26) says that  $Ee^{t_0|\psi_1|^{\alpha/m}} < \infty$ . Note that  $\frac{\alpha}{m} \leq \frac{1}{2}$  since  $0 < \alpha \leq 1$  and

$m \geq 2$ . Moreover,  $\sqrt{2m}v_p = O(\sqrt{\log p}) = o(n^{\beta/2})$ . By the definition of  $\beta$ , we know  $\frac{\beta}{2} = \frac{1}{2} \cdot \frac{\alpha/m}{2-(\alpha/m)}$ . Therefore, it follows from Lemma 2(2) that

$$P\left(\left|\sum_{k=1}^n x_{k1}x_{k2} \cdots x_{km}\right| > \sqrt{2mn}v_p\right) \sim P(|N(0, 1)| > \sqrt{2m}v_p) \sim \frac{2}{\sqrt{4\pi m} v_p} \cdot e^{-mv_p^2}, \tag{34}$$

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

$$v_p \sim \sqrt{\log p};$$

$$-mv_p^2 = -\log(p^m) + \frac{1}{2}(\log \log p + 2 \log(m!\sqrt{m\pi})) - \frac{1}{2}z$$

as  $n \rightarrow \infty$ . Therefore,

$$\lambda_p \sim \frac{p^m}{m!} \frac{2}{\sqrt{4m\pi \log p}} \cdot e^{-mv_p^2} = e^{-z/2}. \tag{35}$$

In particular, this implies that

$$P(X_\alpha > \sqrt{2m}v_p) \sim \frac{m!}{p^m} e^{-z/2} \tag{36}$$

as  $n \rightarrow \infty$  for any  $\alpha \in I$ . Consequently, we have from (30) that

$$b_1 \leq |I| \cdot |B_\alpha| \cdot P(X_\alpha > \sqrt{2m}v_p)^2 = O\left(\frac{1}{p}\right). \tag{37}$$

Now, we estimate  $b_2$ . First,

$$b_2 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(X_\alpha > \sqrt{2m}v_p, X_\beta > \sqrt{2m}v_p)$$

$$= \sum_{\alpha \in I} \sum_{s=1}^{m-1} \sum_{\beta \in I: |\beta \cap \alpha| = s} P(X_\alpha > \sqrt{2m}v_p, X_\beta > \sqrt{2m}v_p). \tag{38}$$

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

$$\{\beta \in I : |\beta \cap \{1, 2, \dots, m\}| = s\}$$

$$= \{(i_1, \dots, i_m) \in \mathbb{Z}^p : 1 \leq i_1 < \dots < i_m \leq p, |(i_1, \dots, i_m) \cap \{1, 2, \dots, m\}| = s\}.$$



Hence,  $|\{\beta \in I : |\beta \cap \{1, 2, \dots, m\}| = s\}| = \binom{m}{s} \cdot \binom{p-m}{m-s} \leq m^s p^{m-s}$ . Review the notations in (22) and (23). In particular,

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

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

$$b_2 \leq |I| \cdot \sum_{s=1}^{m-1} m^s p^{m-s} \cdot \Psi_n^{(s)}(a_n) \tag{39}$$

where

$$a_n := \frac{\sqrt{2m} v_p}{\sqrt{\log p}} \rightarrow \sqrt{2m}$$

as  $n \rightarrow \infty$ . By Lemma 4, for any  $\epsilon > 0$ , we have

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

as  $n$  is large enough. This implies that

$$\begin{aligned} b_2 &\leq p^m \cdot m^m p^{m-1} \cdot p^{-2m+\epsilon} \\ &= m^m p^{-1+\epsilon} \rightarrow 0 \end{aligned} \tag{40}$$

as  $n \rightarrow \infty$  for all  $\epsilon \in (0, 1)$ . Combining (32), (35), (37) and (40), we complete the proof.  $\square$

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

**Lemma 5** *Let  $x_{ij}$ 's be as in Theorem 2 and  $v_p$  be as in (31). Define  $c_n = \sqrt{2m/\log n} v_p$  and  $\xi_1 = x_{11}x_{12} \cdots x_{1m}$ . Then*

$$n^{1+(c_n^2/2)} \sqrt{\log n} \cdot P(|\xi_1| > \sqrt{n/\log n}) \rightarrow 0 \tag{41}$$

as  $n \rightarrow \infty$ .

**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  $x \geq 0$ . Observe that

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

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

$$\log^{\tau_2}(1 + |x_{11}| \cdots |x_{1m}|) \leq \left[\sum_{j=1}^m \log(1 + |x_{1j}|)\right]^{\tau_2}$$

$$\leq m^{\tau_2-1} \sum_{j=1}^m \log^{\tau_2}(1 + |x_{1j}|)$$

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

$$\begin{aligned} Eg(|\xi_1|) &\leq m^{\tau_2-1} \sum_{j=1}^m E[|x_{11}|^{\tau_1} \cdots |x_{1m}|^{\tau_1} \log^{\tau_2}(1 + |x_{1j}|)] \\ &= m^{\tau_2} E[|x_{11}|^{\tau_1} \log^{\tau_2}(1 + |x_{11}|)] \cdot (E|x_{11}|^{\tau_1})^{m-1} \\ &< \infty. \end{aligned} \tag{42}$$

Therefore,

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

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

$$g(\sqrt{n/\log n}) \geq 3^{-\tau_2} n^{\tau_1/2} (\log n)^{\tau_2-(\tau_1/2)} = 3^{-\tau_2} n^{2m\alpha+1} \sqrt{\log n}. \tag{43}$$

In summary,

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

The condition  $p = O(n^\alpha)$  implies that  $\log p \leq \alpha \log n + O(1)$ . Then, we have from (31) that

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

as  $n$  is sufficiently large. Hence,

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

as  $n \rightarrow \infty$ . The assertion (41) is yielded. □

**Lemma 6** *Let the assumptions of Theorem 2 hold. Recall  $v_p$  as in (31). Set  $a_n = \sqrt{2m/\log p} v_p$ . Let  $\Psi_n^{(s)}(a_n)$  be as in (23). Then,  $\max_{1 \leq s \leq m-1} \Psi_n^{(s)}(a_n) = O(p^{-2m+\delta})$  for any  $\delta > 0$ .*

**Proof** It is enough to show  $\Psi_n^{(s)}(a_n) = O(p^{-2m+\delta})$  for each  $1 \leq s \leq m - 1$ , where  $\delta > 0$  is given. Similar to (24) we have that

$$\begin{aligned} \Psi_n^{(s)}(a_n) &\leq P\left(\left|\sum_{k=1}^n \xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)})\right| \geq 2a_n\sqrt{n \log p}\right) \\ &\quad + P\left(\left|\sum_{k=1}^n \xi_k^{(s)}(\eta_k^{(s)} - \zeta_k^{(s)})\right| \geq 2a_n\sqrt{n \log p}\right) \\ &:= A_n + B_n, \end{aligned} \tag{44}$$

where  $\xi_k^{(s)}$ ,  $\eta_k^{(s)}$  and  $\zeta_k^{(s)}$  are as in (22). Define  $V_k = \xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)})/\sqrt{2}$  for  $1 \leq k \leq n$ . Then,  $\sum_{k=1}^n \xi_k^{(s)}(\eta_k^{(s)} + \zeta_k^{(s)}) = \sqrt{2} \sum_{k=1}^n V_k$ . Observe that  $V_k$ 's are i.i.d. random variables with

$$EV_1 = 0 \text{ and } EV_1^2 = 1. \tag{45}$$

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  $x \geq 0$  as in Lemma 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 nonnegative. 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,

$$\begin{aligned} h'(x) &= \frac{1}{(1+x)^2} \left[ \underbrace{(1+x)}_{J_1} \left( \tau_1 x^{\tau_1-1} \log^{\tau_2-1}(1+x) + \underbrace{(\tau_2 - 1) \frac{x^{\tau_1} \log^{\tau_2-2}(1+x)}{1+x}}_{J_2} \right) \right. \\ &\quad \left. - x^{\tau_1} \log^{\tau_2-1}(1+x) \right] \\ &\geq \frac{(\tau_1 - 1)x^{\tau_1} \log^{\tau_2-1}(1+x)}{(1+x)^2} \\ &\geq 0 \end{aligned}$$

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,

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

$$= Eg(2|\xi_1^{(s)}\eta_1^{(s)}|).$$

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 (22),  $\xi_1^{(s)}\eta_1^{(s)} = x_{11}x_{12} \cdots x_{1m}$ . Consequently,

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

by (42). This particularly implies  $E[g(|V_1|)I(|V_1| > \sqrt{n/\log n})] \rightarrow 0$ . Now,

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

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

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

From (43),

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

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

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

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

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

as  $n \rightarrow \infty$ . Therefore, it is seen from (45) and Proposition 1 that

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

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  $x \rightarrow \infty$ . It follows that

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

as  $n \rightarrow \infty$  for any  $\delta > 0$ . Therefore,  $A_n = O(p^{-2m+\delta})$ . Similarly,  $B_n = O(p^{-2m+\delta})$ . The proof follows from (44).  $\square$

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

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

and

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

Obviously,  $X_\alpha$  is independent of  $\{X_\beta; \beta \in I \setminus X_\alpha\}$ . Review (29)–(33) in the proof of Theorem 1. Set  $\alpha_0 = \{1, 2, \dots, m\} \in I$ . It is seen from Lemma 1 that,

$$\left| P(\max_{\alpha \in I} X_\alpha \leq \sqrt{2m}v_p) - e^{-\lambda_p} \right| \leq b_1 + b_2, \tag{46}$$

where  $b_1$  and  $b_2$  are as in Lemma 1 and

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

Write  $\sqrt{2mn}v_p = c_n \cdot \sqrt{n \log n}$ . Immediately  $c_n \rightarrow \sqrt{2m\alpha}$  as  $n \rightarrow \infty$  by (31). Set  $\xi_i = x_{i1}x_{i2} \cdots x_{im}$  for  $1 \leq i \leq n$ . Then,  $E\xi_1 = 0$ ,  $\text{Var}(\xi_1) = 1$  and

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

as  $n \rightarrow \infty$  by Lemma 5. The assumption  $E[|x_{11}|^{\tau_1} \log^{\tau_2}(1 + |x_{11}|)] < \infty$  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

$$\begin{aligned} P\left(\left|\sum_{k=1}^n x_{k1} x_{k2} \cdots x_{km}\right| > \sqrt{2mn}v_p\right) &\sim P(|N(0, 1)| > \sqrt{2m}v_p) \\ &\sim \frac{2}{\sqrt{4\pi m} v_p} \cdot e^{-mv_p^2} \end{aligned} \tag{47}$$

as in (34). Hence,

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

Immediately,

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

as  $n \rightarrow \infty$  for any  $\alpha \in I$ . Similar to (36) and (37), we get

$$b_1 = O\left(\frac{1}{p}\right). \tag{48}$$

Now, we work on  $b_2$ . Recalling (38) and (39), we have

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

where  $a_n := \sqrt{2m/\log p} v_p$  for  $n \geq 1$  and

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

and  $\xi_k^{(s)}, \eta_k^{(s)}$  and  $\zeta_k^{(s)}$  are as in (22). By Lemma 6,

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

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

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

as  $n \rightarrow \infty$  for any  $\delta \in (0, 1)$ . This joining with (46)–(48) completes the proof.  $\square$

**Acknowledgements** The authors would like to thank the referee and an associate editor for their constructive comments and suggestions that have led to improvements in the paper.

## References

1. Ambainis, A., Harrow, A.W.: Random tensor theory: extending random matrix theory to mixtures of random product states. *Commun. Math. Phys.* **310**(1), 25–74 (2012)
2. Amosova, N.N.: On limit theorem for the probabilities of moderate deviations. *Vestn. Leningr. Univ.* **13**, 5–14 (1972)
3. Arratia, R., Goldstein, L., Gordon, L.: Two moments suffice for Poisson approximation: the Chen–Stein method. *Ann. Probab.* **17**, 9–25 (1989)
4. Cai, T., Fan, J., Jiang, T.: Distribution of angles in random packing on spheres. *J. Mach. Learn. Res.* **14**, 1837–1864 (2013)
5. Cai, T., Jiang, T.: Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Stat.* **39**, 1496–1525 (2011)

6. Cai, T., Liu, W., Xia, Y.: Two-sample covariance matrix testing and support recovery in high-dimensional and sparse setting. *J. Am. Stat. Assoc.* **108**, 265–277 (2013)
7. Chen, L.H.Y., Fang, X., Shao, Q.: From Stein identities to moderate deviations. *Ann. Probab.* **41**, 262–293 (2013)
8. Fan, J., Jiang, T.: Largest entry of sample correlation matrix from normal populations with compound symmetry structure. *Ann. Probab.* **47**(5), 3321–3374 (2019)
9. Jiang, T.: The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14**, 865–880 (2004)
10. Li, D., Liu, W., Rosalsky, A.: Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Relat. Fields* **148**, 5–35 (2010)
11. Li, D., Qi, Y., Rosalsky, A.: On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivar. Anal.* **111**, 256–270 (2012)
12. Li, D., Rosalsky, A.: Some strong limit theorems for the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **16**, 423–447 (2006)
13. Linnik, J. V.: On the probability of large deviations for sums of independent variables. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability: Volume II*, pp. 289–306. University California Press, Berkeley (1961)
14. Liu, W., Lin, Z., Shao, Q.: 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 (2008)
15. Lytova, A.: Central limit theorem for linear eigenvalue statistics for a tensor product version of sample covariance matrices. *J. Theor. Probab.* **31**, 1024–1057 (2017)
16. Shao, Q., Zhou, W.: Necessary and sufficient conditions for the asymptotic distribution of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42**, 623–648 (2014)
17. Shi, X., Qiu, R., He, X., Chu, L., Ling, Z., Yang, H.: Anomaly detection and location in distribution networks: a data-driven approach. [arXiv:1801.01669](https://arxiv.org/abs/1801.01669) (2018)
18. Tiepova, D.: Distribution of eigenvalues of sample covariance matrices with tensor product samples. *J. Math. Phys. Anal. Geom.* **13**(1), 1024–1057 (2016)
19. Xiao, H., Wu, W.: Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stoch. Process Their Appl.* **123**, 2899–2920 (2013)
20. Zhou, W.: Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Am. Math. Soc.* **359**, 5345–5363 (2007)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.