The Limiting Distributions of Eigenvalues of Sample Correlation Matrices

Short Title: Eigenvalues of Correlation Matrices

TIEFENG JIANG [1]

June 6, 2003

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 coefficient matrix of X_n. Assuming that x_{ij}'s are independent and identically distributed (x_{ij}'s are required to be only independent when they are normals), we show that the largest eigenvalue of R_n almost surely converges to a constant provided n/p goes to a positive constant. Under two conditions on the ratio n/p, we show that the empirical distribution of eigenvalues of R_n converges weakly to the Marčenko-Pastur law and the semi-circular law, respectively. This work is motivated by testing the hypothesis, assuming population distribution $N_p(\mu, \Sigma)$, that the p variates are uncorrelated.

1 Introduction

Suppose a population distribution is an p-dimensional multi-normal distribution with mean μ, covariance matrix Σ and correlation coefficient matrix R. In modern data, p is quite large and even close to a sample size. For such case, Johnstone [10] recently studied the test $H_0 : \Sigma = I$ assuming $\mu = 0$. That is, the p variates of the population distribution are independent and identically distributed. Given a random sample of size n, the asymptotic distribution of the largest eigenvalue of the sample covariance matrix generated by a data matrix is obtained in [10] by using Random Matrix Theory (Soshnikov [11] later generalized this result). Actually it is shown in [10] that suitable normalizations of such largest eigenvalues converge weakly to the *Tracy-Widom* law of order 1, which is defined through the solution of the (non-linear) Painlevé II differential equations.

[1] Research partially supported by NSF #DMS-0308151.

Key Words: sample correlation coefficient matrix, largest eigenvalue, empirical distributions of eigenvalues, semi-circular law, Marčenko-Pastur law.

AMS (2000) subject classifications: 60F05, 60F15, 62H20.

Note that the null hypothesis in the test is $\Sigma = I$. A more general test is $R = I$, i.e., the p variates of the population are uncorrelated with both mean μ and the p standard deviations of those variates unknown. Suppose we have an $n \times p$ data matrix $X_n = (x_{ij})$, 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 coefficient matrix of X_n; that is, the entry ρ_{ij} is the usual Pearson correlation coefficient between the i-th column and j-th column of X_n. An advantage to working on the sample correlation matrix is that they are invariant under scaling and shifting. In other words, by shifting and scaling each column of a given data matrix, the new matrix generates the same sample correlation matrix as before. Therefore, if a population is $N_p(\mu, \Sigma)$ with μ and Σ unknown, under the null hypothesis that the p components are uncorrelated, the distribution of the sample correlation matrix is the same as that generated by a data matrix with independent standard normals as entries.

In an earlier work, the test statistic $\max_{1 \le i < j \le p} |\rho_{ij}|$ for testing $H_0 : R = I$ is studied by Jiang [9]. According to the PCA, the largest eigenvalue is a natural choice for a test statistic. However, it seems very difficult to obtain the asymptotic distribution of the largest eigenvalue $\lambda_{\max}$ of the sample correlation matrix R_n. In this paper, we obtain some properties of $\lambda_{\max}$ and the empirical distributions of eigenvalues of R_n en route to a good understanding of $\lambda_{\max}$.

We assume that both n and p are large. Traditionally, it is standard to assume that the dimension p is fixed and sample size n is large. For example, a book-length treatment of such data structures can be found in Anderson [1]. For modern data, Johnstone in [10] provides four examples in which the sample sizes and the dimensions of data are very large and comparable. In one case, the dimension is even larger than the sample size. Also, Donoho gives many examples of such type in [6].

Now we state our results formally. Discussions will follow later.

Let $X = (x_{ij})$ be an n by p data matrix, where x_{ij}'s are complex random variables. The p columns of X are denoted by $x_1, x_2, \cdots, x_p$, respectively. 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 $e = (1, 1, \cdots, 1)^T \in \mathbb{R}^n$. The Pearson correlation coefficient between x_i and x_j is

$$\rho_{ij} = \frac{(x_i - \bar{x}_i)^* (x_j - \bar{x}_j)}{\|x_i - \bar{x}_i\| \cdot \|x_j - \bar{x}_j\|}, \quad 1 \le i, j \le p, \tag{1.1}$$

where $\|\cdot\|$ is the usual Euclidean norm. Then, the p by p sample correlation matrix, denoted by R_X, is equal to (ρ_{ij}). Clearly,

$$R_X = Y^* Y, \quad \text{where} \quad Y = \left(\frac{x_1 - \bar{x}_1}{\|x_1 - \bar{x}_1\|}, \frac{x_2 - \bar{x}_2}{\|x_2 - \bar{x}_2\|}, \cdots, \frac{x_p - \bar{x}_p}{\|x_p - \bar{x}_p\|} \right). \tag{1.2}$$

Note that $\bar{x}_i$'s are close to zero by the Law of Large Numbers provided all the variables in the i-th column are i.i.d. with zero mean. For convenience, sometimes we look at a modified version of R_X:

$$\tilde{R}_X = \tilde{Y}^* \tilde{Y}, \quad \text{where } \tilde{Y} = \left(\frac{x_1}{\|x_1\|}, \frac{x_2}{\|x_2\|}, \cdots, \frac{x_p}{\|x_p\|} \right). \tag{1.3}$$

For an $n \times n$ symmetric matrix A, suppose it has eigenvalues $\mu_1, \mu_2, \cdots, \mu_n$. Let $F^A(x)$ be the empirical law of these eigenvalues. Precisely,

$$F^A(x) = \frac{1}{n} \sum_{i=1}^{n} I\{\mu_i \leq x\}, \; x \in \mathbb{R}.$$

Let $X = \{\xi, x_{ij};\ i \geq 1, j \geq 1\}$ be a double array of i.i.d. non-degenerate random variables. For any integer p_n, define $X_n = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p_n}$. For convenience, write

$$R_n = R_{X_n} \text{ and } \tilde{R}_n = \tilde{R}_{X_n}. \tag{1.4}$$

When there is no confusion, we simply write p for p_n to save notation.

Let $\lambda_{\max}(n)$ and $\tilde{\lambda}_{\max}(n)$ be the largest eigenvalues of R_n and $\tilde{R}_n$, respectively. Since X_n is an $n \times p$ matrix, $\tilde{R}_n$ is an $p \times p$ matrix. So if $p > n$, then the $p - n$ smallest eigenvalues of $\tilde{R}_n$ are equal to zero. Considering this, let

$$\tilde{\lambda}_{\min}(n) = \begin{cases} \text{the smallest eigenvalue of } \tilde{R}_n, & \text{if } p \leq n; \\ \text{the } (p-n+1)\text{-th smallest eigenvalues of } \tilde{R}_n, & \text{if } p > n. \end{cases} \tag{1.5}$$

First we study the largest and smallest eigenvalues.

THEOREM 1 *Suppose $n/p \to \gamma \in (0, \infty)$. The following is true.*

(i) If $E|\xi|^4 < \infty$, then $\lambda_{max}(n) \to (1 + \sqrt{\gamma^{-1}})^2$ a.s.;

(ii) If $E|\xi|^4 < \infty$ and $E\xi = 0$, then $\tilde{\lambda}_{max}(n) \to (1 + \sqrt{\gamma^{-1}})^2$ a.s. and $\tilde{\lambda}_{min}(n) \to (1 - \sqrt{\gamma^{-1}})^2$ a.s.

Similarly, let $\lambda_{\min}(n)$ be as in (1.5) when $\tilde{R}_n$ is replaced by R_n. It is reasonable to guess from the second part of (ii) that $\lambda_{\min}(n)$ also converges to $(1 - \sqrt{\gamma^{-1}})^2$ almost surely. It will be interesting to see how to prove this.

Second, the empirical distributions of R_n and $\tilde{R}_n$ converge weakly.

THEOREM 2 *Suppose $n/p \to \gamma \in (0, +\infty)$ and $E|\xi|^2 < \infty$. Then, almost surely, F^{R_n} converges weakly to a deterministic probability distribution with density function*

$$p_\gamma(x) = \begin{cases} \frac{\gamma}{2\pi x}\sqrt{(b-x)(x-a)}, & \textit{if } x \in [a, b]; \\ 0, & \textit{otherwise} \end{cases}$$

and a point mass $(1-\gamma)^+$ at $x=0$, where $a=(1-\gamma^{-1/2})^2$ and $b=(1+\gamma^{-1/2})^2$. The above conclusion is true for $F^{\tilde{R}_n}$ provided $E\xi=0$ in addition.

The distribution with probability density function $p_\gamma(x)$ is called the Marčenko-Pastur law.

When $p\to\infty$ and $n/p\to\infty$, the above law does not hold. Heuristically, assuming Theorem 2, the probability density function is 0 for any $x\neq 1$. This means that a dominated number of eigenvalues of R_n and $\tilde{R}_n$ are close to one. We then consider $T_n := (1/2)\sqrt{n/p}(R_n - I)$ and $\tilde{T}_n := (1/2)\sqrt{n/p}(\tilde{R}_n - I)$. We have that

THEOREM 3 *Suppose $E|\xi|^4<\infty$. Let p_n be such that $p_n/\sqrt{n}\to\infty$ and $p_n/n\to 0$. Then, almost surely F^{T_n} converges weakly to a distribution with density function*

$$p(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{if } |x|\le 1; \\ 0, & \text{otherwise.} \end{cases}$$

The above conclusion also holds for $F^{\tilde{T}_n}$ provided $E\xi=0$ in addition.

The distribution with probability density function $p(x)$ above is referred to as the semi-circular law.

Obviously, $E\exp(t_0|\xi|^2)<\infty$ for some $t_0=t_\xi>0$ for any Gaussian random variable $\xi\sim N(\mu,\sigma^2)$. For a population $N_p(\mu,\Sigma)$, a sample $X_n=(x_{ij};1\le i\le n, 1\le j\le p)$ is obtained to test the hypothesis that the p variates are independent. Note that, assuming the hypothesis is true, R_n is invariant under shifting and scaling. We then can assume, w.l.o.g., that x_{ij}'s are i.i.d. with the law of $N(0,1)$. Therefore, by the previous results, we have that

COROLLARY 1.1 *Suppose $\{x_{ij}; i\ge 1, j\ge 1\}$ are independent and $x_{ij}\sim N(\mu_j,\sigma_j^2)$ for some μ_j and $\sigma_j\neq 0$ for all i and j. Then Theorems 1, 2 and 3 also hold.*

Remark 1. In the above theorems, we assume that there is an infinite double array of i.i.d. random variables $\{x_{ij}; i,j\ge 1\}$. If we have a sequence of finite double arrays $\{\xi, x_{ij}^n; 1\le i\le n, 1\le j\le p\}$ such that these np random variables are i.i.d. for each n, then the above theorems are still valid provided some slightly stronger moment conditions are assumed. For example, Theorem 2 holds if $E|\xi|^3<\infty$.

Our main results show that the matrices R_n and $\tilde{R}_n$ behave very much as $X_n^*X_n/n$. This is the reason we believe that the distribution of largest eigenvalues of R_n and $\tilde{R}_n$ quite likely

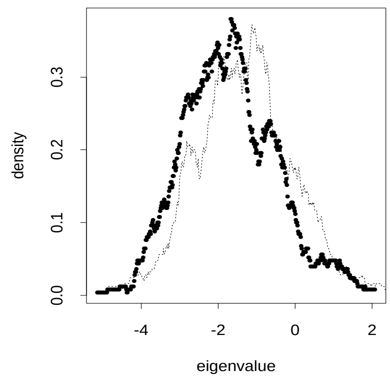

Figure 1: *The two curves formed by two different kinds of points are kernel estimations of empirical p.d.f.'s of largest eigenvalues of covariance and correlation matrices. The blacker one corresponds to sample correlation matrices. The other one corresponds to covariance matrices.*

converge to the Tracy-Widom law as does that of $X_n^* X_n/n$. We make some simulation for such case. See the explanation in the caption of Figure 1. Actually, a 500 random sample of 200 by 200 matrices with independent standard normals as entries are obtained. Normalize all largest eigenvalues of the corresponding sample covariance and correlation matrices by the formula $(\lambda_{\max} - 4n)/(16n)^{1/3}$ with $n = 200$. Kernel estimations of these two sets of 500 normalized values are presented in Figure 1. Observe that the two curves match quite well. Since such normalized largest eigenvalues of Wishart matrices follow roughly the Tracy-Widom law by Johnstone [10], Figure 1 seems to support our belief that the normalized largest eigenvalue of sample correlation matrices also satisfies the Tracy-Widom distribution asymptotically. But a rigorous mathematical proof has to be given to confirm this.

We use some known results about the sample covariance matrix $X_n^* X_n$ to prove our theorems. The idea is approximating the sample correlation matrix R_n by $X_n^* X_n/n$. This process is completed by using some random matrix tools such as rank inequality and difference inequality together with some matrix techniques. All the proofs are given in the next section.

2 Proofs

Looking at (1.4) and (1.3), we see that the correlation coefficient matrix R_n is invariant under shifting and scaling. Precisely, for any constant α_j and $\beta_j \neq 0$, let $z_{ij} = (x_{ij} - \alpha_j)/\beta_j$ for any i and j. Then the corresponding ρ_{ij} for $Z_n := (z_{ij})_{1 \le i \le n, 1 \le j \le p}$ is the same as that for X_n. Particularly, choosing $\alpha_j = E\xi$ and $\beta_j = (\text{var}(\xi))^{1/2}$, then $E(z_{ij}) = 0$ and $\text{var}(z_{ij}) = 1$

for any i and j. It is also easy to see that $\tilde{R}_n$ is invariant under scaling (shifting is not necessary because we assume $E\xi = 0$ in this case). So in the rest of the paper, without loss of generality, we assume that

$\{\xi, x_{ij};\ i,j \geq 1\}$ **are i.i.d. random variables with** $E\xi = 0$ **and var**$(\xi) = 1$.

Proof of Theorem 1. Let $\mu_{\max}(n)$ be the largest eigenvalue of $n^{-1}X_n^*X_n$. Let also $\mu_{\min}(n)$ be as in (1.5) when R_n is replaced by $n^{-1}X_n^*X_n$. Since $E|\xi|^4 < \infty$, by Theorem 2.16 from [4] (also Theorem 3.1 from [13] and Theorem 2 from [3]),

$$\lim_{n\to\infty} \mu_{\max}(n) = (1+\sqrt{\gamma^{-1}})^2 \ \ a.s. \ \text{ and } \ \lim_{n\to\infty} \mu_{\min}(n) = (1-\sqrt{\gamma^{-1}})^2 \ \ a.s. \tag{2.1}$$

To prove the theorem, it suffices to show that

$$\sqrt{\tilde{\lambda}_{\max}(n)} - \sqrt{\mu_{\max}(n)} \to 0 \ \ a.s. \ \text{ and } \sqrt{\tilde{\lambda}_{\min}(n)} - \sqrt{\mu_{\min}(n)} \to 0 \ \ a.s. \tag{2.2}$$

$$\sqrt{\lambda_{\max}(n)} - \sqrt{\mu_{\max}(n)} \to 0 \ \ a.s. \tag{2.3}$$

Recall that $X_n = (x_{ij})_{1\leq i\leq n, 1\leq j\leq p}$. For fixed n, for saving notation, let $x_1,\ x_2 \cdots, x_p$ be the p columns of X_n. Then $R_n = Y_n^*Y_n$ and $\tilde{R}_n = \tilde{Y}_n^*\tilde{Y}_n$, where Y_n is as Y in (1.4) and $\tilde{Y}_n$ is as $\tilde{Y}$ in (1.3) when $X = X_n$. Rewrite $\tilde{Y}_n = (X_n/\sqrt{n})\tilde{D}_n$ where

$$\tilde{D}_n = \operatorname{diag}\left(\frac{\sqrt{n}}{\|x_1\|}, \cdots, \frac{\sqrt{n}}{\|x_p\|}\right). \tag{2.4}$$

We first prove the first limit in (2.2). For any matrix C, let $\|C\|$ be the spectrum norm of the linear operator C. Of course, $\|C\|$ is equal to the squared root of the largest eigenvalue of C^*C. By the triangle inequality of the norm and that $\|C_1C_2\| \leq \|C_1\| \cdot \|C_2\|$ for any C_1 and C_2, we have that

$$\begin{aligned} |\sqrt{\tilde{\lambda}_{\max}(n)} - \sqrt{\mu_{\max}(n)}| &= \|\tilde{Y}_n - (X_n/\sqrt{n})\| \leq \|n^{-1/2}X_n(\tilde{D}_n - I)\| \\ &\leq \|n^{-1/2}X_n\| \cdot \|\tilde{D}_n - I\|. \end{aligned} \tag{2.5}$$

By Lemma 2 from [3], since $E|\xi|^4 < \infty$,

$$\max_{1\leq j\leq p}\left|\frac{\|x_j\|^2}{n} - 1\right| \to 0 \ \ a.s. \tag{2.6}$$

This implies $\|\tilde{D}_n - I\| = \max_{1\leq j\leq p}\left|(n^{1/2}/\|x_j\|) - 1\right| \to 0 \ \ a.s.$ Thus the first part of (2.2) follows from (2.1) and (2.5). The inequalities in (2.5) still hold if the "max" sign is replaced with "min" by Corollary 7.3.8 of Horn and Johnson [8]. So the second limit of (2.2) also follows from (2.1).

Now we turn to prove (2.3).

Recall $e = (1, 1, \cdots, 1)' \in \mathbb{R}^n$. Then $\bar{x}_j e$ is a column of which every entry is equal to $(1/n)\sum_{i=1}^n x_{ij}$. Set $\bar{X}_n = (\bar{x}_1 e, \bar{x}_2 e, \cdots, \bar{x}_p e)$. Rewrite $Y_n = ((X_n - \bar{X}_n)/\sqrt{n})D_n$, where

$$D_n = \text{diag}\left(\frac{\sqrt{n}}{\|x_1 - \bar{x}_1\|}, \cdots, \frac{\sqrt{n}}{\|x_p - \bar{x}_p\|}\right).$$

We claim that

$$\lim_{n\to\infty} \|n^{-1/2}(X_n - \bar{X}_n)\| = 1 + \gamma^{-1/2} \quad a.s. \tag{2.7}$$

Let $\nu_{\max}(n)$ be the largest eigenvalue of $(X_n - \bar{X}_n)^*(X_n - \bar{X}_n)/n$. If (2.7) is true, then by the same arguments as in (2.5), we have that

$$|\sqrt{\lambda_{\max}(n)} - \sqrt{\nu_{\max}(n)}| \le \|n^{-1/2}(X_n - \bar{X}_n)\| \cdot \max_{1\le j\le p}\left|\frac{n^{1/2}}{\|x_j - \bar{x}_j\|} - 1\right|. \tag{2.8}$$

Note that $\|x_j - \bar{x}_j\|^2 = \|x_j\|^2 - n|\bar{x}_j|^2$. It follows from Lemma 2 in [3] that

$$\max_{1\le j\le p}\left|\frac{\|x_j - \bar{x}_j\|^2}{n} - 1\right| \le \max_{1\le j\le p}\left|\frac{\sum_{i=1}^n |x_{ij}|^2}{n} - 1\right| + \max_{1\le j\le p}|\bar{x}_j|^2 \to 0 \quad a.s.$$

under $E|\xi|^4 < \infty$. The assertion (2.3) is then concluded from (2.7) and (2.8).

Now we prove claim (2.7). First, it is not difficult to see that $(X_n - \bar{X}_n)^*(X_n - \bar{X}_n) = X_n^* X_n - n\bar{X}_n^*\bar{X}_n$. Since both $X_n^*X_n$ and $\bar{X}_n^*\bar{X}_n$ are nonnegative definite matrix, by the definition of operator norm $\|\cdot\|$, it is easy to see that $\|X_n^*X_n - n\bar{X}_n^*\bar{X}_n\| \le \|X_n^*X_n\|$, we obtain from (2.1) that

$$\limsup_{n\to\infty} \|n^{-1/2}(X_n - \bar{X}_n)\| \le 1 + \gamma^{-1/2} \quad a.s. \tag{2.9}$$

For a symmetric matrix A, recall that $F^A(x)$ is the empirical distribution function of eigenvalues of A. Set $E_n = n^{-1/2}(X_n - \bar{X}_n)$. By Lemma 2.6 from [4],

$$\|F^{E_n^*E_n} - F^{X_n^*X_n/n}\| \le \frac{2}{p}\text{rank}(\bar{X}_n) \le \frac{2}{p} \to 0$$

as $n \to \infty$, where $\|f\| := \sup_{x\in\mathbb{R}}|f(x)|$ for a function $f(x)$ defined on $\mathbb{R}$. Thus, by Theorem 2.5 from Bai [4], we know that under the condition $E\xi = 0$ and $E\xi^2 = 1$

$$F^{E_n^*E_n} \text{ converges weakly to a distribution function } F \tag{2.10}$$

with $F'(x) = p_\gamma(x)$ as in Theorem 2. On the other hand,

$$\|E_n\|^2 = \|E_n^*E_n\| \ge \left(\frac{1}{p}tr\left((E_n^*E_n)^k\right)\right)^{1/k} = \left(\int_0^\infty x^k F^{E_n^*E_n}(dx)\right)^{1/k}$$

for any $k \geq 1$. By Fatou's lemma and (2.10), we have that

$$\liminf_{n\to\infty} n^{-1/2}\|X_n - \bar{X}_n\| = \liminf_{n\to\infty} \|E_n\| \geq \left(\int_a^b x^k p_\gamma(x)\,dx\right)^{1/(2k)} \quad a.s.$$

for any $k \geq 1$. Note that the right end of the support of $p_\gamma(x)$ is b. Let $k \uparrow \infty$. We obtain that

$$\liminf_{n\to\infty} \|n^{-1/2}(X_n - \bar{X}_n)\| \geq 1 + \gamma^{-1/2} \quad a.s.$$

This together with (2.9) proves (2.7). ■

We need the following lemma to prove Theorem 2.

LEMMA 2.1 *Let $A = (a_{ij})$ be an n by p complex random matrix with a_j as its j-th column. Define $B = (a_1/\|a_1\|, a_2/\|a_2\|, \cdots, a_p/\|a_p\|)$. Then*

$$\frac{1}{p} tr\left(\left(\frac{A}{\sqrt{n}} - B\right)^* \left(\frac{A}{\sqrt{n}} - B\right)\right) = b_1 - 2b_2,$$

where

$$b_1 = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}(|a_{ij}|^2 - 1) \text{ and } b_2 = \frac{1}{p}\sum_{j=1}^{p}\left(\frac{\|a_j\|}{\sqrt{n}} - 1\right).$$

Proof. Let v_j be the j-th column of $n^{-1/2}A - B$. Then $v_j = (\|a_j\| - \sqrt{n})\, a_j/(\sqrt{n}\|a_j\|)$ for $j = 1, 2, \cdots, p$. Thus,

$$\begin{aligned} \mathrm{tr}\left(\left(n^{-1/2}A - B\right)^* \left(n^{-1/2}A - B\right)\right) &= \sum_{j=1}^{p}\|v_j\|^2 = \frac{1}{n}\sum_{j=1}^{p}(\|a_j\| - \sqrt{n})^2 \\ &= \frac{1}{n}\sum_{j=1}^{p}\|a_j\|^2 - 2\sum_{j=1}^{p}\frac{\|a_j\|}{\sqrt{n}} + p. \end{aligned}$$

Note that $\sum_j \|a_j\|^2 = \sum_{i,j}|a_{ij}|^2$. Then the conclusions follows. ■

Proof of Theorem 2. By Theorem 2.5 in Bai [4] (the real case is obtained by Yin [12]), the conclusion of Theorem 2 holds for $F^{X_n^* X_n/n}$ under the condition that $E\xi = 0$ and $E|\xi|^2 = 1$. Again, let $E_n = n^{-1/2}(X_n - \bar{X}_n)$ as in the proof of Theorem 1, Y_n and $\tilde{Y}_n$ be as in (1.4) and (1.3), respectively, where $X = X_n$. Recall $R_n = Y_n^* Y_n$. Let $L(\cdot, \cdot)$ be the Levy distance; see e.g., exercise 2.15 on page 91 from [7]. By the triangle inequality, we have that

$$L(F^{R_n}, F^{X_n^* X_n/n}) \leq L(F^{Y_n^* Y_n}, F^{E_n^* E_n}) + L(F^{E_n^* E_n}, F^{X_n^* X_n/n}).$$

By (2.10), to prove the theorem for R_n, we need to show $L(F^{Y_n^*Y_n}, F^{E_n^*E_n}) \to 0$ *a.s.* Obviously, $tr(Y_n^*Y_n) = p$ and $(X_n - \bar{X}_n)^*(X_n - \bar{X}_n) = X_n^*X_n - n\bar{X}_n^*\bar{X}_n$. Thus

$$\frac{1}{np} tr\left((X_n - \bar{X}_n)^*(X_n - \bar{X}_n)\right) \le \frac{1}{np} tr\left(X_n^*X_n\right) = \frac{1}{np}\sum_{j=1}^{p}\sum_{i=1}^{n}|x_{ij}|^2 \to E|\xi|^2 = 1 \quad a.s.$$

by the law of large numbers. Therefore, to prove the theorem for R_n, by the difference inequality Lemma 2.7 from [4], it is enough to show

$$\frac{1}{p}\mathrm{tr}\left((E_n - Y_n)^*(E_n - Y_n)\right) \to 0 \quad a.s. \tag{2.11}$$

Similarly but easily, to prove the part for $\tilde{R}_n$, it suffices to prove that

$$\frac{1}{p}\mathrm{tr}\left((n^{-1/2}X_n - \tilde{Y}_n)^*(n^{-1/2}X_n - \tilde{Y}_n)\right) \to 0 \quad a.s. \tag{2.12}$$

To do so, by Lemma 2.1, we need to verify that the corresponding b_1 and b_2 go to zero. To avoid confusion, let b_i correspond to (2.11) and $\tilde{b}_i$ correspond to (2.12) for $i = 1, 2$. We prove them by distinguishing these two cases.

(i) *The proof of (2.12).* Evidently,

$$\tilde{b}_1 = \frac{1}{np}\sum_{i=1}^{p}\sum_{j=1}^{n}(|x_{ij}|^2 - 1) \to 0 \quad a.s.$$

by the law of large numbers since $E|\xi|^2 = 1$. To show $\tilde{b}_2 \to 0$, it suffices to show that

$$I_p := \frac{1}{p}\sum_{j=1}^{p}\sqrt{\frac{\sum_{i=1}^{n}|x_{ij}|^2}{n}} \to 1 \quad a.s. \tag{2.13}$$

Since the function $\sqrt{x}$ is concave over $[0, \infty]$,

$$I_p \le \sqrt{\frac{\sum_{i,j}|x_{ij}|^2}{np}} \to 1 \quad a.s. \tag{2.14}$$

by the law of large numbers again. On the other hand,

$$\frac{1}{p}\sum_{j=1}^{p}\sqrt{\frac{\sum_{i=1}^{n}|x_{ij}|^2}{n}} \ge (\min_{1\le j\le p} u_{n,j})^{1/2} \tag{2.15}$$

for any $C \ge 1$, where $u_{n,j} = \sum_{i=1}^{n}|x_{ij}|^2 I\{|x_{ij}| \le C\}/n$. By Lemma 2 from [3], $\min_{1\le j\le p} u_{nj} \to Ex_{11}^2 I\{|x_{11}| \le C\}$ as $n \to \infty$. Letting $n \to \infty$ for both sides of (2.15), followed by letting $C \uparrow +\infty$, we obtain that $\liminf_{n\to\infty} I_p \ge 1$ *a.s.* This and (2.14) yield (2.13).

(ii) *The proof of (2.11).* First,

$$b_1 = \frac{1}{np}\sum_{j=1}^{p}\sum_{i=1}^{n}|x_{ij}-\bar{x}_j|^2 - 1 \quad \text{and} \quad b_2 = \frac{1}{p}\sum_{j=1}^{p}\sqrt{(1/n)\sum_{j=1}^{n}|x_{ij}-\bar{x}_j|^2} - 1.$$

It is easy to check that

$$|b_1 - \tilde{b}_1| = \frac{1}{p}\sum_{j=1}^{p}|\bar{x}_j|^2 \leq \left(\max_{1\leq j\leq p}|\bar{x}_j|\right)^2 \to 0 \quad a.s. \tag{2.16}$$

under $E|\xi|^2 < \infty$ by Lemma 2 from [3] again. Thus $b_1 \to 0$ $a.s.$ since $\tilde{b}_1 \to 0$ $a.s.$ Second, using that $|\sqrt{x} - \sqrt{y}| \leq |x-y|/\sqrt{y}$, we obtain that

$$|b_2 - \tilde{b}_2| \leq \frac{1}{p}\sum_{j=1}^{p}|\bar{x}_j|^2\left(\frac{1}{n}\sum_{i=1}^{n}|x_{ij}|^2\right)^{-1/2} \leq \left(\max_{1\leq j\leq p}|\bar{x}_j|\right)^2 \cdot (\min_{1\leq j\leq p} u_{n,j})^{-1/2}$$

for any $C > 0$ such that $E|\xi|^2 I\{|\xi| \leq C\} > 1/2$. We already shown that $\min_{1\leq j\leq p} u_{nj} \to Ex_{11}^2 I\{|x_{11}| \leq C\}$ as $n \to \infty$. Then $b_2 - \tilde{b}_2 \to 0$ by (2.16). It follows that $b_2 \to 0$ $a.s.$ since $\tilde{b}_2 \to 0$ $a.s.$ ■

Proof of Theorem 3. Recall the definitions in (1.2), (1.3) and (1.4). We prove the conclusion for F^{T_n} and $F^{\tilde{T}_n}$ separately. First, we consider $F^{\tilde{T}_n}$.

(a) Recall $\tilde{T}_n = (1/2)\sqrt{n/p}(\tilde{R}_n - I)$. Let $S_n = X_n^* X_n/n$. Note that $\tilde{R}_n = \tilde{D}_n^* S_n \tilde{D}_n$, where $\tilde{D}_n$ is as in (2.4). For any $\epsilon > 0$, define an $p \times p$ diagonal matrix $\hat{D}_n = \text{diag}(\hat{d}_i)$, where

$$\hat{d}_i = \begin{cases} \sqrt{n}/\|x_i\|, & \text{if } |\sqrt{n}/\|x_i\| - 1| \leq \epsilon\sqrt{p/n}; \\ 1, & \text{otherwise.} \end{cases} \tag{2.17}$$

Let $\tilde{U}_n = (1/2)\sqrt{n/p}(\hat{D}_n^* S_n \hat{D}_n - I)$ and $Q_n = (1/2)\sqrt{n/p}(S_n - I)$. First,

$$L(F^{\tilde{T}_n}, F^{Q_n}) \leq L\left(F^{\tilde{T}_n}, F^{\tilde{U}_n}\right) + L\left(F^{\tilde{U}_n}, F^{Q_n}\right). \tag{2.18}$$

For an $p \times p$ symmetric matrix A, denote by $\lambda_k(A)$ the k-th smallest eigenvalue of A. By Courant-Fischer theorem (see p.179 from [8]),

$$\lambda_k(\tilde{H}_n^* S_n \hat{D}_n) = \inf_{w_1,\cdots,w_{p-k}} \max_{0\neq x\perp w_1,\cdots,w_{p-k}} \frac{(\hat{D}_n x)^* S_n (\hat{D}_n x)}{(\hat{D}_n x)^*(\hat{D}_n x)} \cdot \frac{(\hat{D}_n x)^*(\hat{D}_n x)}{x^* x}.$$

Trivially, $\hat{D}_n$ is invertible almost surely for sufficiently large n. It follows that

$$(\lambda_1(\hat{D}_n))^2 \lambda_k(S_n) \leq \lambda_k(\hat{D}_n^* S_n \hat{D}_n) \leq (\lambda_p(\hat{D}_n))^2 \lambda_k(S_n)$$

for any $k \geq 1$. Since $\hat{D}_n$ is diagonal, $\max_{1\leq j\leq p} |(\lambda_j(\hat{D}_n))^2 - 1| = \max_{1\leq j\leq p} |\hat{d}_j^2 - 1|$. Also, since $p/n \to 0$, $|\hat{d}_j^2 - 1| \leq 3|\hat{d}_j - 1| \leq 3\epsilon\sqrt{p/n}$ as n is sufficiently large. Hence

$$\sup_{1\leq k\leq p} |\lambda_k(\hat{D}_n^* S_n \hat{D}_n) - \lambda_k(S_n)| \leq \max_{1\leq j\leq p} |\hat{d}_j^2 - 1| \cdot \lambda_p(S_n) \leq 3\epsilon\sqrt{\frac{p}{n}} \cdot \lambda_p(S_n)$$

as n is sufficiently large. Let $\tilde{X}_n = (x_{ij})_{1\leq i,j\leq n}$. Since $p/n \to 0$, by the interlacing theorem (see p.189 on [8]) and (2.1), $\lambda_p(S_n) \leq \lambda_n(\tilde{X}_n^* \tilde{X}_n/n) \to 4$ $a.s.$ as $n \to \infty$. By the inequality in line 11 on p.615 from [4],

$$L^3\left(F^{\tilde{U}_n}, F^{Q_n}\right) \leq \frac{1}{p}\left(\frac{1}{2}\sqrt{\frac{n}{p}}\right)^2 \sum_{k=1}^{p} |\lambda_k(\hat{D}_n^* S_n \hat{D}_n) - \lambda_k(S_n)|^2 \leq 60\epsilon^2 \tag{2.19}$$

as n is sufficiently large. Now, by the rank inequality Lemma 2.6 from [4],

$$\begin{aligned} L\left(F^{\tilde{T}_n}, F^{\tilde{U}_n}\right) \leq \frac{1}{p}\text{rank}\left(X_n\tilde{D}_n - X_n\hat{D}_n\right) &\leq \frac{1}{p}\text{rank}(\tilde{D}_n - \hat{D}_n) \\ &\leq \frac{1}{p}\sum_{j=1}^{p} I\left(|\frac{\sqrt{n}}{\|x_j\|} - 1| > \epsilon\sqrt{\frac{p}{n}}\right). \end{aligned}$$

Trivially, $\{x : |x-1| > \delta\} \subset \{x : |x^{-1} - 1| > \delta/2\}$ once $\delta < 1$. Also, $|\sqrt{x} - 1| \leq |x-1|$. Thus

$$L\left(F^{\tilde{T}_n}, F^{\tilde{U}_n}\right) \leq \frac{1}{p}\sum_{j=1}^{p} I\left(|\|x_j\|^2 - n| > \frac{\sqrt{np}\,\epsilon}{2}\right) \tag{2.20}$$

as n is sufficiently large. Recall $\|x_j\|^2 = \sum_{i=1}^n x_{ij}^2$. Since $E|x_{11}|^4 < \infty$, by the Chebyshev inequality, $\alpha := P(|\|x_1\|^2 - n| > \sqrt{np}\,\epsilon/2) = O(p^{-1}) \to 0$ as $n \to \infty$. By the Bernstein inequality (see, e.g., Problem 14 on p.111 from [5]) and (2.20),

$$P(L(F^{\tilde{T}_n}, F^{\tilde{U}_n}) \geq 2\epsilon) \leq 2\exp\left(\frac{-p^2\epsilon^2}{2(p\epsilon + p\alpha(1-\alpha))}\right) \leq 2e^{-p\epsilon/3}$$

as n is sufficiently large. By the Borel-Cantelli lemma, $L\left(F^{\tilde{T}_n}, F^{\tilde{U}_n}\right) \to 0$ $a.s.$ as $n \to \infty$. This together with (2.18) and (2.19) implies that $\limsup_{n\to\infty} L(F^{\tilde{T}_n}, F^{Q_n}) \leq 4\epsilon^{2/3}$ $a.s.$ Letting $\epsilon \downarrow 0$, the desired conclusion follows by the theorem from [2] and Theorem 2.9 from [4], which says that Theorem 3 holds if $\tilde{R}_n$ in the definition of $\tilde{T}_n$ is replaced by S_n under $E|\xi|^4 < \infty$.

(b) Let $\hat{Y}_n = (x_1/\|x_1 - \bar{x}_1\|, \cdots, x_p/\|x_p - \bar{x}_p\|)$. Then by the rank inequality again (Lemma 2.6 from Bai [4]),

$$L(F^{(1/2)\sqrt{n/p}(\hat{Y}_n^*\hat{Y}_n - I)}, F^{T_n}) \leq \frac{1}{p}\text{rank}((\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)) \leq \frac{1}{p} \to 0$$

as $n \to \infty$. So to prove this part, one only needs to show that the conclusion of Theorem 3 holds for $F^{(1/2)\sqrt{n/p}(\hat{Y}_n^* \hat{Y}_n - I)}$. This is done through replacing $\|x_i\|$ in (2.17) by $\|x_i - \bar{x}_i\|$ and following the rest arguments in part (a). The only change is to show that $\alpha := P(|\|x_1 - \bar{x}_1\|^2 - n| \geq \sqrt{np}\epsilon/2) \to 0$. Actually, since $\|x_1 - \bar{x}_1\|^2 - n = \sum_{j=1}^n (|x_{1j}|^2 - 1) - n|\bar{x}_1|^2$,

$$\begin{aligned} \alpha &\leq \frac{4}{np\epsilon^2} \text{Var}(\sum_{j=1}^n (|x_{1j}|^2 - 1) - n|\bar{x}_1|^2) \\ &\leq \frac{8}{np\epsilon^2} \{\text{Var}(\sum_{j=1}^n (|x_{1j}|^2 - 1)) + \text{Var}\left(n|\bar{x}_1|^2\right)\} = O(p^{-1}), \end{aligned}$$

as $n \to \infty$, where the fact that $E|\sum_{j=1}^n x_{1j}|^4 = O(n^2)$ is used in the last step. ■

Acknowledgment. The author thanks Zhidong Bai for a help in sharpening the proof. The author also thanks Christopher Bingham and Jianhong Shen for inspiring discussions.

References

[1] Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Second edition.

[2] Bai, Z. D. and Yin, Y.Q. (1988). Convergence to the semicircle law. Ann. Probab. 16, 863-875.

[3] Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Ann. Probab. 21 1275-1294.

[4] Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. Statistica Sinica. Vol. 9, 611-677.

[5] Chow, Y.S. and Teicher, H. (1997). Probability Theory, *Independence, Interchangeability, Martingales.* Springer, New York, Third edition.

[6] Donoho, D. High-dimensional data analysis: The curses and blessings of dimensionality. Technical report of Department of Statistics, Stanford University.

[7] Durrett, R. (1995). Probability: Theory and Examples. The Duxbury Press, Second edition.

[8] Horn, R. N. and Johnson, C. R. (1999). Matrix Analysis. Cambridge University Press.

[9] Jiang, T. (2002). The asymptotic distributions of the largest entries of sample correlation matrices. To appear in "Annals of Applied Probability".

[10] Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. Ann. Stat. 29, 295-327.

[11] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in a certain classes of sample covariance matrices. J. Stat. Phys., Vol. 108, 1033-1056.

[12] Yin, Y.Q. (1986). LSD' for a class of random matrices. J. Multivariate Anal. 20, 50-68.

[13] Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. Probab. Theory Related Fields. 78, 509-521.

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