

On the Limiting Spectral Density of Symmetric Random Matrices with Correlated Entries

Olga Friesen* and Matthias Löwe

Abstract We analyze the spectral distribution of two different models of symmetric random matrices with correlated entries. While we assume that the diagonals of these random matrices are stochastically independent, the elements of the diagonals are taken to be correlated. Depending on the strength of correlation the limiting spectral distribution is either the famous semicircle law known for the limiting spectral density of symmetric random matrices with independent entries, or some other law related to that derived for Toeplitz matrices by Bryc W, Dembo A, Jiang T (2006) Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann Probab* 34(1):1–38.

1 Introduction

The study of random matrices started in the 1920s with the seminal work of Wishart [16]. His basic motivation was the analysis of data. On the other hand, Wigner used the eigenvalues of random matrices to model the spectra of heavy-nuclei atoms [15]. Nowadays, random matrix theory is a field with many applications from telecommunications to random graphs and with many interesting and surprising results.

A central role in the study of random matrices with growing dimension is played by their eigenvalues. To introduce them let, for any $n \in \mathbb{N}$,

*Research supported by Deutsche Forschungsgemeinschaft via SFB 878 at University of Münster.

O. Friesen (✉) · M. Löwe

Westfälische Wilhelms-Universität Münster, Fachbereich Mathematik, Einsteinstraße 62,
48149 Münster, Germany

e-mail: olga.friesen@uni-muenster.de; maloewe@math.uni-muenster.de

$\{a_n(p, q), 1 \leq p \leq q \leq n\}$ be a real-valued random field. Define the symmetric random $n \times n$ matrix \mathbf{X}_n by

$$\mathbf{X}_n(q, p) = \mathbf{X}_n(p, q) = \frac{1}{\sqrt{n}} a_n(p, q), \quad 1 \leq p \leq q \leq n.$$

We will denote the (real) eigenvalues of \mathbf{X}_n by $\lambda_1^{(n)} \leq \lambda_2^{(n)} \leq \dots \leq \lambda_n^{(n)}$. Let μ_n be the empirical eigenvalue distribution, i.e.

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k^{(n)}}.$$

Wigner proved in his fundamental work [15] that, if $a_n(p, q)$ are independent, normally distributed with mean 0 and variance 1, for off-diagonal elements, and variance 2 on the diagonal, the empirical eigenvalue distribution μ_n converges weakly (in probability) to the so-called semicircle distribution (or law), i.e. the probability distribution ν on \mathbb{R} with density

$$\frac{1}{2\pi} \sqrt{4 - x^2} \mathbb{1}_{[-2, 2]}(x) dx.$$

An important step to show the universality of this result was taken by Arnold [1], who verified that the convergence to the semicircle law also is true, if one replaces the Gaussian distributed random variables by independent and identically distributed (i.i.d.) random variables with a finite fourth moment. Also the identical distribution may be replaced by some other assumptions (see e.g. [8]). There are various ways to prove such a result. Among others, large deviations techniques as developed in [4] can be applied as well as Stieltjes transforms [2] (the latter method can also be applied to obtain results on the speed of convergence, see [12]). A still very powerful instrument is the moment method, originally employed by Wigner. To this end, it is useful to notice that, if Y is some random variable distributed according to this semicircle distribution, its moments are given by

$$\mathbb{E}[Y^k] = \begin{cases} 0, & \text{if } k \text{ is odd,} \\ C_{\frac{k}{2}}, & \text{if } k \text{ is even,} \end{cases}$$

where $C_k, k \in \mathbb{N}$, are the Catalan numbers defined by $C_k = (2k)! / (k!(k+1)!)$.

Recently, it was observed by Erdős et al. [9] that the convergence of the spectral measure towards the semicircle law holds in a local sense. More precisely, this can be proved on intervals with width going to zero sufficiently slowly.

However, the assumption of the entries being independent cannot be renounced without any replacement. Bryc, Dembo and Jiang [5] studied random symmetric Toeplitz matrices and obtained a different limiting distribution. To be more precise, they considered a family $\{X_i, 0 \leq i \leq n-1\}, n \in \mathbb{N}$, of independent and identically

distributed real-valued random variables, and assumed that $\text{Var}(X_1) = 1$. Then, the scaled symmetric Toeplitz matrix \mathbf{T}_n was defined by $\mathbf{T}_n(i, j) = 1/\sqrt{n} |X_{|i-j|}$, $1 \leq i, j \leq n$, i.e.

$$\mathbf{T}_n = \frac{1}{\sqrt{n}} \begin{pmatrix} X_0 & X_1 & X_2 & \dots & X_{n-2} & X_{n-1} \\ X_1 & X_0 & X_1 & \dots & X_{n-3} & X_{n-2} \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots \\ X_{n-2} & X_{n-3} & X_{n-4} & \dots & X_0 & X_1 \\ X_{n-1} & X_{n-2} & X_{n-3} & \dots & X_1 & X_0 \end{pmatrix}. \quad (1)$$

In this situation the empirical spectral distribution of \mathbf{T}_n converges weakly almost surely to some non-random probability measure γ_T as $n \rightarrow \infty$. This measure does not depend on the distribution of X_1 . Moreover, it has existing moments of all orders, is symmetric, and has an unbounded support. The aim of the present note is, to investigate the borderline between convergence to the semicircle law and the convergence to γ_T . To this end we will study random matrices with independent diagonals, where the elements on the diagonals may be correlated. If they are independent, we are, of course, back in the Wigner case, while for complete correlation the matrix is a random Toeplitz matrix.

We will see that depending on the strength of the correlation, the empirical spectral distribution either converges to the semicircle law, or to some mixture of γ_T and the semicircle distribution. We hence have a sort of phase transition. Similar results were obtained in [10] for the case of weak correlations and in [11] for stronger correlations. A particularly nice example is borrowed from statistical mechanics. There the Curie-Weiss model is the easiest model of a ferromagnet. Here a magnetic substance has little atoms that carry a magnetic spin, that is either $+1$ or -1 . These spins interact in cooperative way, the strength of the interaction being triggered by a parameter, the so-called inverse temperature. The model exhibits phase transition from paramagnetic to magnetic behavior (the standard reference for the Curie-Weiss model is [7]). We will see that this phase transition can be recovered on the level of the limiting spectral distribution of random matrices, if we fill their diagonals independently with the spins of Curie-Weiss models. For small interaction parameter, this limiting spectral distribution is the semicircle law, while for a large interaction parameter we obtain a distribution which shows the influence of γ_T .

The rest of this article is organized in the following way. In the two following sections we will fix our notation. Section 2 contains a description of the measures γ_T introduced above, while Sect. 3 describes the kind of matrices we will deal with in a general framework. From here we follow two paths. Section 4 contains our results for convergence towards the semicircle law, while Sect. 5 is devoted to the case of strong correlations along the diagonals. The basic ideas of the proofs, however are so similar, that we can treat them in a unified way. This is done in Sect. 6.

2 The Measure γ_T

The limiting spectral distribution γ_T can be defined by its moments which are described with the help of *Toeplitz volumes*.

Compared to [5], we will use a slightly different notation. This will make it easier to understand the arguments of the following sections. Thus, denote by $\mathcal{PP}(k)$, $k \in \mathbb{N}$, the set of all pair partitions of $\{1, \dots, k\}$. For any $\pi \in \mathcal{PP}(k)$, we write $i \sim_\pi j$ if i and j are in the same block of π . To introduce Toeplitz volumes, we associate to any $\pi \in \mathcal{PP}(k)$ the following system of equations in unknowns x_0, \dots, x_k :

$$\begin{aligned}
 x_1 - x_0 + x_{l_1} - x_{l_1-1} &= 0, & \text{if } 1 \sim_\pi l_1, \\
 x_2 - x_1 + x_{l_2} - x_{l_2-1} &= 0, & \text{if } 2 \sim_\pi l_2, \\
 & \vdots & \\
 x_i - x_{i-1} + x_{l_i} - x_{l_i-1} &= 0, & \text{if } i \sim_\pi l_i, \\
 & \vdots & \\
 x_k - x_{k-1} + x_{l_k} - x_{l_k-1} &= 0, & \text{if } k \sim_\pi l_k.
 \end{aligned} \tag{2}$$

Since π is a pair partition, we in fact have only $k/2$ equations although we have listed k . However, we have $k + 1$ variables. If $\pi = \{\{i_1, j_1\}, \dots, \{i_{k/2}, j_{k/2}\}\}$ with $i_l < j_l$ for any $l = 1, \dots, k/2$, we solve (2) for $x_{j_1}, \dots, x_{j_{k/2}}$, and leave the remaining variables undetermined. We further impose the condition that all variables x_0, \dots, x_k lie in the interval $I = [0, 1]$. Solving the equations above in this way determines a cross section of the cube $I^{k/2+1}$. The volume of this will be denoted by $p_T(\pi)$. To give an example, consider the partition $\pi = \{\{1, 3\}, \{2, 4\}\}$. Solving (2) for $x_3 = x_0 - x_1 + x_2$ and $x_4 = x_1 - x_2 + x_3 = x_0$, we obtain a cross section of I^3 given by

$$\{x_0 - x_1 + x_2 \in I\} \cap \{x_0 \in I\}.$$

This set has the volume $p_T(\pi) = 2/3$.

Returning to the measure γ_T , it is shown in [5] that all odd moments are zero, and for any even $k \in \mathbb{N}$, the k -th moment is given by

$$\int x^k d\gamma_T(x) = \sum_{\pi \in \mathcal{PP}(k)} p_T(\pi).$$

Since $|p_T(\pi)| \leq 1$ for any $\pi \in \mathcal{PP}(k)$ and $\#\mathcal{PP}(k) = (k-1)!!$, we have

$$\left| \int x^k d\gamma_T(x) \right| \leq (k-1)!!.$$

In particular, Carleman's condition holds implying that γ_T is uniquely determined by its moments. The results for the independent as well as the Toeplitz case will follow directly from Theorems 1 and 2 in case we assume the uniform boundedness of the moments of all orders.

3 Matrices with Independent Processes on the Diagonals

We want to study two different models of symmetric matrices with dependent entries. Both models have the common property that entries from different diagonals are independent while on each diagonal we have a stochastic process with a given covariance structure. Therefore, consider for any $n \in \mathbb{N}$ a family $\{a_n(p, q), 1 \leq p \leq q \leq n\}$ of real-valued random variables. Introduce the symmetric random $n \times n$ matrix \mathbf{X}_n with

$$\mathbf{X}_n(p, q) = \mathbf{X}_n(q, p) = \frac{1}{\sqrt{n}} a_n(p, q), \quad 1 \leq p \leq q \leq n.$$

Put $a_n(p, q) = a_n(q, p)$ if $1 \leq q < p \leq n$. Since we will resort to the method of moments, we first of all want to assume that

$$(A1) \quad \mathbb{E}[a_n(p, q)] = 0, \quad \mathbb{E}[a_n(p, q)^2] = 1, \text{ and}$$

$$m_k := \sup_{n \in \mathbb{N}} \max_{1 \leq p \leq q \leq n} \mathbb{E}[|a_n(p, q)|^k] < \infty, \quad k \in \mathbb{N}. \quad (3)$$

Note that the assumption of centered entries can be made without loss of generality if the family $\{a_n(p, q), 1 \leq p \leq q \leq n\}$ consists of identically distributed random variables. Indeed, assuming $\mathbb{E}[a_n(p, q)] = b_n$ for any $1 \leq p \leq q \leq n$, $n \in \mathbb{N}$, and some sequence $(b_n)_{n \in \mathbb{N}}$ such that $b_n = o(n)$ yields the same limiting spectral distribution as in the centered case, if it exists. This follows from the rank inequality for Hermitian matrices (cf. [3], Lemma 2.2). Changing the variance, however, provides a different limit which is a scaled version of that we obtain with assumption (A1). To make the condition of independent diagonals more precise, we suppose that

$$(A2) \quad \text{For any } n \in \mathbb{N}, j \in \{1, \dots, n\}, \text{ and distinct integers } r_1, \dots, r_j \in \{0, \dots, n-1\}, \text{ the families } \{a_n(p, p+r_1), 1 \leq p \leq n-r_1\}, \dots, \{a_n(p, p+r_j), 1 \leq p \leq n-r_j\} \text{ are independent.}$$

So far, we know that if we also have independence among the entries on the same diagonal, we will obtain the semicircle law as the limiting spectral distribution. Although we will violate this assumption in our first model, the quickly decaying dependency structure will ensure that nevertheless, we get the same limiting distribution. In our second model, we will basically assume that the covariance is

the same for any two entries on the same diagonal. If it is equal to the variance, it is not surprising that the resulting limit is the same as in the Toeplitz case. In general, we will find that we have a combination of the Toeplitz distribution and the semicircle law.

4 Quickly Decaying Covariances

In our first model, the dependency structure within the diagonals is determined by the conditions

- (A3) The covariance of two entries on the same diagonal can be bounded by some constant depending only on their distance, i.e. for any $n \in \mathbb{N}$ and $0 \leq \tau \leq n - 1$, there is a constant $c_n(\tau) \geq 0$ such that

$$|\text{Cov}(a_n(p, q), a_n(p + \tau, q + \tau))| \leq c_n(\tau), \quad 1 \leq p \leq q \leq n - \tau,$$

- (A4) The entries on the diagonals have a quickly decaying dependency structure, which will be expressed in terms of the condition

$$\sum_{\tau=0}^{n-1} c_n(\tau) = o(n).$$

Theorem 1. *Assume that the symmetric random matrix X_n satisfies the conditions (A1)–(A4). Then, with probability 1, the empirical spectral distribution of X_n converges weakly to the standard semicircle distribution.*

Remark 1. Note that in order for the semicircle law to hold, it is not possible to renounce condition (A4) without any replacement. To understand this, consider a Toeplitz matrix. We clearly have $c_n(\tau) = \mathcal{O}(1)$, and indeed, the empirical distribution of a sequence of Toeplitz matrices tends with probability 1 to a nonrandom probability measure with unbounded support.

4.1 Examples

We want to give some examples of processes that satisfy the assumptions of Theorem 1. Obviously, this is the case if the entries $\{a_n(p, q), 1 \leq p \leq q \leq n\}$ are independent satisfying (A1). The following three examples deal with finite Markov chains, Gaussian Markov processes and m -dependent processes.

- (i) Assume that $\{x(p), p \in \mathbb{N}\}$ is a stationary Markov chain on a finite state space $S = \{s_1, \dots, s_N\}$, $N \geq 2$. Denote by $\varrho = (\varrho_1, \dots, \varrho_N)$ the stationary distribution, and suppose that

$$\mathbb{E}[x(p)] = \sum_{j=1}^N s_j \varrho(j) = 0, \quad \mathbb{E}[x(p)^2] = \sum_{j=1}^N s_j^2 \varrho(j) = 1.$$

If $\{x(p), p \in \mathbb{N}\}$ is aperiodic and irreducible, we have that for some constant $C > 0$ and some $\alpha \in (0, 1)$,

$$\max_{i,j \in \{1, \dots, N\}} |\mathbb{P}(x(p) = s_i \mid x(1) = s_j) - \varrho(i)| \leq C\alpha^{p-1}, \quad p \in \mathbb{N}.$$

For more details, see [13], Theorem 4.9. In particular, we obtain

$$|\text{Cov}(x(p), x(1))| = \left| \sum_{i,j=1}^N s_i s_j \left(\mathbb{P}(x(p) = s_i \mid x(1) = s_j) - \varrho(i) \right) \varrho(j) \right| \leq C\alpha^{p-1}.$$

Now assume that the processes $\{a(p, p+r), p \in \mathbb{N}\}$, $r \in \mathbb{N}_0$, are independent copies of $\{x(p), p \in \mathbb{N}\}$, and put $a_n(p, q) := a(p, q)$ for any $n \in \mathbb{N}$, $1 \leq p \leq q \leq n$. Condition (A2) then holds by definition, and the uniform moment bound in (A1) is given since we have a bounded support. Furthermore,

$$|\text{Cov}(a_n(p, q), a_n(p+\tau, q+\tau))| \leq c_n(\tau),$$

where $c_n(\tau) = c(\tau) = C\alpha^\tau$. This is assumption (A3). Finally, (A4) follows since $\sum_{\tau=0}^{\infty} c(\tau) < \infty$, implying $\sum_{\tau=0}^{n-1} c(\tau) = o(n)$.

- (ii) Let $\{y(p), p \in \mathbb{N}\}$ be a stationary Gaussian Markov process with mean 0 and variance 1. In addition to this, assume that the process is non-degenerate in the sense that $\mathbb{E}[y(p)|y(q), q \leq p-1] \neq y(p)$. In this case, we can represent $y(p)$ as

$$y(p) = b_p \sum_{j=1}^p d_j \xi_j,$$

where $\{\xi_j, j \in \mathbb{N}\}$ is a family of independent standard Gaussian variables and $b_p, d_1, \dots, d_p \in \mathbb{R} \setminus \{0\}$. We can now calculate

$$\bar{c}(\tau) := \text{Cov}(y(p+\tau), y(p)) = b_{p+\tau} b_p \sum_{i=1}^{p+\tau} \sum_{j=1}^p d_i d_j \mathbb{E}[\xi_i \xi_j] = b_{p+\tau} b_p \sum_{j=1}^p d_j^2.$$

Note that $1 = \mathbb{E}[y(p)^2] = b_p^2 \sum_{j=1}^p d_j^2$. As a consequence, we have

$$\bar{c}(\tau) = \frac{b_{p+\tau}}{b_p} = \frac{b_{p+\tau}}{b_{p+\tau-1}} \frac{b_{p+\tau-1}}{b_{p+\tau-2}} \dots \frac{b_{p+1}}{b_p} = \bar{c}(1)^\tau.$$

To see that $|\bar{c}(\tau)| < 1$, we first compute $\bar{c}(1)y(1) = b_2b_1^2d_1^3\xi_1 = b_2d_1\xi_1$, implying that $y(2) = b_2d_2\xi_2 + \bar{c}(1)y(1)$. Using this identity to calculate the variance, we can take account of the independence of ξ_2 and $y(1)$ to obtain

$$1 = \mathbb{E}[y(2)^2] = b_2^2d_2^2 + \bar{c}(1)^2.$$

Since $b_2, d_2 \neq 0$, we conclude that $|\bar{c}(1)| < 1$. In analogy to the first example, we assume that the processes $\{a(p, p+r), p \in \mathbb{N}\}, r \in \mathbb{N}_0$, are independent copies of $\{y(p), p \in \mathbb{N}\}$, and put $a_n(p, q) := a(p, q)$ for any $n \in \mathbb{N}, 1 \leq p \leq q \leq n$. Conditions (A1) and (A2) obviously hold. Defining $c(\tau) := |\bar{c}(\tau)|$, we further obtain

$$|\text{Cov}(a_n(p, q), a_n(p + \tau, q + \tau))| \leq c(\tau).$$

Since $|\bar{c}(1)| < 1$, we have $\sum_{\tau=0}^{\infty} c(\tau) < \infty$, implying $\sum_{\tau=0}^{n-1} c(\tau) = o(n)$. Thus assumptions (A3) and (A4) are satisfied.

- (iii) Assume that $\{z(p), p \in \mathbb{N}\}$ is a stationary process of m -dependent random variables, i.e. $z(p)$ and $z(q)$ are stochastically independent whenever $|p - q| > m$. Moreover, suppose that $z(1)$ is centered with unit variance, and has existing moments of all orders. Define

$$c(\tau) := |\text{Cov}(z(1), z(\tau + 1))|, \quad \tau \in \mathbb{N}_0.$$

Then, $c(\tau) = 0$ for any $\tau > m$. Thus, $\sum_{\tau=0}^{n-1} c(\tau) = \sum_{\tau=0}^m c(\tau) = o(n)$ for any $n \geq m + 1$. Let $\{a(p, p+r), p \in \mathbb{N}\}, r \in \mathbb{N}_0$, be independent copies of $\{z(p), p \in \mathbb{N}\}$, and $a_n(q, p) := a(p, q)$ for any $n \in \mathbb{N}, 1 \leq p \leq q \leq n$. Then, (A1)–(A4) are satisfied.

5 Constant Covariances

For our second model, we assume that

- (A3') The covariance of two distinct entries on the same diagonal depends only on n , i.e. for any $1 \leq \tau \leq n - 1$ and $1 \leq p, q \leq n - \tau$, we can define

$$c_n := \text{Cov}(a_n(p, q), a_n(p + \tau, q + \tau)),$$

- (A4') The limit $c := \lim_{n \rightarrow \infty} c_n$ exists.

To describe the limiting spectral distribution in this case, we want to resort to pair partitions. However, we need a further notion which proved to be useful in [5] when considering the limiting spectral distribution of Markov matrices.

Definition 1. Let $k \in \mathbb{N}$ be even, and fix $\pi \in \mathcal{PP}(k)$. The *height* $h(\pi)$ of π is the number of elements $i \sim_\pi j$, $i < j$, such that either $j = i + 1$ or the restriction of π to $\{i + 1, \dots, j - 1\}$ is a pair partition.

Note that the property that the restriction of π to $\{i + 1, \dots, j - 1\}$ is a pair partition in particular requires that the distance $j - i - 1 \geq 1$ is even. To give an example how to calculate the height of a partition, take $\pi = \{\{1, 6\}, \{2, 4\}, \{3, 5\}\}$. Considering the block $\{1, 6\}$, we see that the restriction of π to $\{2, 3, 4, 5\}$ is a pair partition, namely $\{\{2, 4\}, \{3, 5\}\}$. However, this is not true for both remaining blocks. Hence, $h(\pi) = 1$.

In the following, we will say that a pair partition $\pi \in \mathcal{PP}(k)$ is *crossing* if there are $i < j < l < m$ such that $i \sim_\pi l$ and $j \sim_\pi m$. Otherwise, we call the pair partition *non-crossing*. The set of all crossing pair partitions of $\{1, \dots, k\}$ is denoted by $\mathcal{CP}(k)$, and the set of all non-crossing pair partitions by $\mathcal{NPP}(k)$. Recall that for even $k \in \mathbb{N}$, the Catalan number $C_{k/2}$ is given by $C_{k/2} = \#\mathcal{NPP}(k)$.

We can now state the main result of this section.

Theorem 2. Assume that the symmetric random matrix X_n satisfies the conditions (A1), (A2), (A3'), and (A4'). Then, with probability 1, the empirical spectral distribution of X_n converges weakly to a deterministic probability distribution ν_c with k -th moment

$$\int x^k d\nu_c(x) = \begin{cases} C_{\frac{k}{2}} + \sum_{\pi \in \mathcal{CP}(k)} p_T(\pi) c^{\frac{k}{2} - h(\pi)}, & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd.} \end{cases}$$

If k is even, we can also write $\int x^k d\nu_c(x) = \sum_{\pi \in \mathcal{PP}(k)} p_T(\pi) c^{k/2 - h(\pi)}$.

Remark 2. As for the limiting distribution in the Toeplitz case, we can verify the Carleman condition to see that ν_c is uniquely determined by its moments.

Remark 3. If $c = 0$, Theorem 2 states that the limiting distribution is the semicircle law since $h(\pi) < k/2$ for any crossing partition $\pi \in \mathcal{CP}(k)$. This result can also be deduced from Theorem 1. Indeed, choose $c_n(\tau) = |c_n|$ for any $\tau \geq 1$, and $c_n(0) = 1$. We then have for any $1 \leq p \leq q \leq n - \tau$,

$$|\text{Cov}(a_n(p, q), a_n(p + \tau, q + \tau))| = c_n(\tau).$$

Furthermore, we obtain $\sum_{\tau=0}^{n-1} c_n(\tau) = 1 + (n - 1)|c_n| = o(n)$ since $\lim_{n \rightarrow \infty} c_n = c = 0$. Consequently, (A3) and (A4) are satisfied.

5.1 Examples

We want to give some examples of processes satisfying the assumptions of Theorem 2.

- (i) Consider a symmetric Toeplitz matrix as in (1). The limiting spectral distribution can be deduced from Theorem 2 as well. Indeed, assuming that the entries are centered with unit variance and have existing moments of any order, we see that all conditions are satisfied with $c = c_n = 1$. Thus, we get

$$\int x^k d\nu_1(x) = \begin{cases} C_{\frac{k}{2}} + \sum_{\pi \in \mathcal{C}\mathcal{P}\mathcal{P}(k)} p_T(\pi) = \sum_{\pi \in \mathcal{P}\mathcal{P}(k)} p_T(\pi), & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd,} \end{cases}$$

as proven in [5].

- (ii) Suppose that for any $n \in \mathbb{N}$, $\{x_n(p), 1 \leq p \leq n\}$ is a family of exchangeable random variables, i.e. the distribution of the vector $(x_n(1), \dots, x_n(n))$ is the same as that of $(x_n(\sigma(1)), \dots, x_n(\sigma(n)))$ for any permutation σ of $\{1, \dots, n\}$. In this case, we can conclude that for any $1 \leq p < q \leq n$, we have

$$\text{Cov}(x_n(p), x_n(q)) = \text{Cov}(x_n(1), x_n(2)) =: c_n.$$

Now assume that $c_n \rightarrow c \in \mathbb{R}$ as $n \rightarrow \infty$. Define for any $n \in \mathbb{N}$, $r \in \{0, \dots, n-1\}$, the process $\{a_n(p, p+r), 1 \leq p \leq n-r\}$ to be an independent copy of $\{x_n(p), 1 \leq p \leq n-r\}$. Then, all conditions of Theorem 2 are satisfied if we ensure that the moment condition (A1) holds. The resulting limiting distribution for different choices of c is depicted in Fig. 1.

An example for a process with exchangeable variables is the Curie-Weiss model with inverse temperature $\beta > 0$. Here, the vector $x_n = (x_n(1), \dots, x_n(n))$ takes values in $\{-1, 1\}^n$, and for any $\omega = (\omega(1), \dots, \omega(n)) \in \{-1, 1\}^n$, we have

$$\mathbb{P}(x_n = \omega) = \frac{1}{Z_{n,\beta}} \exp\left(\frac{\beta}{2n} \left(\sum_{i=1}^n \omega(i)\right)^2\right),$$

where $Z_{n,\beta}$ is the normalizing constant. Since $\mathbb{P}(x_n(1) = -1) = \mathbb{P}(x_n(1) = 1) = 1/2$, we obtain $\mathbb{E}[x_n(1)] = 0$. Further, we clearly have $\mathbb{E}[x_n(1)^2] = 1$. It remains to determine $c = \lim_{n \rightarrow \infty} c_n$. Therefore, we want to make use of the identity

$$c_n = \text{Cov}(x_n(1), x_n(2)) = \mathbb{E}[x_n(1)x_n(2)] = \frac{n}{n-1} \mathbb{E}[m_n^2] - \frac{1}{n-1},$$

where $m_n := 1/n \sum_{i=1}^n x_n(i)$ is the so-called magnetization of the system. Since $|m_n| \leq 1$, we see that m_n is uniformly integrable. Thus, m_n converges in \mathcal{L}^2 to some random variable m if and only if $m_n \rightarrow m$ in probability. In [6], it was verified that $m_n \rightarrow 0$ in probability if $\beta \leq 1$, and $m_n \rightarrow m$ with $m \sim 1/2 \delta_{m(\beta)} + 1/2 \delta_{-m(\beta)}$ for some $m(\beta) > 0$ if $\beta > 1$. The function $m(\beta)$ is

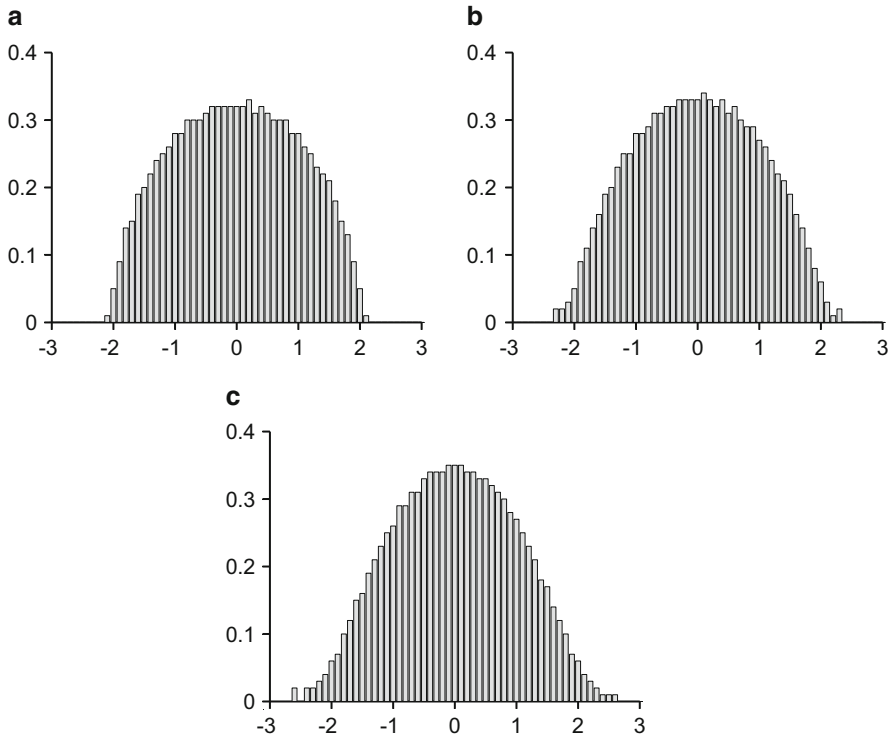


Fig. 1 Histograms of the empirical spectral distribution of 100 realizations of $1,000 \times 1,000$ matrices $\mathbf{X}_{1,000}$ with standard Gaussian entries. (a) $c = 0.25$. (b) $c = 0.5$. (c) $c = 0.75$

monotonically increasing on $(1, \infty)$, and satisfies $m(\beta) \rightarrow 0$ as $\beta \searrow 1$ and $m(\beta) \rightarrow 1$ as $\beta \rightarrow \infty$. We now obtain

$$c = \lim_{n \rightarrow \infty} c_n = \begin{cases} 0, & \text{if } \beta \leq 1, \\ m(\beta)^2, & \text{if } \beta > 1. \end{cases}$$

Thus, the limiting spectral distribution of \mathbf{X}_n is the semicircle law if $\beta \leq 1$, and approximately the Toeplitz limit if β is large. This is insofar not surprising as the different sites in the Curie-Weiss model show little interaction, i.e. behave almost independently, if the temperature is high, or, in other words, β is small. However, if the temperature is low, i.e. β is large, the magnetization of the sites strongly depends on each other. The phase transition at the critical inverse temperature $\beta = 1$ in the Curie-Weiss model is thus reflected in the limiting spectral distribution of \mathbf{X}_n as well.

6 Proof of Theorems 1 and 2

The proofs of both theorems start with the same idea. We need to distinguish them only as soon as the covariances have to be calculated. The main technique we want to apply is the method of moments. The idea is to first determine the weak limit of the expected empirical distribution. Afterwards, concentration inequalities can be used to obtain almost sure convergence.

6.1 The Expected Empirical Spectral Distribution

To determine the limit of the k -th moment of the expected empirical spectral distribution μ_n of \mathbf{X}_n , we write

$$\begin{aligned} \mathbb{E} \left[\int x^k d\mu_n(x) \right] &= \frac{1}{n} \mathbb{E} \left[\text{tr}(\mathbf{X}_n^k) \right] \\ &= \frac{1}{n^{\frac{k}{2}+1}} \sum_{p_1, \dots, p_k=1}^n \mathbb{E} [a(p_1, p_2)a(p_2, p_3) \cdots a(p_{k-1}, p_k)a(p_k, p_1)]. \end{aligned}$$

The main task is now to compute the expectations on the right hand side. However, we have to face the problem that some of the entries involved are independent and some are not. To be more precise, $a(p_1, q_1), \dots, a(p_j, q_j)$ are independent whenever they can be found on different diagonals of \mathbf{X}_n , i.e. the distances $|p_1 - q_1|, \dots, |p_j - q_j|$ are distinct. Hence, a first step in our proof is to consider the expectation $\mathbb{E} [a(p_1, p_2)a(p_2, p_3) \cdots a(p_{k-1}, p_k)a(p_k, p_1)]$, and to identify entries with the same distance of their indices. Therefore, we want to adapt some concepts of [14] and [5] to our situation.

To start with, fix $k \in \mathbb{N}$, and define $\mathcal{T}_n(k)$ to be the set of k -tuples of *consistent pairs*, that is multi-indices (P_1, \dots, P_k) satisfying for any $j = 1, \dots, k$,

1. $P_j = (p_j, q_j) \in \{1, \dots, n\}^2$,
2. $q_j = p_{j+1}$, where $k+1$ is cyclically identified with 1.

With this notation, we find that

$$\frac{1}{n} \mathbb{E} \left[\text{tr}(\mathbf{X}_n^k) \right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in \mathcal{T}_n(k)} \mathbb{E} [a_n(P_1) \cdots a_n(P_k)].$$

To reflect the dependency structure among the entries $a_n(P_1) \dots a_n(P_k)$, we want to make use of the set $\mathcal{P}(k)$ of partitions of $\{1, \dots, k\}$. Thus, take $\pi \in \mathcal{P}(k)$. We say that an element $(P_1, \dots, P_k) \in \mathcal{T}_n(k)$ is a π -consistent sequence if

$$|p_i - q_i| = |p_j - q_j| \iff i \sim_\pi j.$$

According to condition (A2), this implies that $a_n(P_{i_1}), \dots, a_n(P_{i_l})$ are stochastically independent if i_1, \dots, i_l belong to l different blocks of π . The set of all π -consistent sequences $(P_1, \dots, P_k) \in \mathcal{T}_n(k)$ is denoted by $S_n(\pi)$. Note that the sets $S_n(\pi)$, $\pi \in \mathcal{P}(k)$, are pairwise disjoint, and $\bigcup_{\pi \in \mathcal{P}(k)} S_n(\pi) = \mathcal{T}_n(k)$. Consequently, we can write

$$\frac{1}{n} \mathbb{E} [\text{tr}(\mathbf{X}_n^k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{P}(k)} \sum_{(P_1, \dots, P_k) \in S_n(\pi)} \mathbb{E} [a_n(P_1) \cdots a_n(P_k)]. \quad (4)$$

In a next step, we want to exclude partitions that do not contribute to (4) as $n \rightarrow \infty$. These are those partitions satisfying either $\#\pi > k/2$ or $\#\pi < k/2$, where $\#\pi$ denotes the number of blocks of π . We want to treat the two cases separately.

First case: $\#\pi > k/2$. Since π is a partition of $\{1, \dots, k\}$, there is at least one singleton, i.e. a block containing only one element i . Consequently, $a_n(P_i)$ is independent of $\{a_n(P_j), j \neq i\}$ if $(P_1, \dots, P_k) \in S_n(\pi)$. Since we assumed the entries to be centered, we obtain

$$\mathbb{E} [a_n(P_1) \cdots a_n(P_k)] = \mathbb{E} \left[\prod_{i \neq l} a_n(P_i) \right] \mathbb{E} [a_n(P_l)] = 0.$$

This yields

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n(\pi)} \mathbb{E} [a_n(P_1) \cdots a_n(P_k)] = 0.$$

Second case: $r := \#\pi < k/2$. Here, we want to argue that π gives vanishing contribution to (4) as $n \rightarrow \infty$ by calculating $\#S_n(\pi)$. To fix an element $(P_1, \dots, P_k) \in S_n(\pi)$, we first choose the pair $P_1 = (p_1, q_1)$. There are at most n possibilities to assign a value to p_1 , and another n possibilities for q_1 . To fix $P_2 = (p_2, q_2)$, note that the consistency of the pairs implies $p_2 = q_1$. If now $1 \sim_\pi 2$, the condition $|p_1 - q_1| = |p_2 - q_2|$ allows at most two choices for q_2 . Otherwise, if $1 \not\sim_\pi 2$, we have at most n possibilities. We now proceed sequentially to determine the remaining pairs. When arriving at some index i , we check whether i is in the same block as some preceding index $1, \dots, i-1$. If this is the case, then we have at most two choices for P_i and otherwise, we have n . Since there are exactly $r = \#\pi$ different blocks, we can conclude that

$$\#S_n(\pi) \leq n^2 n^{r-1} 2^{k-r} \leq C n^{r+1} \quad (5)$$

with a constant $C = C(r, k)$ depending on r and k .

Now the uniform boundedness of the moments (3) and the Hölder inequality together imply that for any sequence (P_1, \dots, P_k) ,

$$|\mathbb{E} [a_n(P_1) \cdots a_n(P_k)]| \leq \left[\mathbb{E} |a_n(P_1)|^k \right]^{\frac{1}{k}} \cdots \left[\mathbb{E} |a_n(P_k)|^k \right]^{\frac{1}{k}} \leq m_k. \quad (6)$$

Consequently, taking account of the relation $r < k/2$, we get

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n(\pi)} |\mathbb{E}[a_n(P_1) \cdots a_n(P_k)]| \leq C \frac{\#S_n(\pi)}{n^{\frac{k}{2}+1}} \leq C \frac{1}{n^{\frac{k}{2}-r}} = o(1).$$

Combining the calculations in the first and the second case, we can conclude that

$$\frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\substack{\pi \in \mathcal{P}(k), \\ \#\pi = \frac{k}{2}}} \sum_{(P_1, \dots, P_k) \in S_n(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] + o(1).$$

Now assume that k is *odd*. Then the condition $\#\pi = k/2$ cannot be satisfied, and the considerations above immediately yield

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] = 0.$$

It remains to determine the even moments. Thus, let $k \in \mathbb{N}$ be *even*. Recall that we denoted by $\mathcal{PP}(k) \subset \mathcal{P}(k)$ the set of all pair partitions of $\{1, \dots, k\}$. In particular, $\#\pi = k/2$ for any $\pi \in \mathcal{PP}(k)$. On the other hand, if $\#\pi = k/2$ but $\pi \notin \mathcal{PP}(k)$, we can conclude that π has at least one singleton and hence, as in the first case above, the expectation corresponding to the π -consistent sequences will become zero. Consequently,

$$\frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{PP}(k)} \sum_{(P_1, \dots, P_k) \in S_n(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] + o(1). \quad (7)$$

We have now reduced the original set $\mathcal{P}(k)$ to the subset $\mathcal{PP}(k)$. Next we want to fix a $\pi \in \mathcal{PP}(k)$ and concentrate on the set $S_n(\pi)$. The following lemma will help us to calculate that part of (7) which involves non-crossing partitions.

Lemma 1 (cf. [5], Proposition 4.4). *Let $S_n^*(\pi) \subseteq S_n(\pi)$ denote the set of π -consistent sequences (P_1, \dots, P_k) satisfying*

$$i \sim_{\pi} j \implies q_i - p_i = p_j - q_j$$

for all $i \neq j$. Then, we have

$$\#(S_n(\pi) \setminus S_n^*(\pi)) = o\left(n^{1+\frac{k}{2}}\right).$$

Proof. We call a pair (P_i, P_j) with $i \sim_{\pi} j$, $i \neq j$, *positive* if $q_i - p_i = q_j - p_j > 0$ and *negative* if $q_i - p_i = q_j - p_j < 0$. Since $\sum_{i=1}^k q_i - p_i = 0$ by consistency, the existence of a negative pair implies the existence of a positive one. Thus, we can

assume that any $(P_1, \dots, P_k) \in S_n(\pi) \setminus S_n^*(\pi)$ contains a positive pair (P_l, P_m) . To fix such a sequence, we first determine the positions of l and m , and then fix the signs of the remaining differences $q_i - p_i$. The number of possibilities to accomplish this depends only on k and not on n . Now we choose one of n possible values for p_l , and continue with assigning values to the differences $|q_i - p_i|$ for all P_i except for P_l and P_m . Since π is a pair partition, we have at most $n^{k/2-1}$ possibilities for that. Then, $\sum_{i=1}^k q_i - p_i = 0$ implies that

$$0 < 2(q_l - p_l) = q_l - p_l + q_m - p_m = \sum_{\substack{i \in \{1, \dots, k\}, \\ i \neq l, m}} p_i - q_i.$$

Since we have already chosen the signs of the differences $|q_i - p_i|$, $i \neq l, m$, as well as their absolute values, we know the value of the sum on the right hand side. Hence, the difference $q_l - p_l = q_m - p_m$ is fixed. We now have the index p_l , all differences $|q_i - p_i|$, $i \in \{1, \dots, k\}$, and their signs. Thus, we can start at P_l and go systematically through the whole sequence (P_1, \dots, P_k) to see that it is uniquely determined. Consequently, our considerations lead to

$$\#(S_n(\pi) \setminus S_n^*(\pi)) \leq Cn^{\frac{k}{2}} = o\left(n^{1+\frac{k}{2}}\right). \quad \square$$

A consequence of Lemma 1 and relation (6) is the identity

$$\frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{PP}(k)} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] + o(1). \quad (8)$$

As already mentioned, the sets $S_n^*(\pi)$ help us to deal with the set $\mathcal{NPP}(k)$ of non-crossing pair partitions.

Lemma 2. *Let $\pi \in \mathcal{NPP}(k)$. For any $(P_1, \dots, P_k) \in S_n^*(\pi)$, we have*

$$\mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = 1.$$

Proof. Let $l < m$ with $l \sim_\pi m$. Since π is non-crossing, the number $l - m - 1$ of elements between l and m must be even. In particular, there is $l \leq i < j \leq m$ with $i \sim_\pi j$ and $j = i + 1$. By the properties of $S_n^*(\pi)$, we have $a_n(P_i) = a_n(P_j)$, and the sequence $(P_1, \dots, P_l, \dots, P_{i-1}, P_{i+2}, \dots, P_m, \dots, P_k)$ is still consistent. Applying this argument successively, all pairs between l and m can be eliminated and we see that the sequence $(P_1, \dots, P_l, P_m, \dots, P_k)$ is consistent, that is $q_l = p_m$. Then, the identity $p_l = q_m$ also holds. In particular, $a_n(P_l) = a_n(P_m)$. Since this argument applies for arbitrary $l \sim_\pi m$, we obtain

$$\mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \prod_{\substack{l < m, \\ l \sim_{\pi} m}} \mathbb{E}[a_n(P_l) a_n(P_m)] = 1. \quad \square$$

By Lemma 2, we can conclude that

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{NPP}(k)} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{NPP}(k)} \#S_n^*(\pi).$$

The following lemma allows us to finally calculate the term on the right hand side.

Lemma 3. *For any $\pi \in \mathcal{NPP}(k)$, we have*

$$\lim_{n \rightarrow \infty} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} = 1.$$

Proof. Since π is non-crossing, we can find a nearest neighbor pair $i \sim_{\pi} i + 1$. Now fix $(P_1, \dots, P_k) \in S_n^*(\pi)$, and write $P_l = (p_l, p_{l+1})$, $l = 1, \dots, k$, where $k + 1$ is identified with 1. Then the properties of $S_n^*(\pi)$ ensure that $(p_i, p_{i+1}) = (p_{i+2}, p_{i+1})$. Hence, we can eliminate the pairs P_i, P_{i+1} to obtain a sequence $(P_1^{(1)}, \dots, P_{k-2}^{(1)}) := (P_1, \dots, P_{i-1}, P_{i+2}, \dots, P_k)$ which is still consistent. Denote by π' the partition obtained from π by deleting the block $\{i, i + 1\}$, and relabeling any $l \geq i + 2$ to $l - 2$. Since π is non-crossing, we have $\pi' \in \mathcal{NPP}(k - 2)$. Moreover, $(P_1^{(1)}, \dots, P_{k-2}^{(1)}) \in S_n^*(\pi')$. Thus we see that any $(P_1, \dots, P_k) \in S_n^*(\pi)$ can be reconstructed from a tuple $(P_1^{(1)}, \dots, P_{k-2}^{(1)}) \in S_n^*(\pi')$ and a choice of p_{i+1} . The latter admits $n - (k - 2)/2$ possibilities since $\{i, i + 1\}$ forms a block on its own in π . Consequently,

$$\frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} = \frac{\#S_n^*(\pi')}{n^{\frac{k}{2}}} + o(1). \quad (9)$$

Now if $k = 2$, we get $S_n^*(\pi) = \{(p, q), (q, p)\} : p, q \in \{1, \dots, n\}$, implying $\#S_n^*(\pi)/n^2 = 1$. For arbitrary even $k \in \mathbb{N}$, the statement of Lemma 3 follows then by induction using the identity in (9). \square

Taking account of the relation $\#\mathcal{NPP}(k) = C_{k/2}$, we now arrive at

$$\begin{aligned} & \frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] \\ &= C_{\frac{k}{2}} + \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{CPP}(k)} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] + o(1), \end{aligned} \quad (10)$$

with $\mathcal{CP}\mathcal{P}(k)$ being the set of all crossing pair partitions of $\{1, \dots, k\}$. At this point, we have to distinguish between Theorems 1 and 2. Indeed, to obtain the semicircle law, we need to show that the sum over all crossing partitions is negligible in the limit. However, the limiting distribution in Theorem 2 indicates that we do have a contribution.

6.2 Convergence of the Expected Empirical Spectral Distribution in Theorem 1

The convergence of the expected empirical spectral distribution to the semicircle distribution follows directly from relation (10) and

Lemma 4. *For any crossing $\pi \in \mathcal{CP}\mathcal{P}(k)$, we have*

$$\sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = o\left(n^{\frac{k}{2}+1}\right).$$

Proof. Let π be crossing and consider a sequence $(P_1, \dots, P_k) \in S_n^*(\pi)$. Write $P_l = (p_l, p_{l+1})$. Note that if there is an $l \in \{1, \dots, k\}$ with $l \sim_\pi l + 1$, where $k + 1$ is identified with 1, we immediately have $a_n(P_l) = a_n(P_{l+1})$. In particular,

$$\mathbb{E}[a_n(P_l)a_n(P_{l+1})] = 1.$$

The sequence $(P_1^{(1)}, \dots, P_{k-2}^{(1)}) := (P_1, \dots, P_{l-1}, P_{l+2}, \dots, P_k)$ is still consistent, and

$$\mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \mathbb{E}\left[a_n(P_1^{(1)}) \cdots a_n(P_{k-2}^{(1)})\right].$$

Define $\pi^{(1)} \in \mathcal{CP}\mathcal{P}(k-2)$ to be the pair partition induced by π after eliminating the indices l and $l + 1$. In particular, $(P_1^{(1)}, \dots, P_{k-2}^{(1)}) \in S_n^*(\pi^{(1)})$. Since there are at most n choices for p_{l+1} when $(P_1^{(1)}, \dots, P_{k-2}^{(1)})$ is fixed, we have for any $(Q_1, \dots, Q_{k-2}) \in S_n^*(\pi^{(1)})$,

$$\#\{(P_1, \dots, P_k) \in S_n^*(\pi) : (P_1^{(1)}, \dots, P_{k-2}^{(1)}) = (Q_1, \dots, Q_{k-2})\} \leq n.$$

Let r denote the maximum number of pairs of indices that can be eliminated in this way. Since π is crossing, there are at least two pairs left and hence, $r \leq k/2 - 2$. Define $\pi^{(r)} \in \mathcal{CP}\mathcal{P}(k - 2r)$ and $(P_1^{(r)}, \dots, P_{k-2r}^{(r)}) \in S_n^*(\pi^{(r)})$ to be the partition and the sequence left after this elimination. By induction, we conclude that for any $(Q_1, \dots, Q_{k-2r}) \in S_n^*(\pi^{(r)})$, we have the estimate

$$\#\{(P_1, \dots, P_k) \in S_n^*(\pi) : (P_1^{(r)}, \dots, P_{k-2r}^{(r)}) = (Q_1, \dots, Q_{k-2r})\} \leq n^r.$$

Since $\mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \mathbb{E}[a_n(P_1^{(r)}) \cdots a_n(P_{k-2r}^{(r)})]$, we obtain

$$\begin{aligned} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} |\mathbb{E}[a_n(P_1) \cdots a_n(P_k)]| \\ \leq n^r \sum_{(Q_1, \dots, Q_{k-2r}) \in S_n^*(\pi^{(r)})} |\mathbb{E}[a_n(Q_1) \cdots a_n(Q_{k-2r})]|. \end{aligned} \quad (11)$$

Choose $i \sim_{\pi^{(r)}} i+j$ such that j is minimal. For any sequence $(Q_1, \dots, Q_{k-2r}) \in S_n^*(\pi^{(r)})$, put $Q_l = (q_l, q_{l+1})$, $l = 1, \dots, k-2r$. We want to count the number of such sequences given that q_i and q_{i+j+1} are fixed. Therefore, we start with choosing one of n possible values for q_{i+1} . Then, the fact that i is equivalent to $i+j$ ensures that we can also deduce the value of

$$q_{i+j} = q_{i+1} - q_i + q_{i+j+1}.$$

Hence, Q_i and Q_{i+j} are fixed. Since j is minimal, any element in $\{i+1, \dots, i+j-1\}$ is equivalent to some element outside the set $\{i, \dots, i+j\}$. There are n possibilities to fix $Q_{i+1} = (q_{i+1}, q_{i+2})$ because q_{i+1} is already fixed. Proceeding sequentially, we have n possibilities for the choice of any pair Q_l with $l \in \{i+2, \dots, i+j-2\}$, and there is only one choice for Q_{i+j-1} since q_{i+j} is already chosen. We thus made n^{j-2} choices to fix all pairs Q_l , $l \in \{i+1, \dots, i+j-1\}$. For any Q_l with $l \in \{1, \dots, k-2r\} \setminus \{i, \dots, i+j\}$, there are at most n possibilities if it is not equivalent to one pair that has already been chosen. Otherwise, there is only one possibility. Since there were $k/2 - r - j$ new equivalence classes left, we have at most $n^{k/2-r-j}$ choices for those pairs. Hence, assuming that the elements q_i and q_{i+j+1} are fixed, we have at most

$$n n^{j-2} n^{\frac{k}{2}-r-j} = n^{\frac{k}{2}-r-1}$$

possibilities to choose the rest of the sequence $(Q_1, \dots, Q_{k-2r}) \in S_n^*(\pi^{(r)})$. Note that $|\mathbb{E}[a_n(Q_l)a_n(Q_m)]| \leq (\mathbb{E}[a_n(Q_l)^2])^{1/2}(\mathbb{E}[a_n(Q_m)^2])^{1/2} = 1$. Since $\pi^{(r)}$ is a pair partition, we thus get

$$|\mathbb{E}[a_n(Q_1) \cdots a_n(Q_{k-2r})]| \leq |\mathbb{E}[a_n(Q_i)a_n(Q_{i+j})]|.$$

By assumption (A3), the expectation on the right hand side depends only on the absolute value of the difference

$$\min\{q_i, q_{i+1}\} - \min\{q_{i+j}, q_{i+j+1}\} = \max\{q_i, q_{i+1}\} - \max\{q_{i+j}, q_{i+j+1}\}.$$

Now the definition of $S_n^*(\pi^{(r)})$ ensures that $q_i - q_{i+1} = q_{i+j+1} - q_{i+j}$. In particular, $\min\{q_i, q_{i+1}\} - \min\{q_{i+j}, q_{i+j+1}\} = q_i - q_{i+j+1} = q_{i+j} - q_{i+1}$, and

$$\mathbb{E}[a_n(Q_i)a_n(Q_{i+j})] = c_n(|q_i - q_{i+j+1}|).$$

Consequently, estimating the term in (11) further, we obtain

$$\begin{aligned} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} |\mathbb{E}[a_n(P_1) \cdots a_n(P_k)]| &\leq n^{\frac{k}{2}-1} \sum_{q_i, q_{i+j+1}=1}^n c_n(|q_{i+j+1} - q_i|) \\ &\leq C n^{\frac{k}{2}} \sum_{\tau=0}^{n-1} c_n(\tau) = o\left(n^{\frac{k}{2}+1}\right), \end{aligned}$$

since $\sum_{\tau=0}^{n-1} c_n(\tau) = o(n)$ by condition (A4). \square

6.3 Convergence of the Expected Empirical Spectral Distribution in Theorem 2

We again start with the identity in (10). Since we consider only pair partitions, we know that the expectation on the right hand side is of the form

$$\mathbb{E}[a_n(p_1, q_1)a_n(p_1 + \tau_1, q_1 + \tau_1)] \cdots \mathbb{E}[a_n(p_r, q_r)a_n(p_r + \tau_r, q_r + \tau_r)],$$

for $r := k/2$ and some choices of $p_1, q_1, \tau_1, \dots, p_r, q_r, \tau_r \in \mathbb{N}$. In order to calculate this expectation, assumption (A3') indicates that we only need to distinguish for any $i = 1, \dots, k$, whether we have $\tau_i = 0$ or not. In the first case, we get the identity $\mathbb{E}[a_n(p_i, q_i)a_n(p_i + \tau_i, q_i + \tau_i)] = 1$, in the second we can conclude that $\mathbb{E}[a_n(p_i, q_i)a_n(p_i + \tau_i, q_i + \tau_i)] = c_n$. Fix some pair partition $\pi \in \mathcal{PP}(k)$, and take $(P_1, \dots, P_k) \in S_n^*(\pi)$. Motivated by these considerations, we put $P_i = (p_i, q_i)$, and define

$$\begin{aligned} m(P_1, \dots, P_k) &:= \#\{1 \leq i < j \leq k : a_n(P_i) = a_n(P_j)\} \\ &= \#\{1 \leq i < j \leq k : (p_i, q_i) = (q_j, p_j)\}. \end{aligned}$$

Obviously, we have $0 \leq m(P_1, \dots, P_k) \leq k/2$. With this notation, we find that

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{l=0}^{k/2} c_n^{\frac{k}{2}-l} \#A_n^{(l)}(\pi), \quad (12)$$

where

$$A_n^{(l)}(\pi) := \{(P_1, \dots, P_k) \in S_n^*(\pi) : m(P_1, \dots, P_k) = l\}.$$

The following lemma states that if a pair P_i, P_j contributes to $m(P_1, \dots, P_k)$, then we can assume that the block $\{i, j\}$ in π is not crossed by any other block.

Lemma 5. *Let $\pi \in \mathcal{PP}(k)$ and fix $i \sim_\pi j$, $i < j$. Define*

$$S_n^*(\pi; i, j) := \{(P_1, \dots, P_k) \in S_n^*(\pi) : P_i = (p_i, q_i), P_j = (p_j, q_j), p_i = q_j, q_i = p_j\}.$$

Assume that there is some $i' \sim_\pi j'$ such that $i < i' < j$, and either $j' < i$ or $j < j'$. Then,

$$\#S_n^*(\pi; i, j) = o\left(n^{\frac{k}{2}+1}\right).$$

Proof. To fix some $(P_1, \dots, P_k) \in S_n^*(\pi; i, j)$, we first choose a value for $p_i = q_j$ and $q_i = p_j$. This allows for at most n^2 possibilities. Hence, P_i and P_j are fixed. Now consider the pairs $P_{i+1}, \dots, P_{i'-1}$. p_{i+1} is uniquely determined by consistency. For q_{i+1} , there are at most n choices. Then, $p_{i+2} = q_{i+1}$. If $i+2 \sim_\pi i+1$, we have one choice for q_{i+2} . Otherwise, there are at most n . Proceeding in the same way, we see that we have n possibilities whenever we start a new equivalence class. Similarly, we can assign values to the pairs $P_j, \dots, P_{j'+1}$ in this order. Now $P_{j'}$ is determined by consistency. When fixing $P_{i-1}, \dots, P_i, P_k, \dots, P_{j+1}$, we again have n choices for any new equivalence class. To sum up, we are left with at most

$$n^2 n^{\frac{k}{2}-2} = n^{\frac{k}{2}}$$

possible values for an element in $S_n^*(\pi; i, j)$. □

Recall Definition 1 where we introduced the notion of the *height* $h(\pi)$ of a pair partition π . Lemma 5 in particular implies that only those $(P_1, \dots, P_k) \in S_n^*(\pi)$ with

$$0 \leq m(P_1, \dots, P_k) \leq h(\pi)$$

contribute to the limit of (12). Indeed, if $m(P_1, \dots, P_k) > h(\pi)$, we can find some $i \sim_\pi j$, $i < j$, such that $(P_1, \dots, P_k) \in S_n^*(\pi; i, j)$ and neither $j = i+1$ nor is the restriction of π to $\{i+1, \dots, j-1\}$ a pair partition. Hence, the crossing property in Lemma 5 is satisfied, and (P_1, \dots, P_k) is contained in a set that is negligible in the limit. The identity in (12) thus becomes

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{l=0}^{h(\pi)} c_n^{\frac{k}{2}-l} \#B_n^{(l)}(\pi) + o(1),$$

where

$$B_n^{(l)}(\pi) := \{(P_1, \dots, P_k) \in S_n^*(\pi) : m(P_1, \dots, P_k) = l; \\ a_n(P_i) = a_n(P_j), i < j \Rightarrow j = i+1 \text{ or } \pi|_{\{i+1, \dots, j-1\}} \text{ is a pair partition}\}.$$

In the next step, we want to simplify the expression above further by showing that $B_n^{(l)}(\pi) = \emptyset$ whenever $0 \leq l < h(\pi)$. This is ensured by

Lemma 6. *Let $\pi \in \mathcal{PP}(k)$. For any $(P_1, \dots, P_k) \in S_n^*(\pi)$, we have*

$$m(P_1, \dots, P_k) \geq h(\pi).$$

Proof. If $h(\pi) = 0$, there is nothing to prove. Thus, suppose that $h(\pi) \geq 1$ and take some $i \sim_\pi j$, $i < j$, such that either $j = i + 1$ or $j - i - 1 \geq 2$ is even and the restriction of π to $\{i + 1, \dots, j - 1\}$ is a pair partition. Fix $(P_1, \dots, P_k) \in S_n^*(\pi)$, and write $P_l = (p_l, p_{l+1})$ for any $l = 1, \dots, k$. We need to verify that $p_{i+1} = p_j$. If we achieve this, the definition of $S_n^*(\pi)$ will also ensure that $p_i = p_{j+1}$. As a consequence, the π -block $\{i, j\}$ will contribute to $m(P_1, \dots, P_k)$. Since there are $h(\pi)$ such blocks, we will obtain $m(P_1, \dots, P_k) \geq h(\pi)$ for any choice of $(P_1, \dots, P_k) \in S_n^*(\pi)$.

If $j = i + 1$, we immediately obtain $p_{i+1} = p_j$. To show this property in the second case, note that the sequence $(P_{i+1}, \dots, P_{j-1})$ solves the following system of equations:

$$\begin{aligned} p_{i+2} - p_{i+1} + p_{l_1+1} - p_{l_1} &= 0, & \text{if } i + 1 \sim_\pi l_1, \\ p_{i+3} - p_{i+2} + p_{l_2+1} - p_{l_2} &= 0, & \text{if } i + 2 \sim_\pi l_2, \\ & \vdots \\ p_{i+m+1} - p_{i+m} + p_{l_m+1} - p_{l_m} &= 0, & \text{if } i + m \sim_\pi l_m, \\ & \vdots \\ p_j - p_{j-1} + p_{l_{j-i-1}+1} - p_{l_{j-i-1}} &= 0, & \text{if } j - 1 \sim_\pi l_{j-i-1}. \end{aligned}$$

Start with solving the first equation for p_{i+2} which yields

$$p_{i+2} = p_{i+1} - p_{l_1+1} + p_{l_1}.$$

Then, insert this in the second equation, and solve it for p_{i+3} to obtain

$$p_{i+3} = p_{i+1} - p_{l_1+1} + p_{l_1} - p_{l_2+1} + p_{l_2}.$$

In the $j - i - 1$ -th step, we substitute $p_{j-1} = p_{i+(j-i-1)}$ in the $j - i - 1$ -th equation, and solve it for $p_j = p_{i+(j-i-1)+1}$. We then have

$$p_j = p_{i+1} - \sum_{m=1}^{j-i-1} (p_{l_m+1} - p_{l_m}).$$

Since the restriction of π to $\{i+1, \dots, j-1\}$ is a pair partition, we can conclude that the sets $\{l_1, \dots, l_{j-i-1}\}$ and $\{i+1, \dots, j-1\}$ are equal. Hence, we obtain $\sum_{m=1}^{j-i-1} (p_{l_m+1} - p_{l_m}) = p_j - p_{i+1}$, implying $p_j = p_{i+1}$. \square

With the help of Lemma 6, we thus arrive at

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \frac{\#B_n^{(h(\pi))}(\pi)}{n^{\frac{k}{2}+1}} c_n^{\frac{k}{2}-h(\pi)} + o(1).$$

Note that any element $(P_1, \dots, P_k) \in S_n^*(\pi)$ satisfying the condition

$$a_n(P_i) = a_n(P_j), i < j \Rightarrow j = i + 1 \text{ or } \pi|_{\{i+1, \dots, j-1\}} \text{ is a pair partition,} \quad (13)$$

fulfills the condition $m(P_1, \dots, P_k) = h(\pi)$ as well. Indeed, (13) guarantees that $m(P_1, \dots, P_k) \leq h(\pi)$, and Lemma 6 ensures that $m(P_1, \dots, P_k) \geq h(\pi)$. Thus, we can write

$$B_n^{(h(\pi))}(\pi) = \{(P_1, \dots, P_k) \in S_n^*(\pi) :$$

$$a_n(P_i) = a_n(P_j), i < j \Rightarrow j = i + 1 \text{ or } \pi|_{\{i+1, \dots, j-1\}} \text{ is a pair partition}\}.$$

Now any element in the complement of $B_n^{(h(\pi))}(\pi)$ satisfies for some $i \sim_\pi j$ the crossing assumption in Lemma 5. This yields

$$\frac{\#(B_n^{(h(\pi))}(\pi))^c}{n^{\frac{k}{2}+1}} = o(1).$$

Since $B_n^{(h(\pi))}(\pi) \cup (B_n^{(h(\pi))}(\pi))^c = S_n^*(\pi)$, we obtain that

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} c_n^{\frac{k}{2}-h(\pi)} + o(1). \quad (14)$$

To calculate the limit on the right-hand side, we have

Lemma 7 (cf. [5], Lemma 4.6). *For any $\pi \in \mathcal{PP}(k)$, it holds that*

$$\lim_{n \rightarrow \infty} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} = p_T(\pi),$$

where $p_T(\pi)$ is the Toeplitz volume defined by solving the system of equation (2).

Proof. Fix $\pi \in \mathcal{PP}(k)$. Note that if $P = \{(p_i, p_{i+1}), i = 1, \dots, k\} \in S_n^*(\pi)$, then x_0, x_1, \dots, x_k with $x_i = p_{i+1}/n$ is a solution of the system of equation (2). On the other hand, if $x_0, x_1, \dots, x_k \in \{1/n, 2/n, \dots, 1\}$ is a solution of (2) and $p_{i+1} = nx_i$, then either $\{(p_i, p_{i+1}), i = 1, \dots, k\} \in S_n^*(\pi)$ or $\{(p_i, p_{i+1}), i = 1, \dots, k\} \in S_n(\eta)$ for some partition $\eta \in \mathcal{P}(k)$ such that $i \sim_\pi j \Rightarrow i \sim_\eta j$, but $\#\eta < \#\pi$.

In (2), we have $k + 1$ variables and only $k/2$ equations. Denote the $k/2 + 1$ undetermined variables by $y_1, \dots, y_{k/2+1}$. We thus need to assign values from the set $\{1/n, 2/n, \dots, 1\}$ to $y_1, \dots, y_{k/2+1}$, and then to calculate the remaining $k/2$ variables from the equations. Since the latter are also supposed to be in the range $\{1/n, 2/n, \dots, 1\}$, it might happen that not all values for the undetermined variables are admissible. Let $p_n(\pi)$ denote the admissible fraction of the $n^{k/2+1}$ choices for $y_1, \dots, y_{k/2+1}$. By our remark at the beginning of the proof and estimate (5), we have that

$$\lim_{n \rightarrow \infty} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} = \lim_{n \rightarrow \infty} p_n(\pi),$$

if the limits exist. Now we can interpret $y_1, \dots, y_{k/2+1}$ as independent random variables with a uniform distribution on $\{1/n, 2/n, \dots, 1\}$. Then, $p_n(\pi)$ is the probability that the computed values stay within the interval $(0, 1]$. As $n \rightarrow \infty$, $y_1, \dots, y_{k/2+1}$ converge in law to independent random variables uniformly distributed on $[0, 1]$. Hence, $p_n(\pi) \rightarrow p_T(\pi)$. \square

Applying Lemma 7 and assumption (A4') to Eq. (14), we arrive at

$$\lim_{n \rightarrow \infty} \frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1, \dots, P_k) \in S_n^*(\pi)} \mathbb{E}[a_n(P_1) \cdots a_n(P_k)] = p_T(\pi) c^{\frac{k}{2}-h(\pi)}.$$

Substituting this result in (10), we find that for any even $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] = C_{\frac{k}{2}} + \sum_{\pi \in \mathcal{CP}(k)} p_T(\pi) c^{\frac{k}{2}-h(\pi)}.$$

To obtain the alternative expression for the even moments in Theorem 2, note that the considerations above were not restricted to crossing partitions. In particular, we can start from identity (8) instead of (10) to see that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] = \lim_{n \rightarrow \infty} \sum_{\pi \in \mathcal{PP}(k)} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} c_n^{\frac{k}{2}-h(\pi)} = \sum_{\pi \in \mathcal{PP}(k)} p_T(\pi) c^{\frac{k}{2}-h(\pi)}.$$

6.4 Almost Sure Convergence

The aim of this part of the proof is to show almost sure convergence in both situations, that of Theorem 1 and that of Theorem 2. Therefore, we want to follow the ideas used in [5], Proposition 4.9, to verify the same for Toeplitz matrices. This is possible since the arguments only rely on the fact that the diagonals are independent. The particular dependency structure can be neglected. We want to start with the proof of a concentration inequality for the moments of the spectral measure. The obtained bound will enable us to use the Borel-Cantelli lemma.

Lemma 8. *Suppose that conditions (A1) and (A2) hold. Then, for any $k, n \in \mathbb{N}$,*

$$\mathbb{E} \left[\left(\text{tr}(\mathbf{X}_n^k) - \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] \right)^4 \right] \leq C n^2.$$

Proof. Fix $k, n \in \mathbb{N}$. Using the notation

$$P = (P_1, \dots, P_k) = ((p_1, q_1), \dots, (p_k, q_k)), \quad a_n(P) = a_n(P_1) \cdots a_n(P_k),$$

we have that

$$\begin{aligned} & \mathbb{E} \left[\left(\text{tr}(\mathbf{X}_n^k) - \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] \right)^4 \right] \\ &= \frac{1}{n^{2k}} \sum_{\pi^{(1)}, \dots, \pi^{(4)} \in \mathcal{P}(k)} \sum_{\substack{P^{(i)} \in \mathcal{S}_n(\pi^{(i)}) \\ i=1, \dots, 4}} \mathbb{E} \left[\prod_{j=1}^4 (a_n(P^{(j)}) - \mathbb{E}[a_n(P^{(j)})]) \right]. \end{aligned} \quad (15)$$

Now consider a partition π of $\{1, \dots, 4k\}$. We say that a sequence $(P^{(1)}, \dots, P^{(4)})$ is π -consistent if each $P^{(i)}, i = 1, \dots, 4$, is a consistent sequence and

$$|q_l^{(i)} - p_l^{(i)}| = |q_m^{(j)} - p_m^{(j)}| \iff l + (i-1)k \sim_\pi m + (j-1)k.$$

Let $\mathcal{S}_n(\pi)$ denote the set of all π -consistent sequences with entries in $\{1, \dots, n\}$. Then, (15) becomes

$$\begin{aligned} & \mathbb{E} \left[\left(\text{tr}(\mathbf{X}_n^k) - \mathbb{E}[\text{tr}(\mathbf{X}_n^k)] \right)^4 \right] \\ &= \frac{1}{n^{2k}} \sum_{\pi \in \mathcal{P}(4k)} \sum_{(P^{(1)}, \dots, P^{(4)}) \in \mathcal{S}_n(\pi)} \mathbb{E} \left[\prod_{j=1}^4 (a_n(P^{(j)}) - \mathbb{E}[a_n(P^{(j)})]) \right]. \end{aligned} \quad (16)$$

We want to analyze the expectation on the right hand side. Therefore, fix a $\pi \in \mathcal{P}(4k)$. We call π a *matched partition* if

1. Any block of π contains at least two elements,
2. For any $i \in \{1, \dots, 4\}$, there is a $j \neq i$ and $l, m \in \{1, \dots, k\}$ with

$$l + (i - 1)k \sim_{\pi} m + (j - 1)k.$$

In case π does not satisfy (i), we have a singleton $\{l + (i - 1)k\}$, implying that $\mathbb{E}[a_n(P^{(i)})] = 0$. As a consequence,

$$\sum_{(P^{(1)}, \dots, P^{(4)}) \in \mathcal{S}_n(\pi)} \mathbb{E} \left[\prod_{j=1}^4 (a_n(P^{(j)}) - \mathbb{E}[a_n(P^{(j)})]) \right] = 0. \quad (17)$$

If π does not satisfy (ii), we can conclude that for some $i \in \{1, \dots, 4\}$, the term $a_n(P^{(i)}) - \mathbb{E}[a_n(P^{(i)})]$ is independent of $a_n(P^{(j)}) - \mathbb{E}[a_n(P^{(j)})]$, $j \neq i$. Thus, (17) holds in this case as well. To sum up, we only have to consider matched partitions to evaluate the sum in (16). Let π be such a partition and denote by $r = \#\pi$ the number of blocks of π . Note that condition (i) implies $r \leq 2k$. We want to count all π consistent sequences $(P^{(1)}, \dots, P^{(4)})$. Therefore, first choose one of at most n^r possibilities to fix the r different equivalence classes. Afterwards, we fix the elements $p_1^{(1)}, \dots, p_1^{(4)}$, which can be done in n^4 ways. Since now the differences $|q_l^{(i)} - p_l^{(i)}|$ are uniquely determined by the choice of the corresponding equivalence classes, we can proceed sequentially to see that there are at most two choices left for any pair $P_l^{(i)}$. To sum up, we have at most

$$2^{4k} n^4 n^r = C n^{r+4}$$

possibilities to choose $(P^{(1)}, \dots, P^{(4)})$. If now $r \leq 2k - 2$, we can conclude that

$$\#\mathcal{S}_n(\pi) \leq C n^{2k+2}. \quad (18)$$

It remains to consider the cases in which $r = 2k - 1$ and $r = 2k$, respectively. To begin with, let $r = 2k - 1$. Then, we have either two equivalence classes with three elements or one equivalence class with four. Since π is matched, there must exist an $i \in \{1, \dots, 4\}$ and an $l \in \{1, \dots, k\}$ such that $P_l^{(i)}$ is not equivalent to any other pair in the sequence $P^{(i)}$. Without loss of generality, we can assume that $i = 1$. In contrast to the construction of $(P^{(1)}, \dots, P^{(4)})$ as above, we now alter our procedure as follows: We fix all equivalence classes except of that $P_l^{(1)}$ belongs to. There are n^{r-1} possibilities to accomplish that. Now we choose again one of n^4 possible values for $p_1^{(1)}, \dots, p_1^{(4)}$. Hereafter, we fix $q_m^{(1)}$, $m = 1, \dots, l - 1$, and then start from $q_k^{(1)} = p_1^{(1)}$ to go backwards and obtain the values of $p_k^{(1)}, \dots, p_{l+1}^{(1)}$. Each of these steps leaves at most two choices to us, that is 2^{k-1} choices in total. But

now, $P_l^{(1)}$ is uniquely determined since $p_l^{(1)} = q_{l-1}^{(1)}$ and $q_l^{(1)} = p_{l+1}^{(1)}$ by consistency. Thus, we had to make one choice less than before, implying (18).

Now, let $r = 2k$. In this case, each equivalence class has exactly two elements. Since we consider a matched partition, we can find here as well an $l \in \{1, \dots, k\}$ such that $P_l^{(1)}$ is not equivalent to any other pair in the sequence $P^{(1)}$. But in addition to that, we also have an $m \in \{1, \dots, k\}$ such that, possibly after relabeling, $P_m^{(2)}$ is neither equivalent to any element in $P^{(1)}$ nor to any other element in $P^{(2)}$. Thus, we can use the same argument as before to see that this time, we can reduce the number of choices to at most $C n^{r+2} = C n^{2k+2}$. In conclusion, (18) holds for any matched partition π . To sum up our results, we obtain that

$$\begin{aligned} & \mathbb{E} \left[\left(\text{tr}(\mathbf{X}_n^k) - \mathbb{E} \left[\text{tr}(\mathbf{X}_n^k) \right] \right)^4 \right] \\ &= \frac{1}{n^{2k}} \sum_{\substack{\pi \in \mathcal{P}(4k), \\ \pi \text{ matched}}} \sum_{(P^{(1)}, \dots, P^{(4)}) \in \mathcal{S}_n(\pi)} \mathbb{E} \left[\prod_{j=1}^4 \left(a_n(P^{(j)}) - \mathbb{E} \left[a_n(P^{(j)}) \right] \right) \right] \leq C n^2, \end{aligned}$$

which is the statement of Lemma 8. \square

From Lemma 8 and Chebyshev's inequality, we can now conclude that for any $\varepsilon > 0$ and any $k, n \in \mathbb{N}$,

$$\mathbb{P} \left(\left| \frac{1}{n} \text{tr}(\mathbf{X}_n^k) - \mathbb{E} \left[\frac{1}{n} \text{tr}(\mathbf{X}_n^k) \right] \right| > \varepsilon \right) \leq \frac{C}{\varepsilon^4 n^2}.$$

Applying the Borel-Cantelli lemma, we see that

$$\frac{1}{n} \text{tr}(\mathbf{X}_n^k) - \mathbb{E} \left[\frac{1}{n} \text{tr}(\mathbf{X}_n^k) \right] \rightarrow 0, \quad \text{a.s.} \quad (19)$$

Let Y be a random variable distributed according to the semicircle law or according to the distribution given by the moments in Theorem 2, depending on whether we want to prove Theorems 1 or 2. The convergence of the moments of the expected empirical distributions and relation (19) yield

$$\frac{1}{n} \text{tr}(\mathbf{X}_n^k) \rightarrow \mathbb{E}[Y^k], \quad \text{a.s.}$$

Since the distribution of Y is uniquely determined by its moments, we obtain almost sure weak convergence of the empirical spectral distribution of \mathbf{X}_n to the distribution of Y .

References

1. Arnold, L.: On wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **19**, 191–198 (1971)
2. Bai, Z.D.: Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.* **21**(2), 625–648 (1993)
3. Bai, Z.: Methodologies in spectral analysis of large-dimensional random matrices, a review. *Stat. Sin.* **9**(3), 611–677 (1999). With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author
4. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Relat. Fields* **108**(4), 517–542 (1997)
5. Bryc, W., Dembo, A., Jiang, T.: Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.* **34**(1), 1–38 (2006)
6. Ellis, R., Newman, C.: Fluctuations in Curie-Weiss exemplis. In: Osterwalder, K., Seiler, E. (eds.) *Mathematical Problems in Theoretical Physics. Lecture Notes in Physics*, vol. 80, pp. 313–324. Springer, Berlin/Heidelberg (1978)
7. Ellis, R.S.: *Entropy, Large Deviations, and Statistical Mechanics. Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*, vol. 271. Springer, New York (1985)
8. Erdős, L.: Universality of Wigner random matrices: a survey of recent results. *Uspekhi Mat. Nauk* **66**(3(399)), 67–198 (2011)
9. Erdős, L., Schlein, B., Yau, H.T.: Local semicircle law and complete delocalization for Wigner random matrices. *Commun. Math. Phys.* **287**(2), 641–655 (2009)
10. Friesen, O., Löwe, M.: The semicircle law for matrices with independent diagonals. *J. Theor. Probab.* (2011, Preprint, To appear)
11. Friesen, O., Löwe, M.: A phase transition for the limiting spectral distribution of random matrices. *Electron. J. Probab.* **114**(17), 1–17 (2013)
12. Götze, F., Tikhomirov, A.: Rate of convergence to the semi-circular law. *Probab. Theory Relat. Fields* **127**, 228–276 (2003)
13. Levin, D.A., Peres, Y., Wilmer, E.L.: *Markov Chains and Mixing Times*. American Mathematical Society, Providence (2006)
14. Schenker, J., Schulz-Baldes, H.: Semicircle law and freeness for random matrices with symmetries or correlations. *Math. Res. Lett.* **12**, 531–542 (2005)
15. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. *Ann. Math.* **67**, 325–328 (1958)
16. Wishart, J.: The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* **20**, 32–52 (1928)