Abstract

We prove that Kendall’s Rank correlation matrix converges to the Marčenko Pastur law, under the assumption that observations are i.i.d random vectors $X_1, \ldots, X_n$ with components that are independent and absolutely continuous with respect to the Lebesgue measure. This is the first result on the empirical spectral distribution of a multivariate $U$-statistic.

1 Introduction

Estimating the association between two random variables $X, Y \in \mathbb{R}$ is a central statistical problem. As such many methods have been proposed, most notably Pearson’s correlation coefficient. While this measure of association is well suited to the Gaussian case, it may be inaccurate in other cases. This observation has led statisticians to consider other measures of associations such as Spearman’s $\rho$ and Kendall’s $\tau$ that can be proved to be more robust to heavy-tailed distributions (see, e.g., [LHY+12]). In a multivariate setting, covariance and correlation matrices are preponderant tools to understand the interaction between variables. They are also used as building blocks for more sophisticated statistical questions such as principal component analysis or graphical models.

The past decade has witnessed an unprecedented and fertile interaction between random matrix theory and high-dimensional statistics (see [PA14] for a recent survey). Indeed, in high-dimensional settings, traditional asymptotics where the sample size tends to infinity fail to capture a delicate interaction between sample size and dimension and random matrix theory has allowed statisticians and practitioners alike to gain valuable insight on a variety of multivariate problems.

The terminology “Wishart matrices” is often, though sometimes abusively, used to refer to $p \times p$ random matrices of the form $X^\top X/n$, where $X$ is an $n \times p$ random matrix with independent rows (throughout this paper we restrict our attention to real random matrices). The simplest example arises where $X$ has i.i.d standard Gaussian entries but the main characteristics are shared by a much wider class of random matrices. This universality phenomenon manifests itself in various aspects of the limit distribution, and in particular in the limiting behavior of the empirical spectral distribution of the matrix. Let $W = X^\top X/n$ be a $p \times p$ Wishart matrix and denote by $\lambda_1, \ldots, \lambda_p$ its eigenvalues; then the empirical spectral distribution $\hat{\mu}_p$ of $W$ is the distribution on $\mathbb{R}$ defined as the following mixture of Dirac point masses at the $\lambda_j$s:

$$\hat{\mu}_p = \frac{1}{p} \sum_{k=1}^p \delta_{\lambda_k}.$$  

Assuming that the entries of $X$ are independent, centered and of unit variance, it can be shown that $\mu_p$ converges weakly to the Marčenko-Pastur distribution under weak moment conditions (see [EKYY12] for the weakest condition).

While this development alone has led to important statistical advances, it fails to capture more refined notions of correlations, notably more robust ones involving ranks and therefore dependent observations. A first step in this direction was made by [YK86], where the matrix $X$ is assumed to have independent rows

---

*NYU, Courant Institute of Mathematical Sciences: bandeira@cims.nyu.edu
†MIT, Department of Mathematics: lodhia@math.mit.edu
‡MIT, Department of Mathematics: rigollet@math.mit.edu
with isotropic distribution. More recently, this result was extended in [BZ08, O’R12] and covers for example the case of Spearman’s \( \rho \) matrix that is based on ranks, which is also a Wishart matrix of the form \( X^\top X/n \).

The main contribution of this paper is to derive the limiting distribution of Kendall’s \( \tau \) matrix, a cousin of Spearman’s \( \rho \) matrix but which is not of the Wishart type but rather a matrix whose entries are \( U \)-statistics. Kendall’s \( \tau \) matrix is a very popular surrogate for correlation matrices but an understanding the fluctuations of its eigenvalues is still missing. Interestingly, Marčenko-Pastur results have been used as heuristics, without justification, precisely for Kendall’s \( \tau \) in the context of certain financial applications [CCL+15].

As it turns out, the limiting distribution of \( \hat{\mu}_p \) is not exactly Marčenko-Pastur, but rather an affine transformation of it. Our main theorem below gives the precise form of this transformation.

**Theorem 1.** Let \( X_1, \ldots, X_n \), be \( n \) independent random vectors in \( \mathbb{R}^p \) whose components \( X_i(k) \) are independent random variables that have a density with respect to the Lebesgue measure on \( \mathbb{R} \). Then as \( n \to \infty \) and \( \frac{p}{n} \to \gamma > 0 \) the empirical spectral distribution of \( \tau \) converges in probability to \[ \frac{2}{3} Y + \frac{1}{3}, \]
where \( Y \) is distributed according to the standard Marčenko-Pastur law with parameter \( \gamma \) (see Theorem 4 for the appropriate definition).

Figure 1 illustrates numerically the result of Theorem 1.
2 Kendall’s Tau

The (univariate) Kendall τ statistic [Ess24, Lin25, Lin29, Ken38] is defined as follows. Let \((Y_1, Z_1), \ldots, (Y_n, Z_n)\) be \(n\) independent samples of a pair \((Y, Z) \in \mathbb{R} \times \mathbb{R}\) of real-valued random variables. Then the (empirical) Kendall τ between \(Y\) and \(Z\) is defined as

\[
\tau(Y, Z) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}(Y_i - Y_j) \cdot \text{sign}(Z_i - Z_j).
\]

The statistic \(\tau\) takes values in \([-1, 1]\) and it is not hard to see that it can be expressed as

\[
\tau = \frac{1}{\binom{n}{2}} \left( \# \{\text{concordant pairs}\} - \# \{\text{discordant pairs}\} \right),
\]

where a pair \((i, j)\) is said to be concordant if \(Y_i - Y_j\) and \(Z_i - Z_j\) have the same sign and discordant otherwise.

It is known that the Kendall \(\tau\) statistic is asymptotically Gaussian (see, e.g., [Ken38]). Specifically, if \(Y\) and \(Z\) are independent, then as \(n \to \infty\),

\[
\sqrt{n} \tau(Y, Z) \xrightarrow{d} \mathcal{N} \left(0, \frac{4}{9}\right).
\]

This property has been central to construct independence tests between two random variables \(X\) and \(Y\) (see, e.g., [KG90]).

Kendall’s \(\tau\) statistic can be extended to the multivariate case. Let \(X_1, \ldots, X_n\), be \(n\) independent copies of a random vector \(X \in \mathbb{R}^p\), with independent coordinates \(X(1), \ldots, X(p)\). The (empirical) Kendall \(\tau\) matrix of \(X\) is defined to be the \(p \times p\) matrix whose entries \(\tau_{kl}\) are given by

\[
\tau_{kl} := \tau(X(k), X(l)) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign} \left(X_i(k) - X_j(k)\right) \cdot \text{sign} \left(X_i(l) - X_j(l)\right) \quad 1 \leq k, l \leq p.
\]

Note that the \(\tau\) can be written as the sum of \(\binom{n}{2}\) rank-one random matrices:

\[
\tau = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign} \left(X_i - X_j\right) \otimes \text{sign} \left(X_i - X_j\right),
\]

where the sign function is taken entrywise.

It is easy to see that \(\tau_{ii} = 1\) for all \(i\). Together with (1), it implies that the matrix

\[
\bar{\tau} = \frac{3}{2} \tau - \frac{1}{2} I_p
\]

is such that \(\mathbb{E}[\sqrt{n}\tau] \to I_p\) and \(\text{Var}[\tau_{ij}] \to \mathbb{I}\{i \neq j\}\) as \(n \to \infty\). This suggests that if the empirical spectral distribution of \(\tau\) converges to a Marčenko-Pastur distribution, it should be a standard Marčenko-Pastur distribution. This heuristic argument supports the affine transformation arising in Theorem 1. However, the matrix \(\tau\) is not Wishart and the Marčenko-Pastur limit distribution does not follow from standard arguments. Nevertheless, Kendall’s \(\tau\) is a U-statistic which are known to satisfy the weakest form of universality, namely a Central Limit Theorem under general conditions [Hoe48, dIPG99]. In this paper, we show that in the case of the Kendall \(\tau\) matrix, this universality phenomenon extends to the empirical spectral distribution.

3 Proof of Theorem 1

For any pair \((i, j)\) such that \(1 \leq i, j \leq n\), let \(A_{(i,j)}\) be a rank-one matrix defined by

\[
A_{(i,j)} = \text{sign} \left(X_i - X_j\right) \otimes \text{sign} \left(X_i - X_j\right),
\]

and recall from (3) that

\[
\tau = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} A_{(i,j)}.
\]
Akin to most asymptotic results on U-statistics, we utilize a variant of Hoeffding’s (a.k.a. Efron-Stein, a.k.a ANOVA) decomposition [Hoe48]:

\[
\mathbf{A}_{(i,j)} = \mathbf{\bar{A}}_{(i,j)} + \mathbf{\bar{A}}_{(.,j)} + \mathbf{I}_p \tag{4}
\]

where

\[
\mathbf{\bar{A}}_{(i,\cdot)} = \mathbb{E}[\mathbf{A}_{(i,j)} - \mathbf{I}_p | X_i] , \quad \mathbf{\bar{A}}_{(\cdot,j)} = \mathbb{E}[\mathbf{A}_{(i,j)} - \mathbf{I}_p | X_j],
\]

and \(\mathbf{\bar{A}}_{(i,j)} = \mathbf{A}_{(i,j)} - \mathbf{\bar{A}}_{(.,j)} - \mathbf{\bar{A}}_{(i,\cdot)} - \mathbf{I}_p\). It is easy to check that each of first three matrices in the right-hand side of (4) is centered and that the four matrices are orthogonal to each other with respect to the inner product \(\mathbb{E} \text{Tr}(A^T B)\) as long as \(i \neq j\).

**Lemma 2.** For \(k \in [p]\), let \(F_k\) denote the cumulative distribution function of \(X(k)\). Fix \(i \in [n]\) and let \(U_i \in \mathbb{R}^p\) is a random vector with \(k\)th coordinate given by \(U_i(k) = F_k(X_i(k)) \sim \text{Unif}([0,1])\). Then

\[
\mathbf{\bar{A}}_{(i,\cdot)} = \mathbf{\bar{A}}_{(\cdot,i)} = \mathbb{D}_0((2U_i - 1_p) \otimes (2U_i - 1_p)) .
\]

**Proof.** For any \(i \in [n]\), observe that since the components of \(X\) have a density, then \(\mathbb{P}(X_i(k) = X_j(k) | X_i) = 0\) so that

\[
\mathbb{E}[\text{sign}(X_i(k) - X_j(k)) | X_i] = \mathbb{P}(X_i(k) > X_j(k)) | X_i) - \mathbb{P}(X_i(k) < X_j(k)) | X_i) = 2F_k(X_i(k)) - 1 .
\]

Therefore, since the coordinates of \(X\) are independent, we have

\[
\mathbf{\bar{A}}_{(i,\cdot)} = \mathbb{E}[\text{sign}(X_i - X_j) \otimes \text{sign}(X_i - X_j) | X_i] - \mathbf{I}_p
\]

\[
= \mathbb{D}_0(\mathbb{E}[\text{sign}(X_i - X_j) | X_i] \otimes \mathbb{E}[\text{sign}(X_i - X_j) | X_i])
\]

\[
= \mathbb{D}_0((2U_i - 1_p) \otimes (2U_i - 1_p))
\]

\(\square\)

By symmetry, we have \(\mathbf{\bar{A}}_{(\cdot,j)} = \mathbf{\bar{A}}_{(j,\cdot)}\). Thus

\[
2 \sum_{1 \leq i < j \leq n} \{\mathbf{\bar{A}}_{(i,j)} + \mathbf{\bar{A}}_{(j,i)}\} = \sum_{1 \leq i,j \leq n} \{\mathbf{\bar{A}}_{(i,j)} + \mathbf{\bar{A}}_{(j,i)}\} - \sum_{i=1}^n \{\mathbf{\bar{A}}_{(i,i)} + \mathbf{\bar{A}}_{(\cdot,i)}\} = 2(n-1) \sum_{i=1}^n \mathbf{\bar{A}}_{(i,i)} .
\]

Together with Lemma 2, this yields

\[
\frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \{\mathbf{\bar{A}}_{(i,j)} + \mathbf{\bar{A}}_{(j,i)}\} = \frac{2}{n} \sum_{i=1}^n \mathbb{D}_0((2U_i - 1_p) \otimes (2U_i - 1_p)) . \tag{6}
\]

Next, note that, the coordinates of each \(U_i, i = 1, \ldots, n\) are mutually independent so that \(\mathbb{E}[2U_i - 1_p] = 0\) and

\[
\mathbb{E}[(2U_i - 1_p) \otimes (2U_i - 1_p)] = \mathbb{E}[(2T - 1)^2]I_p = \frac{1}{3}I_p , \tag{7}
\]

where \(T \sim \text{Unif}([0,1])\). Thus Theorem 4 implies as \(n \to \infty\) and \(\frac{p}{n} \to \gamma > 0\), the empirical spectral distribution of

\[
\frac{2}{n} \sum_{i=1}^n (2U_i - 1_p) \otimes (2U_i - 1_p)
\]

converges in probability to \((2/3)Y\), where \(Y\) is distributed according to the standard Marčenko-Pastur law with parameter \(\gamma\). Moreover,

\[
\left\| \frac{2}{n} \sum_{i=1}^n \text{diag}((2U_i - 1_p) \otimes (2U_i - 1_p)) - \frac{2}{3}I_p \right\|_\text{op} = 2 \max_{1 \leq k \leq p} \left| \frac{1}{n} \sum_{i=1}^n ((2U_i(k) - 1)^2 - \mathbb{E}[(2U_i(k) - 1)^2]) \right| \to 0
\]

in probability as \(n \to \infty\) and \(\frac{p}{n} \to \gamma > 0\) by Chebyshev’s inequality. Together with (6), it yields the following result.
Proposition 3. As \( n \to \infty \) and \( \frac{1}{n} \to \gamma > 0 \), the empirical spectral distribution \( \tilde{\mu}_p \) of

\[
\frac{1}{(2)} \sum_{1 \leq i < j \leq n} \left\{ \tilde{A}_{(i, \cdot)} + \tilde{A}_{(\cdot, j)} + I_p \right\}
\]

converges in probability to \( \frac{2}{3} Y + \frac{1}{3} \), where \( Y \) is distributed according to the standard Marčenko-Pastur law with parameter \( \gamma \).

Let \( \tilde{\mu}_p \) denote the empirical spectral distribution of \( \tau \). We are going to show that the Levy distance between \( \tilde{\mu}_p \) and \( \tilde{\mu}_p \) converges to zero which implies Theorem 1 by Proposition 3. To that end, observe that by (4) and Lemma 5 convergence of the Levy distance follows if we show that

\[
\frac{1}{p} \mathbb{E} \parallel \tau - \frac{1}{(2)} \sum_{1 \leq i < j \leq n} \{ \tilde{A}_{(i, \cdot)} + \tilde{A}_{(\cdot, j)} + I_p \} \parallel^2_F = \frac{1}{p} \mathbb{E} \parallel \frac{1}{(2)} \sum_{1 \leq i < j \leq n} \tilde{A}_{(i, j)} \parallel^2 \to 0.
\]

Moreover, by orthogonality of Hoeffding’s decomposition, we have \( \mathbb{E} \parallel \tilde{A}_{(i, j)} \parallel^2_F \leq \mathbb{E} \parallel A_{(i, j)} \parallel^2_F = p^2 \). Therefore

\[
\mathbb{E} \parallel \frac{1}{(2)} \sum_{1 \leq i < j \leq n} \tilde{A}_{(i, j)} \parallel^2_F \leq p^2 \mathcal{C} \frac{1}{n^2},
\]

for some constant \( C > 0 \) so that (8) holds. This concludes the proof of Theorem 1.

Acknowledgements: The authors would like to thank Alice Guionnet for helpful comments and discussions. A. S. B. was supported by Grant NSF-DMS-1541100. Part of this work was done while A. S. B. was at the Department of Mathematics at MIT. A. L. was supported by Grant NSF-MS-1307704. P.R was supported by Grants NSF-DMS-1541099, NSF-DMS-1541100 and DARPA-BAA-16-46 and a grant from the MIT NEC Corporation.

A The standard Marčenko-Pastur law

We include here the definition of the standard Marčenko-Pastur law and a bound on the distance between empirical spectral distributions of two matrices.

Theorem 4 (Marčenko-Pastur law [BS10, Theorem 3.6]). Let \( X_1, \ldots, X_n \) be independent copies of a random vector \( X \in \mathbb{R}^p \) such that

\[
\mathbb{E}[X] = 0, \quad \mathbb{E}[X \otimes X] = I_p.
\]

Suppose that \( n \to \infty \), \( \frac{1}{n} \to \gamma > 0 \) and define \( a = (1 - \sqrt{\gamma})^2 \), and \( b = (1 + \sqrt{\gamma})^2 \). Then the empirical spectral distribution of the matrix

\[
\frac{1}{n} \sum_{i=1}^n X_i \otimes X_i,
\]

converges almost surely to the standard Marčenko-Pastur law which has density:

\[
p_\gamma(x) = \begin{cases} 
\frac{1}{2\pi x\gamma} \sqrt{(b-x)(x-a)}, & \text{if } a \leq x \leq b, \\
0, & \text{otherwise},
\end{cases}
\]

and has a point mass \( 1 - \frac{1}{\gamma} \) at the origin if \( \gamma > 1 \).

Lemma 5 ([BS10, Corollary A.41]). Let \( A \) and \( B \) be two \( p \times p \) normal matrices, with empirical spectral distributions \( \tilde{\mu}_A \) and \( \tilde{\mu}_B \). Then

\[
L(\tilde{\mu}_A, \tilde{\mu}_B)^3 \leq \frac{1}{p} \| A - B \|_F^2,
\]

where \( L(\tilde{\mu}_A, \tilde{\mu}_B) \) is the Lévy distance between the distribution functions \( \tilde{\mu}_A \) and \( \tilde{\mu}_B \).
References


