The asymptotic Distributions of the Largest Entries of Sample Correlation Matrices

TIEFENG JIANG
July 4, 2002

Abstract Let $X_n = (x_{ij})$ be an $n$ by $p$ data matrix, where the $n$ rows form a random sample of size $n$ from a certain $p$-dimensional population distribution. Let $R_n = (\rho_{ij})$ be the $p \times p$ sample correlation matrix of $X_n$; that is, the entry $\rho_{ij}$ is the usual Pearson’s correlation coefficient between the $i$-th column of $X_n$ and $j$-th column of $X_n$. For contemporary data both $n$ and $p$ are large. When the population is a multivariate normal we study the test that $H_0$: the $p$ variates of the population are uncorrelated. A test statistic is chosen as $L_n = \max_{i \neq j} |\rho_{ij}|$. The asymptotic distribution of $L_n$ is derived by using the Chen-Stein Poisson approximation method. Similar results for non-Gaussian case are also derived.

1 Introduction

Let $X_n = (x_{ij})$ be an $n$ by $p$ data matrix, where the $n$ rows are observations from a certain multivariate distribution and each of $p$ columns is an $n$ observation from a variable of the population distribution. Let $\rho_{ij}$ be the Pearson correlation coefficient between the $i$-th and $j$-th columns of $X_n$. That is

$$
\rho_{ij} = \frac{\sum_{k=1}^{n} (x_{k,i} - \bar{x}_i)(x_{k,j} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{k,i} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^{n} (x_{k,j} - \bar{x}_j)^2}}.
$$

(1.1)

where $\bar{x}_i = (1/n) \sum_{k=1}^{n} x_{k,i}$. Then $R_n := (\rho_{ij})$ is a $p$ by $p$ symmetric matrix. It is called the sample correlation matrix generated by $X_n$.

Suppose the population is a multivariate normal distribution with mean vector $\mu$, covariance matrix $\Sigma$ and correlation coefficient matrix $R$. When the sample size $n$ and the dimension $p$ are large and comparable, Johnstone [13] studied the test with null hypothesis $H_0 : \Sigma = I$ under assumption that $\mu = 0$, where $I$ is the identity matrix. The null hypothesis is equivalent to that the population distribution is the product of $p$ univariate standard

\footnote{School of Statistics, University of Minnesota.}

\textbf{Key Words} sample correlation matrices, maxima, Chen-Stein method, moderate deviations.

normal distributions. The test statistic is chosen as the maximum eigenvalue of the sample covariance matrix $X_n^T X_n$ according to the method Principal Component Analysis (PCA). It is proved that the asymptotic distribution of the maximum eigenvalue is the Tracy-Widom law.

When both $n$ and $p$ are large we consider the test with null hypothesis

$$H_0: R = I.$$  \hfill (1.2)

Equivalently, the population distribution is a product of univariate normal distribution $N(\mu_i, \sigma_i^2)$’s for some unknown $\mu_i$’s and unknown $\sigma_i$’s. The difference between this test and the one in Johnstone [13] mentioned above is that all $\mu_i$’s don’t have to be identical and all $\sigma_i$’s don’t have to be identical, either. Besides, we don’t assume that $\mu_i$’s and $\sigma_i$’s are known. Our test seems to be more natural and practical. The maximum eigenvalue $\lambda_{\text{max}}$ of the sample correlation matrix $R_n$ can be taken as the test statistic according to PCA. But the distribution of $\lambda_{\text{max}}$ is not clear so far although there is an evidence that $\lambda_{\text{max}}$ may also follow the Tracy-Widom law asymptotically as shown in [12].

In this paper, we don’t pursue the maximum eigenvalue $\lambda_{\text{max}}$ as the test statistic because of its complexity. Instead we choose the following intuitive one:

$$L_n = \max_{1 \leq i < j \leq p} |\rho_{ij}|,$$

where $\rho_{ij}$ is as in (1.1). Evidently, $L_n$ is close to zero under the null hypothesis that the $p$ variables of $N_p(0, \Sigma)$ are uncorrelated. We will derive the strong law and limiting distribution of $L_n$. In fact, we will prove some general results; the observations $x_{ij}$’s don’t have to be Gaussian. We state our results precisely next.

Suppose $\{\xi, x_{ij}, i, j = 1, 2, \cdots\}$ are i.i.d. random variables. Let $X_n = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$. Let $x_1, x_2, \cdots, x_p$ be the $p$ columns of $X_n$. Then $X_n = (x_1, x_2, \cdots, x_p)$. Let $\bar{x}_k$ be the sample average of $x_k$, that is, $\bar{x}_k = (1/n) \sum_{i=1}^n x_{ik}$. We write $x_i - \bar{x}_i$ for $x_i - \bar{x}_i e$, where $e = (1, 1, \cdots, 1)^T \in \mathbb{R}^n$. Then, $\rho_{ij}$, the Pearson correlation coefficient in (1.1) between $x_i$ and $x_j$ can be rewritten as

$$\rho_{ij} = \frac{(x_i - \bar{x}_i)^T (x_j - \bar{x}_j)}{\|x_i - \bar{x}_i\| \cdot \|x_j - \bar{x}_j\|}, \quad 1 \leq i, j \leq p, \tag{1.3}$$

where $\|\cdot\|$ is the usual Euclidean norm. Obviously, $\rho_{ii} = 1$ for each $i$.

First, we obtain a strong limit theorem as follows.

**Theorem 1** Suppose $E|\xi|^{18 - \epsilon} < \infty$ for any $\epsilon > 0$. If $n/p \to \gamma \in (0, \infty)$, then

$$\lim_{n \to \infty} \sqrt{\frac{n}{\log n}} L_n = 2 \; \text{a.s.}$$
The above strong law of $L_n$ does not depend on $p$ although $X_n$ is an $n$ by $p$ matrix. For limiting distribution the following holds.

**Theorem 2** Suppose that $E|\xi|^{18+\epsilon} < \infty$ for some $\epsilon > 0$. If $n/p \to \gamma$, then

$$P\left(nL_n^2 - 4 \log n + \log(\log n) \leq y\right) \to e^{-Ke^{-y/2}}$$

as $n \to \infty$ for any $y \in \mathbb{R}$, where $K = (\gamma^2 \sqrt{8\pi})^{-1}$.

The limiting distributions appeared in the above Theorem is called the extreme distribution of type I.

For constants $a_i \in \mathbb{R}^2$ and $b_i \in \mathbb{R}^1$, $i = 1, 2, \ldots, p$, it is easy to see that the matrix $(a_1x_1 + b_1e, a_2x_2 + b_2e, \ldots, a_px_p + b_pe)$ and $X_n = (x_1, x_2, \ldots, x_p)$ generate the common correlation matrix $R_n$. Also, if $\xi \sim N(0, 1)$, then $Ee^{t\xi^2} < \infty$ for all $t < 1/2$. We immediately have the following result.

**Corollary 1.1** Suppose $\{x_{ij}; i \geq 1, j \geq 1\}$ are independent and $x_{ij} \sim N(\mu_j, \sigma_j^2)$ for some $\mu_j$ and $\sigma_j \neq 0$ for all $i$ and $j$. Let the sample correlation matrix $R_n$ be obtained from $X_n := (x_{ij}; 1 \leq i \leq n, 1 \leq j \leq p)$. Then the conclusions of Theorems 1 and 2 also hold.

Let

$$F(x) = \int_{-\infty}^{x} e^{-Ke^{-y/2}} dy,$$

where $K = (\gamma^2 \sqrt{2\pi})^{-1}$. Let $e_p$ stand for the $(100p)$-th percentile of the cumulative distribution function $F(x)$, that is, the unique solution $e_p$ for the equation $F(e_p) = p$. By using the above corollary, we are able to give a rejection region for the hypothesis test stated in (1.2). If the alternative hypothesis is $H_a : \mathbb{R} \neq I$, then $H_0$ is rejected if $nL_n^2 - 4 \log n + \log(\log n) \notin (e_{\alpha/2}, e_{1-\alpha/2})$ corresponding to a significance level $\alpha \in (0, 1)$.

The following Theorem 3 is used in the proof of Theorems 3.1 and 3.2. It is a non-asymptotic inequality on the moderate deviation of partial sums of independent random variables. Though sums of independent random variables are well-understood we did not notice a similar result in the literature, e.g., [18] and [19]. The usual moderate results such as those in [14] and Theorem 3.7.1 on page 109 from [7] are not applicable in our case. The reason is that we do not have identical distribution assumption. Second, asymptotic bounds do not work in our proof because our case involves an uniform bound of infinitely many such probabilities. This is evident from Lemma 2.1 in Section 3. There is a similar situation in the large deviation case. The Chernoff bound (see, e.g., (c) of Remarks on page
27 from [7]) is a non-asymptotic bound of sums of i.i.d. random variables. But the classical
Cramér type large deviation is a limiting result. The Chernoff bound is used in the proof
of theorems in [9] and [10] for the same reason of proving our Theorems 1 and 2 via the
following Theorem 3.

**Theorem 3** Let \( \{\eta_i, 1 \leq i \leq n\} \) be a sequence of independent random variables with
mean zero. Assume \( \sup_{1 \leq i \leq n} E|\eta_i|^\beta < \infty \) for some \( \beta \geq 2 \). Then

\[
P \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i \geq t \right) \leq \frac{M_\beta}{n^{\rho \beta - 1}} + K_n e^{-t^2/(2M_2)},
\]

where

\[
M_s := \frac{1}{n} \sum_{i=1}^{n} E|\eta_i|^s \text{ for } s \in (0, \beta], \text{ and } K_n := \exp \left\{ \frac{t^3 n^{\rho - (1/2)}}{3M_2^2} e^{2t' n^{\rho - (1/2)}/M_2} \right\}
\]

and \( t' = t - M_\beta n^{\rho(1-\beta) + (1/2)} \) for any \( \rho > 0 \).

In our applications, \( K_n \sim 1, t' \sim t \) and \( M_2 = 1 \). Also, \( M_\beta/n^{\rho \beta - 1} \) in (1.4) is smaller than
the term next to it. So the probability is roughly bounded by \( e^{-t^2/2} \).

The main tool used in proving Theorems 1 and 2 are the Chen-Stein Poisson approxi-
mation method and probabilities of moderate deviations by Amosova [1], and Rubin and
Sethuraman [20]. They are listed in Appendix.

In traditional random matrix theories, eigenvalues are the primary concern. See, e.g.,
[17] and [5]. This paper together with [11], in which the maxima of entries of certain Haar-
distributed matrices were studied for an imaging analysis problem, suggests that the study
of maxima of entries of matrices are also important.

Now we state the outline of this paper. A couple of lemmas are given in Section 2 for
the preparation of the proofs of main results. We prove all main results in Section 3. In the
last section some known results used in the proofs of our theorems are listed.

## 2 Auxiliary Lemmas

Three lemmas are needed before we go to the proof of main results. The proof of the
following relies on Theorem 3, which will be proved at last. There is no circular reasoning.

**Lemma 2.1** Let \( \{\xi_k, \eta_k, \eta'_k, k = 1, 2, \ldots n\} \) be i.i.d. random variables with mean 0 and
variance 1. Let \( \{u_n; n \geq 1\} \) be a sequence of positive numbers such that \( u_n/\sqrt{n \log n} \rightarrow a \in \)
(0, \infty). If E|\xi_i|^q < \infty for some q > (a^2 + 2)^2/2, then
\[ P \left( \sum_{k=1}^{n} \xi_k \eta_k \geq u_n, \sum_{k=1}^{n} \xi_k \eta_k \geq u_n \right) = O \left( n^{b-a^2} \right) \] (2.1)
as \( n \to \infty \) for any \( b > 0 \).

**Proof.** The two events in (2.1) are conditionally independent given \( \xi_k \)'s. Denote by \( P^1 \) and \( E^1 \) such conditional probability and expectation, respectively. Then the probability in (2.1) is
\[
E \left[ P^1 \left( \left| \sum_{k=1}^{n} \xi_k \eta_k \right| \geq u_n \right)^2 \right]. \tag{2.2}
\]

Set
\[
A_n(s) = \left\{ \frac{1}{n} \sum_{k=1}^{n}(|\xi_k|^s - E|\xi_k|^s) \leq \delta \right\}
\]
for \( s \geq 2 \) and \( \delta \in (0,1) \). Choose \( \beta \in (a^2 + 2, 2q/(a^2 + 2)) \) and \( r = (a^2 + 2)/2 \). Let \( \zeta_k = |\xi_k|^\beta - E|\xi_k|^\beta \) for \( k = 1, 2, \ldots, n \). By the Chebyshev inequality and Theorem A.1
\[
P(A_n(\beta)^c) = P \left( \sum_{k=1}^{n} \zeta_k \right) > n\delta \leq (n\delta)^{-r} E \left[ \sum_{k=1}^{n} \zeta_k \right]^r = O \left( n^{-f(r)} \right) \tag{2.3}
\]
as \( n \to \infty \), where \( f(r) = r/2 \) if \( r \geq 2 \); \( = r - 1 \) if \( 1 \leq r \leq 2 \). Let \( \{\zeta'_{k}; 1 \leq k \leq n\} \) be an independent copy of \( \{\zeta_{k}; 1 \leq k \leq n\} \). Then, \( P \left( \left| \sum_{k=1}^{n} \zeta_k \right| \leq n\delta/2 \right) \geq 1/2 \) for sufficiently large \( n \). It follows that
\[
P \left( \left| \sum_{k=1}^{n} \zeta_k \right| > n\delta \right) \leq 2P \left( \left| \sum_{k=1}^{n} (\zeta_k - \zeta'_{k}) \right| > n\delta/2 \right) = O \left( n^{-f(r)} \right) \tag{2.4}
\]
by repeating (2.3). Since \( \sum_{k=1}^{n} (\zeta_k - \zeta'_{k}) \) is a sum of i.i.d. symmetric random variables, by Lemma A.2, there is a constant \( C > 0 \) such that
\[
P(A_n(\beta)^c) \leq CP(\max_{1 \leq i \leq n} |\zeta_i - \zeta'_{i}| \geq n\delta/C) + O \left( n^{1-r} \right)
\leq 2CnP(|\zeta_i| \geq n\delta/(2C)) + O \left( n^{1-r} \right) = O \left( n^{1-r} \right), \tag{2.5}
\]
where we use the fact \( E|\zeta_1|^r < \infty \) since \( E|\xi|^\beta < \infty \) in the last inequality. Thus, it follows from (2.2) that
\[
E \left[ P^1 \left( \left| \sum_{k=1}^{n} \xi_k \eta_k \right| \geq u_n \right)^2 \right] 
\leq E \left[ P^1 \left( \left| \sum_{k=1}^{n} \xi_k \eta_k \right| \geq u_n \right)^2 I_{A_n(2)\cap A_n(\beta)} \right] + O \left( n^{1-r} \right). \tag{2.6}
\]
Now we apply Theorem 3 to the last probability in (2.6). Note $E^t(\xi_k\eta_k) = 0$ and $E^t|\xi_k\eta_k|^s = |\xi_k|^s(E|\xi_1|^t)$ for any $s > 0$. In particular, $E^t(\xi_k\eta_k)^2 = \xi_k^2$. Thus

$$M_s = \frac{1}{n} \left( \sum_{k=1}^{n} |\xi_k|^s \right) \cdot E|\xi_1|^s.$$

Choose $t = u_n/\sqrt{n}$ and $\rho \in ((a^2 + 2)/(2\beta),1/2)$. It is easy to verify that

$$t - t' \to 0 \quad \text{and} \quad K_n \to 1$$

as $n \to \infty$. Then there is a constant $C > 0$ such that the probability in (2.6) under the restriction $A_n(2) \cap A_n(\beta)$ is less than

$$C \left( n^{-a^2/2} + n^{-a^2/(2(1+\delta)^2)} \right)$$

as $n$ is sufficiently large. Note that, $O(n^{1-r}) = O(n^{-a^2/2})$. Consequently the last term in (2.6) is negligible. Therefore, the probability in (2.1) is $O(n^{-a^2/(2(1+\delta)^2)})$ by (2.2) and (2.6). The proof then follows by choosing $\delta$ small enough. ■

For any matrix $A = (a_{i,j})$, the notation $\|A\|$ is used to stand for the maximum norm of $A$; that is, $\max |a_{i,j}|$ over all indices $i$ and $j$.

**Lemma 2.2** Recall $x_i$ in (1.3). Let $h_i = \|x_i - \bar{x}_i\|/\sqrt{n}$ for each $i$. Then

$$\|nR_n - X_n^TX_n\| \leq (b_1^2 + 2b_1)W_nb_3^{-2} + nb_3^{-2}b_4^2,$$

where

$$b_1 = \max_{1 \leq i \leq n} |h_i - 1|, \quad W_n = \max_{1 \leq i < j \leq n} |x_i^T x_j|, \quad b_3 = \min_{1 \leq i \leq n} h_i, \quad b_4 = \max_{1 \leq i \leq n} |\bar{x}_i|.$$

**Proof.** The $(i,j)$-entry of $R_n$ is

$$\rho_{ij} = \frac{(x_i - \bar{x}_i)^T(x_j - \bar{x}_j)}{\|x_i - \bar{x}_i\| \cdot \|x_j - \bar{x}_j\|} = \frac{x_i^T x_j - n\bar{x}_i \bar{x}_j}{nh_i h_j}.$$

The $(i,j)$-entry of $X_n^TX_n$ is $x_i^T x_j$. So

$$|n\rho_{ij} - x_i^T x_j| \leq |(h_i h_j)^{-1} - 1| \cdot |x_i^T x_j| + n \frac{|\bar{x}_i|}{h_i} \cdot \frac{|\bar{x}_j|}{h_j}.$$

Take maximum for both sides, we obtain

$$\|nR_n - X_n^TX_n\| \leq \max_{1 \leq i < j \leq n} |(h_i h_j)^{-1} - 1| \cdot \max_{1 \leq i < j \leq n} |x_i^T x_j| + n \left( \max_{1 \leq i \leq n} \frac{|\bar{x}_i|}{h_i} \right)^2.$$

Write $1 - (h_i h_j)^{-1} = (h_i - 1)(h_j - 1) + (h_i - 1) + (h_j - 1)$. Then the desired inequality follows. ■

Next we estimate $b_i$'s.
**Lemma 2.3** Suppose $n/p \to \gamma \in (0, \infty)$. If $E|\xi|^{4/(1-\alpha)} < \infty$ for some $\alpha \in (0,1/2)$ then

$$n^\alpha b_1 \to 0 \ a.s., \ b_3 \to 1 \ a.s. \ and \ n^\alpha b_4 \to 0 \ a.s.$$  

**Proof.** The second limit follows from the first one. Easily, $\|x_i - \bar{x}_i\|^2 = x_i^T x_i - n|\bar{x}_i|^2$. Using the fact that $|x-1| \leq |x^2-1|$ for any $x > 0$, we have that

$$n^\alpha b_1 \leq \max_{1 \leq i \leq n} \left| \frac{x_i^T x_i - n}{n^{1-\alpha}} \right| + \left( n^{\alpha/2} \max_{1 \leq i \leq n} |\bar{x}_i| \right)^2. \quad (2.7)$$

Note $x_i^T x_i = \sum_{k=1}^n x_{ki}^2$. By Lemma A.5, the first and the second maxima above go to zero when $E|\xi|^{4/(1-\alpha)} < \infty$. So the first limit is proved. The limit that $n^\alpha b_4 \to 0 \ a.s.$ is proved by noting the relationship between $n^\alpha b_4$ and the right most term in (2.7). 

The analysis of $W_n$ is given in the next section.

### 3 Proof of Main Results

The proofs of Theorems 1 and 2 rely on an analysis of covariance matrix $X_n^T X_n$. The $(i,j)$-entry of $X_n^T X_n$ is $\sum_{k=1}^n x_{ki} x_{kj}$. Recall

$$W_n = \max_{1 \leq i < j \leq n} \left| \sum_{k=1}^n x_{ki} x_{kj} \right|.$$ 

The first step in proving our main theorems is approximating $R_n$ by $X_n^T X_n$ as shown in Lemmas 2.2 and 2.3. This will be seen clearly in the proof of Theorems 1 and 2. The second step is deriving the corresponding results for $X_n^T X_n$. We actually will prove the following two lemmas.

**Lemma 3.1** Suppose that $E \xi = 0$ and $E|\xi|^{18-\epsilon} < \infty$ for any $\epsilon > 0$. If $n/p \to \gamma \in (0, \infty)$, then

$$(i) \ \limsup_{n \to \infty} \frac{W_n}{\sqrt{n \log n}} \leq 2 \ a.s. \quad (ii) \ \liminf_{n \to \infty} \frac{W_n}{\sqrt{n \log n}} \geq 2 \ a.s.$$ 

The above lemma actually says that $W_n/\sqrt{n \log n} \to 2 \ a.s.$ as $n \to \infty$. The reason we did not combine (i) and (ii) as a single limit is that the proof of the combined one is relatively long. We will prove the two parts separately.
**Lemma 3.2** Suppose that $E\xi = 0$ and $E|\xi|^{16+\epsilon} < \infty$ for some $\epsilon > 0$. If $n/p \to \gamma \in (0, \infty)$, then

$$P \left( \frac{W_n^2 - \alpha_n}{n} \leq y \right) \to e^{-Ke^{-\gamma/2}}$$

as $n \to \infty$ for any $y \in \mathbb{R}$, where $\alpha_n = 4n \log n - n \log(\log n)$ and $K = (\gamma^2 \sqrt{8\pi})^{-1}$.

Assuming Lemmas 3.1 and 3.2 we next prove Theorems 1 and 2. The proof of the former two lemmas are given later.

**Proof of Theorems 1 and 2.** Choose $\alpha = 1/3$. Under the condition that $E|\xi|^{6} < \infty$ we have from Lemmas 2.2 and 2.3 that

$$|nL_n - W_n| \leq \|nR_n - X_n^T X_n\| \leq 4n^{-1/3}W_n + 2n^{1/3} \text{ a.s.} \quad (3.1)$$

as $n$ is sufficiently large. By Lemma 3.1, $nL_n - W_n = O(n^{1/6}(\log n)^{1/2})$ a.s. Theorem 1 then follows immediately from Lemma 3.1. Now, Theorem 1 and Lemma 3.1 imply that $nL_n + W_n = O(\sqrt{n \log n})$. Consequently,

$$n^2L_n^2 - W_n^2 = O(n^{2/3}(\log n)) \text{ a.s.}$$

Theorem 2 then follows from Lemma 3.2. \hfill \Box

Now we turn to prove the two lemmas.

**Proof of (i) Lemma 3.1.** Given $\delta \in (0, 1)$, let $w_n = (2 + \delta)\sqrt{n \log n}$. By Lemma A.3, under the condition that $E|\xi|^{d} < \infty$ for some $d > 2 + (2 + \delta)^2$,

$$\max_{1 \leq i \neq j < \infty} P(|y_{ij}^{(l)}| > w_l) = P(|y_{12}^{(l)}| > w_{l}) = O \left( \frac{1}{l^{2+\delta}} \right) \quad (3.2)$$

as $l$ is large, where $y_{ij}^{(l)} := \sum_{k=1}^{l} x_{ki}x_{kj}$, $i,j \geq 1$. We also use the fact that

$$1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-t^2/2} \, dt \sim \frac{1}{\sqrt{2\pi} x} e^{-x^2/2} \quad (3.3)$$

as $x \to +\infty$. Evidently, for any integer $m > 4/\delta$,

$$\max_{n^m \leq l \leq (n+1)^m} W_l \leq \max_{1 \leq i \neq j \leq (n+1)^m} \left( \max_{n^m \leq l \leq (n+1)^m} |y_{ij}^{(l)}| \right) \leq \max_{1 \leq i \neq j \leq (n+1)^m} |y_{ij}^{(m)}| + r_n, \quad (3.4)$$

where

$$r_n = \max_{1 \leq i \neq j \leq (n+1)^m} \max_{n^m \leq l \leq (n+1)^m} |y_{ij}^{(l)} - y_{ij}^{(m)}|. \quad (3.5)$$
By (3.2),

\[ P \left( \max_{1 \leq i \neq j \leq (n+1)^m} |y_{ij}^{(m^*)}| > w_{n^m} \right) \leq (n + 1)^{2m} P \left( |y_{12}^{(m^*)}| > w_{n^m} \right) = O \left( n^{-\delta m} \right). \]

Since \( \sum_n n^{-\delta m} < \infty \), by the Borel-Cantelli lemma,

\[ \limsup_{n \to \infty} \frac{\max_{1 \leq i \neq j \leq (n+1)^m} |y_{ij}^{(m^*)}|}{\sqrt{n^m \log(n^m)}} \leq 2 + \delta \text{ a.s.} \quad (3.6) \]

Now let’s estimate \( r_n \) as in (3.5).

Let \( \{z_1, z_2, \ldots\} \) be i.i.d. random variables with the same law as \( x_{11}x_{12} \) with partial sums \( S_0 = 0 \) and \( S_k = \sum_{i=1}^k z_i \). Clearly, \( E z_1 = 0 \) and \( E z_1^2 = 1 \). Observe that the distribution of \( y_{ij}^{(l)} - y_{ij}^{(m^*)} \) is equal to that of \( S_{l-n^m} \) for all \( l \geq n^m \). Thus,

\[ P(r_n \geq \delta \sqrt{n^m \log(n^m)}) \]
\[ \leq n^{2m} P \left( \max_{1 \leq k \leq (n+1)^m - n^m} |S_k| \geq \delta \sqrt{n^m \log(n^m)} \right) \]
\[ \leq 2n^{2m} P \left( |S_{(n+1)^m - n^m}| \geq (\delta/2) \sqrt{n^m \log(n^m)} \right) \quad (3.7) \]

as \( n \) is sufficiently large, where Ottaviani’s inequality (see Exercise 16 on page 74 in [6]) is used in the last inequality. Set \( k_n = (n+1)^m - n^m \). Note that, for fixed \( m \) and \( \delta \), \( (\delta/2) \sqrt{n^m \log(n^m)} \geq (2+\delta) \sqrt{k_n \log k_n} \) as \( n \) is sufficiently large. By (3.3) and Lemma A.3, the probability in (3.7) is equal to \( O(\exp(-(m-1)(2+\delta)^2(\log n)/2)) \) provided \( E|\xi|^d < \infty \) for some \( d > 2 + (2+\delta)^2 \). Therefore,

\[ P(r_n \geq \delta \sqrt{n^m \log(n^m)}) = O(n^{-u}), \]

where \( u = (m-1)(2+\delta)^2/2 - 2m > 1 \) since \( m \) is chosen such that \( m > 4/\delta \). By the Borel-Cantelli lemma again,

\[ \limsup_{n \to \infty} \frac{r_n}{\sqrt{n^m \log(n^m)}} \leq \delta \text{ a.s.} \quad (3.8) \]

By (3.4), (3.6) and (3.8), we obtain that

\[ \limsup_{n \to \infty} \frac{\max_{1 \leq i \leq (n+1)^m} W_i}{\sqrt{n^m \log(n^m)}} \leq 2 + 2\delta \text{ a.s.} \]

for any sufficiently small \( \delta > 0 \). This implies inequality (i) in Lemma 3.1. ■
Proof of (ii) of Lemma 3.1. We continue to use the notations in the proof of (i) of Lemma 3.1. For any $\delta \in (0, 1)$, define $v_n = (2 - \delta)\sqrt{n \log n}$. We first claim that

$$P(W_n \leq v_n) = O\left(\frac{1}{n^{\delta'}}\right)$$  \hspace{1cm} (3.9)

as $n \to \infty$ for some positive constant $\delta'$ depending on $\delta$ and $\xi$ only. If this is true, take an integer $m$ such that $m > 1/\delta'$. Then $P(W_n^m \leq v_n^m) = O(1/n^{\delta'm})$. Since $\sum_{n} n^{-\delta'm} < \infty$, by the Borel-Cantelli Lemma, we have that

$$\liminf_{n \to \infty} \frac{W_n^m}{\sqrt{n^m \log(n^m)}} \geq 2 - \delta \quad a.s.$$  \hspace{1cm} (3.10)

for any $\delta \in (0, 1)$. Recall the definition of $r_n$ in (3.4), we have that

$$\inf_{n^m \leq k \leq (n+1)^m} W_k \geq W_n^m - r_n.$$ 

By (3.8) and (3.10), we have that

$$\liminf_{n \to \infty} \frac{\inf_{n^m \leq k \leq (n+1)^m} W_k}{\sqrt{n^m \log(n^m)}} \geq 2 - 2\delta \quad a.s.$$ 

for any $\delta$ small enough. This implies (ii) of Lemma 3.1.

Now we turn to prove claim (3.9) by Lemma A.4.

Take $I = \{(i, j); 1 \leq i < j \leq p\}$. For $\alpha = (i, j) \in I$, set $B_\alpha = \{(k, l) \in I; \text{ one of } k \text{ and } l = i \text{ or } j \text{, but } (k, l) \neq \alpha\}$, $\eta_\alpha = |y_{ij}^{(n)}|$, $t = v_n$ and $A_\alpha = A_{ij} = \{|y_{ij}^{(n)}| > v_n\}$. By Lemma A.4,

$$P(W_n \leq v_n) \leq e^{-\lambda_n} + b_{1,n} + b_{2,n}.$$  \hspace{1cm} (3.11)

Evidently

$$\lambda_n = \frac{p(p-1)}{2}P(A_{12}), \quad b_{1,n} \leq 2p^3 P(A_{12})^2 \quad \text{and} \quad b_{2,n} \leq 2p^3 P(A_{12}A_{13}).$$  \hspace{1cm} (3.12)

Remember that $y_{12}^{(n)}$ is a sum of i.i.d. random variables with mean 0 and variance 1. Recall (3.3). By Lemma A.3, 

$$P(A_{12}) \sim \frac{1}{(2-\delta)\sqrt{2\pi \log n}} \cdot \frac{1}{n^{(2-\delta)/2}}$$  \hspace{1cm} (3.13)

as $n \to \infty$ provided $E|\eta|^6 < \infty$. Note that $P(A_{12}A_{13}) = P(|y_{12}^{(n)}| \geq v_n, |y_{13}^{(n)}| \geq v_n)$ and $v_n/\sqrt{n \log n} \to 2 - \delta$. By Lemma 2.1, $P(A_{12}A_{13}) = O(n^{b-4\delta^2}/2) < 18$. Choosing both $b$ and $\delta$ small enough, we obtain

$$e^{-\lambda_n} \leq e^{-n^s}, \quad b_{1,n} \leq \frac{1}{\sqrt{n}} \quad \text{and} \quad b_{2,n} \leq \frac{1}{\sqrt{n}}$$  \hspace{1cm} (3.14)
for sufficiently large \( n \). Then (3.9) follows from (3.11) and (3.14).

**Proof of Lemma 3.2.** We need to show that
\[
P(\max_{1 \leq i < j \leq p} |y_{ij}| \leq \sqrt{\alpha_n + ny}) \to e^{-K e^{-y/2}},
\]
where \( y_{ij} = \sum_{k=1}^n x_{ki} x_{kj} \). Now we apply Lemma A.4 to prove (3.15). Take \( I = \{(i, j); 1 \leq i < j \leq p\} \). For \( \alpha = (i, j) \in I \), set \( X_\alpha = |y_{ij}| \) and \( B_\alpha = \{(k, l) \in I; \text{ one of } k \text{ and } l = i \text{ or } j, \text{ but } (k, l) \neq \alpha\} \). Choose \( t = \sqrt{\alpha_n + ny} \). We first calculate \( \lambda = \lambda_n \) in the theorem. Since \( \{y_{ij}; (i, j) \in I\} \) are identically distributed,
\[
\lambda_n = \sum_{1 \leq i < j \leq p} P(|y_{ij}| > \sqrt{\alpha_n + ny}) = \frac{p^2 - p}{2} P\left(\frac{|y_{12}|}{\sqrt{n}} > \sqrt{\frac{\alpha_n}{n} + y}\right).
\]
Observe that \( y_{12} \) is a sum of i.i.d. random variables with mean 0 and variance 1. Since \( \sqrt{\alpha_n/n} + y \sim 2\sqrt{\log n} \), by Lemma A.3, we have that
\[
P\left(\frac{|y_{12}|}{\sqrt{n}} > \sqrt{\frac{\alpha_n}{n} + y}\right)
= P\left(\frac{y_{12}}{\sqrt{n}} > \sqrt{\frac{\alpha_n}{n} + y}\right) + P\left(\frac{-y_{12}}{\sqrt{n}} > \sqrt{\frac{\alpha_n}{n} + y}\right) \sim e^{-y/2} \sqrt{\frac{2\pi n}{\alpha_n}}
\]
provided \( E|\xi|^q < \infty \) for some \( q > c^2 + 2 = 6 \). Thus
\[
\lambda_n \to \frac{e^{-y/2}}{\gamma^2 \sqrt{8\pi}}.
\]
Obviously, \( X_\alpha \) is independent of \( \{X_\beta; \beta \in I\setminus B_\alpha\} \) for any \( \alpha = (i, j) \in I \). To complete (3.15), by Lemma A.4, we have to verify that \( b_1 \to 0 \) and \( b_2 \to 0 \) as \( n \to \infty \). It is easy to check that the size of \( B_\alpha \) is less than \( 2p \). Thus
\[
b_1 \leq \frac{1}{2} (p^2 - p) \cdot (2p) \cdot P\left(\frac{|y_{12}|}{\sqrt{n}} > \sqrt{\frac{\alpha_n}{n} + y}\right)^2 = O\left(\frac{1}{n}\right)
\]
by (3.17). Also, by symmetry,
\[
b_2 \leq p (p^2 - p) P(|y_{12}| > \sqrt{\frac{\alpha_n}{n} + ny}, |y_{13}| > \sqrt{\frac{\alpha_n}{n} + ny}).
\]
Here, \( \sqrt{\alpha_n + ny}/\sqrt{n \log n} \to 2 \). By Lemma 2.1, the above probability is \( O(n^{b-4}) \) for any \( b > 0 \) provided \( E|\xi|^q < \infty \) for some \( q > (a^2 + 2)^2/2 = 18 \). Now choose \( b < 1 \), then \( b_2 \to 0 \).

By Lemma A.4, (3.15) is concluded.

\[\square\]
Proof of Theorem 3. Define

$$\tilde{\eta}_i = \eta_i I(\eta_i \leq n^\rho) - E\eta_i I(\eta_i \leq n^\rho)$$

for $\rho > 0$. It is easy to see that

$$P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}_i \geq t\right) \leq P\left(\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}_i \right) + r_n \geq t\right) + \sum_{i=1}^{n} P(\eta_i \geq n^\rho),$$

where $r_n = (1/\sqrt{n}) \sum_{i=1}^{n} E\eta_i I(\eta_i > n^\rho)$ because $E\eta_i = 0$. Clearly, by Markov’s inequality,

$$\sum_{i=1}^{n} P(\eta_i \geq n^\rho) \leq \frac{M_3}{n^{\rho^2-1}} \text{ and } |r_n| \leq \frac{M_3}{n^{\rho^2-(1/2)}}.$$

Thus, to prove the lemma, it suffices to show that

$$P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}_i \geq t'\right) \leq K_n e^{-t'^2/(2M_2)}, \quad (3.20)$$

By the Chebyshev inequality and independence, we obtain that

$$P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}_i \geq t'\right) \leq e^{-\theta t'/\sqrt{n}} \prod_{i=1}^{n} E e^{\theta \eta_i} / \sqrt{n} \quad (3.21)$$

for any $\theta > 0$. Since $e^x \leq 1 + x + (x^2/2) + (|x|^3/6) e^{|x|}$ for any $x \in \mathbb{R}$,

$$E e^{\theta \tilde{\eta}_i} / \sqrt{n} \leq 1 + \frac{\theta^2}{2n} E \tilde{\eta}_i^2 + \frac{\theta^3}{6n^{3/2}} E \left(|\tilde{\eta}_i|^3 \exp(\theta |\tilde{\eta}_i|/\sqrt{n}) \right). \quad (3.22)$$

Obviously, $E \tilde{\eta}_i^2 \leq E \eta_i^2$ and $|\tilde{\eta}_i| \leq 2n^\rho$. It follows that

$$E \left(|\tilde{\eta}_i|^3 \exp(\theta |\tilde{\eta}_i|/\sqrt{n}) \right) \leq 2n^\rho e^{2\theta n^\rho - (1/2)} E |\eta_i|^2.$$

Since $1 + x \leq e^x$ for any $x \in \mathbb{R}$, by (3.22), we have that

$$E e^{\theta \tilde{\eta}_i} / \sqrt{n} \leq \exp \left\{ \frac{\theta^2}{2n} E (\eta_i^2) + \frac{\theta^3 E |\eta_i|^2}{3n^{(3/2)-\rho}} e^{2\theta n^\rho - (1/2)} \right\}.$$

Substituting this back to (3.21), we obtain

$$P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}_i \geq t'\right) \leq \exp \left( -\theta t' + \frac{M_2}{2} \theta^2 + \frac{\theta^3 M_2}{3n^{(1/2)-\rho}} e^{2\theta n^\rho - (1/2)} \right)$$

for any $\theta > 0$. Then (3.20) follows by choosing $\theta = t' / M_2$. The proof is completed. $\blacksquare$
4 Appendix

For proofs of the main theorems, we quote some results from literature in this section.

The following is a corollary of the Marcinkiewicz-Zygmund inequality, see, e.g., corollary 2 on page 368 for $p \geq 2$ and theorem 2 on page 367 for $p \in [1, 2)$ from [6].

**Lemma A.1** If $\{\eta_n, n \geq 1\}$ are i.i.d. random variables with $E\eta_1 = 0$, $E|\eta_1|^p < \infty$, $p \geq 1$, and $S_n = \sum_{i=1}^{n} \eta_i$. Then,

$$E|S_n|^p = \begin{cases} O(n^{p/2}), & \text{if } p \geq 2; \\ O(n), & \text{if } 1 \leq p < 2. \end{cases}$$

The following is lemma 2.2 from [16], which is a useful version of the maximal inequality of Hoffmann-Jøgensen, see [8] or proposition 6.7 from [15].

**Lemma A.2** Let $\{\eta_k, 1 \leq k \leq n\}$ be independent symmetric random variables and $S_n = \sum_{k=1}^{n} \eta_k$. Then, for each integer $j \geq 1$, there exist positive numbers $C_j$ and $D_j$ depending only on $j$ such that for all $t > 0$,

$$P(|S_n| \geq 2jt) \leq C_j P(\max_{1 \leq k \leq n} |\eta_k| \geq t) + D_j (P(|S_n| \geq t))^j.$$

The following lemma is from [1]. It is a refinement of Theorem 2 from [20]. See also page 254 from [19].

**Lemma A.3** Let $\{\eta_1, \eta_2, \cdots\}$ be a sequence of independent random variables with $E\eta_i = 0$ and $E\eta_i^2 = \sigma_i^2 > 0$ for all $i$. Define $S_n = \sum_{i=1}^{n} \eta_i$ and $B_n = \sum_{i=1}^{n} \sigma_i^2$. Suppose

$$\liminf_n \left\{ \frac{B_n}{n} \right\} > 0 \quad \text{and} \quad \limsup_n \left\{ \frac{1}{n} \sum_{i=1}^{n} E|\eta_i|^q \right\} < \infty$$

for some $q > 2$. Let $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-t^2/2} dt$. Then

$$P \left( \frac{S_n}{\sqrt{B_n}} > x \right) \frac{1}{1 - \Phi(x)} \to 1$$

uniformly on $[0, c \sqrt{\log n}]$ for any $c \in (0, \sqrt{q - 2})$ as $n \to \infty$.

The following Poisson approximation result is essentially a special case of Theorem 1 in [3], which is again a special case of the general Chen-Stein Poisson approximation method. One application of the following lemma is studying behaviors of maxima of random variables. See, for example, [9] and [10].
**Lemma A.4** Let $I$ be an index set and $\{B_\alpha, \alpha \in I\}$ be a set of subsets of $I$, that is, $B_\alpha \subset I$ for each $\alpha \in I$. Let also $\{\eta_\alpha, \alpha \in I\}$ be random variables. For a given $t \in \mathbb{R}$, set $\lambda = \sum_{\alpha \in I} P(\eta_\alpha > t)$. Then

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

where

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t)P(\eta_\beta > t),$$

$$b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t),$$

$$b_3 = \sum_{\alpha \in I} E[P(\eta_\alpha > t|\sigma(\eta_\beta, \beta \notin B_\alpha)) - P(\eta_\alpha > t)],$$

and $\sigma(\eta_\beta, \beta \notin B_\alpha)$ is the $\sigma$-algebra generated by $\{\eta_\beta, \beta \notin B_\alpha\}$. Particularly, if $\eta_\alpha$ is independent of $\{\eta_\beta, \beta \notin B_\alpha\}$ for each $\alpha$, then $b_3 = 0$.

The following is lemma 2 from [4].

**Lemma A.5** Let $\{v, v_{ij}, i, j = 1, 2, \cdots\}$ be a double array of i.i.d. random variables and let $\alpha > 1/2$, $\beta > 0$ and $M > 0$ be constants. Then as $n \to \infty$,

$$\max_{1 \leq j \leq Mn^\beta} \left| \sum_{i=1}^{n} (v_{ij} - E v) \right| \to 0 \text{ a.s.,}$$

if and only if $E|v|^{(\beta+1)/\alpha} < \infty$.

**Acknowledgment.** The author thanks Iain Johnstone very much for suggesting the problems studied in this paper.

**References**


crash: tjiang@stat.umn.edu
Mailing Address: School of Statistics, 313 Ford Hall, 224 Church Street S.E., Minneapolis, MN55455