Elect. Comm. in Probab. **14** (2009), 412–421

ELECTRONIC COMMUNICATIONS in PROBABILITY

SPECTRUM OF RANDOM TOEPLITZ MATRICES WITH BAND STRUCTURE

VLADISLAV KARGIN
Mathematics Department, Stanford University, Palo Alto, CA 94305
email: kargin@stanford.edu

Submitted May 6, 2009, *accepted in final form* September 6, 2009

AMS 2000 Subject classification: 15A52
Keywords: random Toeplitz matrices

Abstract

This paper considers the eigenvalues of symmetric Toeplitz matrices with independent random entries and band structure. We assume that the entries of the matrices have zero mean and a uniformly bounded 4th moment, and we study the limit of the eigenvalue distribution when both the size of the matrix and the width of the band with non-zero entries grow to infinity. It is shown that if the bandwidth/size ratio converges to zero, then the limit of the eigenvalue distributions is Gaussian. If the ratio converges to a positive limit, then the distributions converge to a non-Gaussian distribution, which depends only on the limit ratio. A formula for the fourth moment of this distribution is derived.

1 Introduction

In recent interesting papers, (Bryc et al., 2006) and (Hammond and Miller, 2005), it was shown that the distribution of Toeplitz matrices with i.i.d. entries converges to a universal law as the size of the matrix grows. The limit law is not Gaussian, which is all the more surprising if we compare this result with the result of (Bose and Mitra, 2002) which says that the limit law is Gaussian for circulants, a closely related ensemble of matrices.

This comparison suggests that the difference in the behavior of the limit law is due to entries in the upper-right and lower-left corners of the matrices. For this reason, it is interesting to study what happens if the entries in these corners are set equal to zero, that is, if the matrices have a band structure.

It turns out that if the ratio of the band width to the matrix size decreases to zero as the matrices grow, then it is possible to adjust the matrices in such a way that the resulting matrix is a circulant, and the effect of the modification on the eigenvalue distribution is negligible. This allows us to show that in this case the limit eigenvalue distribution is Gaussian.

In contrast, if the ratio of the band width to the matrix size approaches a positive limit as the size of the matrix grows, then the limit eigenvalue law is not Gaussian. The proof of this result is by the method of moments and by an explicit calculation of the kurtosis of the limit law. It turns out that it is different from the Gaussian value of 3 for all width/size ratios except one. This exceptional ratio needs a calculation of higher moments. Numerical estimation suggests that even in this case the limit distribution is not Gaussian.

Besides papers mentioned above, the spectrum of random Toeplitz matrices is investigated in (Bose and Sen, 2007), which studies the case when the entries of Toeplitz matrices are non-centered, in (Massey et al., 2007), where the main interest is in random Toeplitz matrices with palyndromic structure, and in (Meckes, 2007), which studies the norm of random Toeplitz matrices. In addition, (Chatterjee, 2009) shows that the fluctuations of the eigenvalue distribution around the limit are Gaussian.

The rest of the paper is organized as follows. Section 2 is about the case when the ratio of the band width to the matrix size decreases as the size of the matrix grows, and Section 3 is about the case when this ratio tends to a non-zero limit.

2 Vanishing band to size ratio

We consider finite symmetric Toeplitz matrices that have the band structure. In other words, we assume that entries of an N-by-N symmetric matrix X_N, which we denote X_{ij}, depend only on the difference $i-j$ and that they are zero if the difference is large, i.e. $X_{ij} = a_{|i-j|}$ and $a_k = 0$ if $k > M$. It is understood that a_k depends on N. We will suppress the dependence in order to make the notation less cumbersome.

If $\lambda_1, \ldots, \lambda_N$ are eigenvalues of the matrix X_N, counted with multiplicities, then we define the *empirical probability measure of eigenvalues* as

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}.$$

We are interested in the behavior of measures μ_N as both N and M grow to infinity. Let the cumulative distribution function of μ_N be denoted as $F_N(x)$.

Our first result is that if the size of the matrix grows faster than the width of the band, then the spectral distributions $F_N(x)$ converge to the Gaussian distribution.

Theorem 2.1. *Let X_N be a sequence of the random symmetric Toeplitz matrices with the band of width $M(N)$. Assume that the non-zero entries are independent real random variables such that $Ea_k = 0$, $Ea_k^2 = 1/M$ and $\sup_{k,N} E\left|\sqrt{M}a_k\right|^4 < C < \infty$. If both $M(N) \to \infty$ and $M/N \to 0$ as $N \to \infty$, then for every x,*

$$E\left[F_N(\sqrt{2}x)\right] \to (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2} dt,$$

and $\operatorname{Var}\left[F_N(x)\right] \to 0$.

Our strategy of proof is by relating eigenvalue distributions of Toeplitz matrices and circulants. Namely, we will modify the upper-right and lower-left corners of the matrix X_N so that the resulting matrix Y_N is a circulant and the eigenvalue distribution of Y_N is close to the eigenvalue distribution of X_N. See (Gray, 1972) for an explanation of this method in the deterministic setting. We will adapt this method to the case of random Toeplitz matrices.

Proof: Set all entries of Y_N on the N-th and $(-N)$-th diagonal to $a_N := a_1$, all entries on the $(N-1)$-st and $-(N-1)$-st diagonals to $a_{N-1} := a_2$, and so on. Finally set all entries on the

$(N-M+1)$-st and $-(N-M+1)$-st diagonals equal to $a_{N-M+1} := a_M$. All other entries of Y_N are the same as those of X_N.

By choosing N sufficiently large, we can ensure that $N > 2M$. Then, the change affects only zero entries of the matrix X_N. The total number of entries changed is equal to $M(M+1)$.

Let us introduce a measure of distance between probability distributions (cf. (Trotter, 1984)). If $F(x)$ and $G(x)$ are two distributions, then the *distance in distribution* is defined by

$$d(F,G)^2 = \inf E(X-Y)^2,$$

where the infimum is over all random variables X and Y, which are defined over the same probability space and which have the distributions F and G. The usefulness of this distance stems from the fact that if two $d\left(F_N, G_N\right) \to 0$, then

$$\int f(x)\, dF_N(x) - \int f(x)\, dG_N(x) \to 0$$

for every continuous function $f(x)$. For the proof, see (Trotter, 1984).

We are going to use this result in the following way. We will take the spectral distribution of Y_N as G_N, and we will show that $d\left(F_N, G_N\right) \to 0$ in probability. This implies that for each t,

$$F_N(t) \to G_N(t) \tag{1}$$

in probability. Note that $F_N(t)$ and $G_N(t)$ are positive random variables bounded by 1. For each N, these two random variables are defined on the same probability space. This fact, and the convergence in probability in (1) imply that

$$EF_N(t) \to EG_N(t),$$

and

$$E\left(F_N(t) - G_N(t)\right)^2 \to 0.$$

Hence, in order to prove the theorem it will remain to show that

$$E\left[G_N(\sqrt{2}x)\right] \to (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2} dt,$$

and that $Var\left[G_N(x)\right] \to 0$.

In order to prove that $d\left(F_N, G_N\right) \to 0$ in probability, we use a result from (Hoffman and Wielandt, 1953). Let the Frobenius norm of a matrix A be defined by

$$\|A\|_2^2 = \frac{1}{N} \mathrm{Tr}\left(A^* A\right),$$

where N is the size of the matrix A.

Lemma 2.2. *Let $\Lambda(X)$ and $\Lambda(Y)$ be empirical distributions of eigenvalues of matrices X and Y. Then*

$$d(\Lambda(X), \Lambda(Y)) \le \|X - Y\|_2.$$

For proof see (Hoffman and Wielandt, 1953) or (Trotter, 1984).

For random matrices X_N and Y_N, we compute:

$$E\left\|X_N - Y_N\right\|_2^2 = \frac{(M+1)}{N}.$$

This expression converges to zero as $N \to \infty$.
By using the independence of random variables a_i, we can also compute

$$\begin{aligned} E\left\|X_N - Y_N\right\|_2^4 &= \left[E\left\|X_N - Y_N\right\|_2^2\right]^2 + \frac{4}{N^2}\left(E\left|a_1\right|^4 + 4E\left|a_2\right|^4 + \ldots + M^2 E\left|a_M\right|^4\right) \\ &\quad -\frac{4}{N^2}\left(\left(E\left|a_1\right|^2\right)^2 + 4\left(E\left|a_2\right|^2\right)^2 + \ldots + M^2\left(E\left|a_M\right|\right)^2\right) \end{aligned}$$

Hence, if we use the uniform boundedness of $E\left(\sqrt{M}a_i\right)^4$ and the equality $E\left|a_i\right|^2 = M^{-1}$, we find that

$$\mathrm{Var}\left(\left\|X_N - Y_N\right\|_2^4\right) \leq C\frac{M^2}{N^2} \to 0 \text{ as } N \to \infty.$$

It follows that $\left\|X_N - Y_N\right\|_2^2$ and, therefore, $d\left(F_N, G_N\right)$ converges to zero in probability.
It remains to investigate $G_N(t)$, the empirical distribution function of eigenvalues of Y_N. These matrices are circulants and we can use the results and methods of (Bose and Mitra, 2002) in order to show that $E\left[G_N\left(\sqrt{2}x\right)\right]$ converges pointwise to the standard Gaussian law and $\mathrm{Var}\left\{G_N(x)\right\} \to 0$. For completeness, we give a detailed argument.
The eigenvalues of Y_N can be computed explicitly as

$$\lambda_{k,N} = a_0 + 2\sum_{l=1}^{M} a_l \cos\left(\frac{2\pi kl}{N}\right). \tag{2}$$

Let us write the random function $G_N(x)$ as follows:

$$G_N(x) = \frac{1}{N}\sum_{k=1}^{N} 1_{(-\infty,x]}\left(\lambda_{k,N}\right).$$

Let $\{x\} = x - [x]$, where $[x]$ denotes the integer which is closest to x with ties broken in favor of smaller integers.

Lemma 2.3. *Assume* $\{2k/N\} > \varepsilon > 0$. *Then*

$$\left|\Pr\left\{\lambda_{k,N} \leq \sqrt{2}x\right\} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt\right| \leq \frac{c(\varepsilon)}{\sqrt{M}}, \tag{3}$$

where $c(\varepsilon)$ *does not depend on* k, M, *or* N.

Proof: Let θ_k denotes $2\pi k/N$. Formula (2) implies that each $\lambda_{k,N}$ is the sum of $M+1$ independent random variables with zero mean and variance equal to

$$\frac{4}{M}\cos^2\left(l\theta_k\right) = \frac{2}{M}\left[1 + \cos\left(2l\theta_k\right)\right]$$

for $l = 1, ..., M$, and equal to $1/M$ for $l = 0$. The variance of this sum is

$$\mathrm{Var}\left(\lambda_{k,N}\right) = 2 + \left\{-\frac{1}{M} + \frac{2}{M}\sum_{l=1}^{M}\cos\left(2l\theta_k\right)\right\}.$$

If $\{2k/N\} > \varepsilon > 0$, then

$$\frac{2}{M}\sum_{l=1}^{M}\cos\left(2l\theta_k\right) = \frac{1}{M}\frac{-1+\cos\theta+\cos M\theta-\cos(M+1)\theta}{1-\cos\theta} \leq \frac{C}{M\varepsilon}$$

which implies that $\mathrm{Var}\left(\lambda_{k,N}\right) \to 2$ as N and M grow to infinity. By the Central Limit Theorem, $\lambda_{k,N}$ converges to the Gaussian distribution with zero mean and variance 2. Since by asssumption the 4th moment of $\sqrt{M}a_k$ is uniformly bounded, therefore the Berry-Esseen bound is applicable and we can conclude that for those k, which satisfy the condition $\{2k/N\} > \varepsilon > 0$, inequality (3) is true. QED.

Using the lemma, we can write

$$\left|\frac{1}{N}\sum_{k=1}^{N}\left[\Pr\left\{\lambda_{k,N} \leq \sqrt{2}x\right\} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt\right]\right| \leq 2\varepsilon + \frac{c(\varepsilon)}{\sqrt{M}}.$$

Hence, for every $\delta > 0$ we can choose ε sufficiently small and then M sufficiently large and ensure that

$$\left|\frac{1}{N}\sum_{k=1}^{N}\Pr\left\{\lambda_{k,N} \leq \sqrt{2}x\right\} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt\right| \leq \delta.$$

This shows that as M and N grow, $E\left[G_N\left(\sqrt{2}x\right)\right]$ converges pointwise to the standard Gaussian law.

The next step is to investigate the variance/covariance structure of $b_{k,N} = 1_{(-\infty,x]}\left(\lambda_{k,N}\right)$. Hence, we need to estimate the expressions:

$$\begin{aligned} & E\left(b_{k,N}b_{l,N}\right) - E\left(b_{k,N}\right)E\left(b_{l,N}\right) \qquad (4)\\ = \; & \Pr\left\{\lambda_{k,N} \leq x; \lambda_{l,N} \leq x\right\} - \Pr\left\{\lambda_{k,N} \leq x\right\}\Pr\left\{\lambda_{l,N} \leq x\right\}. \end{aligned}$$

Again we can take an arbitrarily small $\varepsilon > 0$ and require that $\{(k-l)/N\} > \varepsilon$. An argument as in the proof of Proposition 10.3.2 on p. 344 in (Brockwell and Davis, 1991) shows that the distribution of $\left\{\lambda_{k,N}, \lambda_{l,N}\right\}$ is asymptotically Gaussian with the limit covariance matrix proportional to the identity matrix. Moreover, by assumption the 4th moment of $\sqrt{M}a_k$ is uniformly bounded, and therefore by the multivariate Berry-Esseen theorem, the speed of the convergence to the Gaussian law is bounded by $c(\varepsilon)/\sqrt{M}$. Hence, the difference in (4) can be bounded from above by $c(\varepsilon)/\sqrt{M}$. This implies that for sufficiently large M the covariance of $b_{k,N}$ and $b_{l,N}$ can be made arbitrarily small ($\leq \delta$) for all pairs k and l, such that $\{(k-l)/N\} > \varepsilon$. Therefore,

$$\mathrm{Var}\left\{\frac{1}{N}\sum b_{k,N}\right\} \leq \frac{1}{N^2}\left(2N\varepsilon + \delta N^2\right) \leq 2\delta$$

for all sufficiently large N. Since δ is an arbitrary, we can conclude that $\mathrm{Var}\left\{G_N(x)\right\} \to 0$ as N and M grow to infinity. This completes proof of Theorem 2.1.

3 Positive band to size ratio

Now let us ask what happens if the ratio of the band width to the matrix size remains constant as the size of the matrix grows.

Theorem 3.1. *Let X_N be a sequence of random symmetric N-by-N Toeplitz matrices with the band of width $M(N)$. Assume that the non-zero entries are independent real random variables with symmetric distributions such that $Ea_k^2 = 1/M$ and that for every n, $\sup_{k,N} E\left|\sqrt{M}a_k\right|^n < c(n) < \infty$. If both $M(N) \to \infty$ and $N/M \to c < \infty$ as $N \to \infty$, then there exists a distribution function $\Psi_c(x)$ such that*

$$E\int x^n dF_N(x) \to \int x^n d\Psi_c(x)$$

for every positive integer n. The distribution $\Psi_c(x)$ is non-Gaussian for all $c \neq 2\frac{1}{4}$.

As we will see in the process of proof, the variance of the limit distribution is $2-1/c$, and its fourth moment is $12 - \frac{32}{3}\frac{1}{c}$ if $c > 2$, and $-\frac{4}{3}c^2 + 8c - 4$, if $c \leq 2$.
It is very likely that the distribution is non-Gaussian for all c. In particular, numerical evaluations suggest that if $c = 2\frac{1}{4}$, then the 6th moment of the limit distribution is very close to $400/27 = 15 - 5/27$, and therefore, the limit distribution is not Gaussian. (This result is obtained by a Monte-Carlo evaluation of the volumes of polytopes $P(\pi)$ defined below. The standard error of the Monte-Carlo estimate of the 6th moment is 0.003.)
Before we start the proof, let us develop some machinery of the method of moments. Clearly,

$$E\int x^k dF_N(x) = \frac{1}{N}E\mathrm{Tr}\left(X^k\right)$$

We can write $N^{-1}\mathrm{Tr}\left(X^k\right)$ as follows:

$$\begin{aligned} N^{-1}\mathrm{Tr}\left(X^k\right) &= N^{-1}\sum_{a,b,\ldots,y,z} X_{ab}X_{bc}\ldots X_{yz}X_{za} \\ &= N^{-1}\sum_{a,b,\ldots,y,z} a_{b-a}a_{c-b}\ldots a_{z-y}a_{a-z} \\ &= N^{-1}\sum a_{x_1}a_{x_2}\ldots a_{x_{k-1}}a_{x_k}, \end{aligned} \tag{5}$$

where $a, b, \ldots, y, z$ are k numbers that take values between 1 and N. They show the row position of a particular entry X_{ab}. We will call these numbers "positions", so, for example, a is the starting position. The indices x_i are differences of positions. We will call them "step sizes" since they show how the "positions" of elements X_{ij} change. They also have an additional meaning since they show the diagonal in which the entry X_{ij} is located. In particular, the random variable a_{x_i} is non-zero only if the step size x_i is between $-M$ and M.
Write:

$$N^{-1}Tr\left(X^k\right) = N^{-1}\sum_a \sum_{x_1,\ldots,x_k} \theta\left(a, x_1, \ldots, x_k\right) a_{x_1}a_{x_2}\ldots a_{x_{k-1}}a_{x_k}, \tag{6}$$

where the summation is over all a between 1 and N, and all $x_1, \ldots, x_k$ such that every x_i is between $-M$ and M. Here $\theta\left(a, x_1, \ldots, x_k\right) = 1$ if a and $x_1, \ldots, x_k$ determine a valid sequence of $a, b, \ldots, y, z$, and $\theta\left(a, x_1, \ldots, x_k\right) = 0$ otherwise.

In the following we write $\overrightarrow{x}$ for the sequence $x_1, \ldots, x_k$. Let E_k denote the value of the limit $\lim_{N\to\infty} E\left[N^{-1}Tr\left(X^k\right)\right]$. Since the distributions of a_i are symmetric by assumption, therefore $E_k = 0$ if k is odd.

Consider E_{2k}. Let π denote a partition of the $2k$ indices of the elements x_i, such that indices i and j belong to the same block if and only if $x_i = x_j$ or $x_i = -x_j$. By an argument as in (Bryc et al., 2006) and (Hammond and Miller, 2005), it is easy to see that only partitions with two-element blocks give a non-negligible contribution to E_{2k} as M and N approach infinity and M/N approaches a constant. These partitions are pairings of $2k$ indices x_i. Moreover, (Bryc et al., 2006) and (Hammond and Miller, 2005) showed that another simplification is possible. It is as follows.

If x_i is in the same pair as x_j, then it must be that $x_i = x_j$ or $x_i = -x_j$. In the first case, we will write $\varepsilon_\tau(x_i) = \varepsilon_\tau\left(x_j\right) = 1$, where τ denote the pairing. In the second case, we will write $\varepsilon_\tau(x_i) = \varepsilon_\tau\left(x_j\right) = -1$.

It turns out that we can restrict attention to the case in which $\varepsilon_\tau(x_i) = -1$ for all i. The contribution of other sequences is negligible.

Let $\overrightarrow{x} = \left(x_1, \ldots, x_{2k}\right) \in \tau$ mean that the sequence $x_1, \ldots, x_{2k}$ agrees with the pairing τ and that the choice of ε_τ is $\varepsilon_\tau(x_i) = -1$ for all x_i.

We can think about the position a and the sequence $x_1, \ldots, x_{2k}$ as a random walk on the integer lattice $\mathbb{Z}^2$. The horizontal axis denotes time and the vertical axis denotes the position of a particle. The random walk starts at the point $(0, a)$ and at each step makes a jump of size x_i. Note that pair $\left(a, \overrightarrow{x}\right)$ is valid (i.e., $\theta\left(a, \overrightarrow{x}\right) = 1$) if and only if the random walk stays between the bounds of 1 and N.

In order to show non-Gaussianity of the limit distribution, it is sufficient to compute the limits of the second and the fourth moments. It turns out that for $c > c_0$ the limit kurtosis is greater than 3, and for $c < c_0$ the kurtosis is smaller than 3.

Lemma 3.2. *If* $\lim(N/M) = c$ *as both* M *and* N *approach* ∞*, then the variance of* μ_N *converges to* $2 - 1/c$ *as* $N \to \infty$.

Proof: We consider two cases. In the first case, $c \geq 2$. In this case, we need to consider three possible regions for the initial position a. The first one is $1 \leq a \leq M$. The second region is $M < a < N - M$, and the third one is $N - M \leq a \leq N$. If the initial position is in the first region, then x_1 can take values from $-a + 1$ to M, and $x_2 = -x_1$. Hence the total number of valid combinations $\left(a, \overrightarrow{x}\right)$ is asymptotically equivalent to

$$\int_0^M (M + a)\, da = \frac{3M^2}{2}.$$

By symmetry, this is also true for the third region.

For the second region, x_1 can take values from $-M$ to M, and the total number of combinations $\left(a, \overrightarrow{x}\right)$ is asymptotically equivalent to

$$\int_M^{N-M} 2M\, da \sim 2(c-2)M^2.$$

Hence the total number of all possible $\left(a, \overrightarrow{x}\right)$ for all regions is asymptotically equivalent to $(2c - 1)M^2$.

We should divide this by $NM \sim cM^2$. Hence, we obtain that the variance of μ_N converges to $2 - 1/c$.

In the second case, $c < 2$. Then again we have three possible regions for the initial position. The first one is $1 \le a \le N - M$, the second one is $N - M < a < M$, and the third one is $M \le a < N$. For the first region, x_1 can be between $-a + 1$ and M. Hence, the number of possible combinations $\left(a, \overrightarrow{x}\right)$ is

$$\begin{aligned} &\sim \int_0^{N-M} (M + a)\, da = (c-1) M^2 + \frac{(c-1)^2}{2} M^2 \\ &= \left(\frac{c^2 - 1}{2}\right) M^2. \end{aligned}$$

The same estimate holds for the third region by symmetry.

For the second region, x_1 can be between $-a + 1$ and $N - a$. Hence, we estimate the number of possible combinations $\left(a, \overrightarrow{x}\right)$ as

$$\int_{N-M}^{M} N\, da = c(2 - c) M^2.$$

Hence, the total number of all possible $\left(a, \overrightarrow{x}\right)$ is

$$\sim (2c - 1) M^2.$$

After dividing by MN, we find that the variance of μ_N converges to $2 - 1/c$. QED.

Lemma 3.3. *If $c \ge 2$, then the fourth moment of the measure μ_N converges to*

$$12 - \frac{32}{3}\frac{1}{c}.$$

If $c \in [1, 2]$, then it converges to

$$-\frac{4}{3}c^2 + 8c - 4.$$

Proof: Let $c \ge 4$. We have five essentially different regions for the initial position a: 1) $1 \le a \le M$; 2) $M < a \le 2M$; 3) $2M < a < N - 2M$; 4) $N - 2M \le a < N - M$, and 5) $N - M \le a < N$. In addition, we have 3 different pairings:

$$x_1 x_2 x_3 x_4,$$

$$x_1 x_2 x_3 x_4,$$

and

$$x_1 x_2 x_3 x_4.$$

The proof proceeds by straightforward computations in each of the possible cases. We start with the first pairing. In region 1, the variables x_1 and x_3 can take values from $-a + 1$ to M, and the variables x_2 and x_4 are determined by x_1 and x_3. In regions 2, 3, and 4, the variables x_1 and x_3

can take values from $-M$ to M. And region 5 is symmetric to region 1. Hence, the total number of $\left(a,\overrightarrow{x}\right)$ combinations is asymptotically equivalent to

$$2\int_0^M (M+a)^2\,da + \int_M^{N-M} (2M)^2\,da = \left(4c - \frac{10}{3}\right)M^3.$$

Next, we consider the second pairing. In region 1, the variable x_1 can take values from $-a+1$ to M. If $x_1 \le 0$, then x_2 have to be between $-x_1-a+1$ and M; and if $x_1 > 0$, then x_2 have to be between $-a+1$ and M. Hence the number of $\left(a,\overrightarrow{x}\right)$ combinations is asymptotically equivalent to

$$\int_0^M\int_{-a}^0 \left(M+x_1+a\right)dx_1da + \int_0^M\int_0^M (M+a)\,dx_1da = \frac{13}{6}M^3.$$

In region 2, the variable x_1 can take values from $-M$ to M. If $x_1 > -a+M$, then the variable x_2 can take values between $-M$ and M, and if x_1 is between $-M$ and $-a+M$, then the variable x_2 can take values between $-a-x_1+1$ and M. Hence the number of possible $\left(a,\overrightarrow{x}\right)$ combinations is asymptotic to

$$\int_M^{2M}\int_{-M}^{M-a} \left(M+x_1+a\right)dx_1da + \int_M^{2M}\int_{M-a}^{M} (2M)\ dx_1da = \frac{23}{6}M^3.$$

In region 3, the variables x_1 and x_2 can take values from $-M$ to M. Hence, the number of possible $\left(a,\overrightarrow{x}\right)$ combinations is asymptotic to

$$\int_{2M}^{N-2M} (2M)^2\,da \sim 4(c-4)M^3.$$

Regions 4 and 5 are symmetric to regions 2 and 1, respectively. Hence the total number of possible $\left(a,\overrightarrow{x}\right)$ combinations for the second pairing is

$$\left(\frac{13}{3}+\frac{23}{3}+4c-16\right)M^3 = (4c-4)M^3.$$

Finally, we consider the third pairing. In region 1, the variable x_1 can take values from $-a+1$ to M. If $x_1 \le M-a$, then x_2 can take values between $-a-x_1+1$ and M, and if $x_1 > M-a$, then x_2 is between $-M$ and M. Hence the number of (a,x) combinations is

$$\int_0^M\int_{-a}^{M-a} \left(M+a+x_1\right)dx_1da + \int_0^M\int_{M-a}^{M} (2M)\,dx_1da = \frac{5}{2}M^3.$$

In region 2, the variable x_1 can take values from $-M$ to M. If $x_1 \le M-a$, then x_2 is between $-a-x_1+1$ and M. If $x_1 > M-a$, then x_2 is between $-M$ and M. Hence, the number of possible (a,x) combinations is

$$\int_M^{2M}\int_{-M}^{M-a} \left(M+a+x_1\right)dx_1da + \int_M^{2M}\int_{M-a}^{M} (2M)\,dx_1da = \frac{23}{6}M^3.$$

In region 3, the variables x_1 and x_2 can take values from $-M$ to M, hence the number of possible (a,x) combinations is

$$\int_{2M}^{N-2M} (2M)^2\,da \sim 4(c-4)M^3.$$

(Note that the calculations for regions 2 and 3 are the same as for the second pairing, while the calculation for region 1 is different.) Regions 4 and 5 are symmetric to regions 2 and 1, respectively. Hence, the total number of possible (a, x) combinations for the third pairing is asymptotic to

$$\left(5 + \frac{23}{3} + 4c - 16\right) M^3 = \left(4c - \frac{10}{3}\right) M^3.$$

The total number of all possible (a, x) combinations for all pairings is

$$\left(4c - \frac{10}{3} + 4c - \frac{12}{3} + 4c - \frac{10}{3}\right) M^3 = \left(12c - \frac{32}{3}\right) M^3.$$

Dividing this by NM^2, we obtain the limit of the fourth moment of the distribution μ_N is

$$12 - \frac{32}{3}\frac{1}{c}.$$

Similar calculations can be done for the cases when $c \in [3,4]$, $c \in [2,3]$, and $c \in [1,2]$. The limits of the fourth moment in these cases are $12 - \frac{32}{3c}$ in the first two cases, and $-\frac{4}{3}c^2 + 8c - 4$ in the third one. QED.

Lemma 3.4. *Let assumptions of Theorem 3.1 hold. Then for all integer $k \geq 0$, the expression $E\left[N^{-1} Tr\left(X^k\right)\right]$ has a limit as $N \to \infty$, $M \to \infty$, and $M/N \to c > 0$.*

Proof: We know that for odd k, $E\left[N^{-1}\mathrm{Tr}\left(X^k\right)\right] = 0$. We have also computed the limit for $k = 2$ and 4 in the lemmas above. For arbitrary even k, we have the following argument. Let $k = 2l$. We have shown above that to compute $E\left[N^{-1}\mathrm{Tr}\left(X^k\right)\right] = 0$, we can restrict attention to those terms in the expansion of the trace that correspond to pairings of $2l$ indices. All other terms give asymptotically negligible contribution.

Consider a pairing π of $2l$ indices. We will think about it as a product of l disjoint transpositions in the symmetric group S_{2l} operating on numbers $\{1, 2, \ldots, 2l\}$. For this pairing we define a $2l$-by-l matrix M_π in the following way: Let us order the transpositions in π according to the smallest elements in these transpositions. For example, for $\pi = (16)(23)(45)$, the first transposition is $\tau_1 = (16)$, the second one is $\tau_2 = (23)$, and the third one is $\tau_3 = (45)$. Let the k-th transposition be $\tau_k = (i_k, j_k)$ where $i_k < j_k$. Then define matrix elements $M_{i_k k} = 1$ and $M_{j_k k} = -1$. Set all other entries in the k-th column of matrix M_π equal to zero. This defines the matrix M_π.

Next define another matix $\widetilde{M}_\pi$ as follows. Let T is a $2l$-by-$2l$ lower-triangular matrix such that $T_{ij} = 1$ if $i \geq j$. Define $\widetilde{M}_\pi := TM_\pi$. Intuitively, the first row of $\widetilde{M}_\pi$ is the same as the first row of M_π. The second row of $\widetilde{M}_\pi$ is the sum of the first and the second row of M_π, the third row of $\widetilde{M}_\pi$ is the sum of the first, second, and the third rows of M_π, and so on.

Consider the space $\mathbb{R}^{l+1}$, where each point has coordinates $\left(a, x_1, ..., x_l\right)$. Let $P_N(\pi)$ be the convex polytope in $\mathbb{R}^{l+1}$ defined by the following set of $3l$ inequalities (not all of which are independent from others):

$$\begin{aligned} \mathbf{1}_{2l} &\leq a\mathbf{1}_{2l} + \widetilde{M}_\pi \overrightarrow{x} \leq N\mathbf{1}_{2l}, \\ -M\mathbf{1}_l &\leq \overrightarrow{x} \leq M\mathbf{1}_l, \end{aligned}$$

where $\mathbf{1}_k$ denotes vector with k components all equal to 1.

It is easy to see that the meaning of these inequalities is that the positions of a random walk with the starting position a and steps $x_1, ..., x_l$ are all between 1 and N and that the steps are restricted

to be between $-M$ and M. Note that these are the same requirements that we need to impose to make sure that $\theta\left(a, \overrightarrow{x}\right) = 1$ in (6). This implies that the contribution of terms corresponding to the pairing π towards $E\left[N^{-1}\mathrm{Tr}\left(X^k\right)\right]$ equals the number of integer points that belong to the polytope $P_N(\pi)$ multiplied by $N^{-1}M^{-l}$.
Let $\widetilde{P}_N(\pi)$ be the polytope $P_N(\pi)$ rescaled by $1/M$. Then, as $N \to \infty$ the polytopes $\widetilde{P}_N(\pi)$ converge to a convex polytope $P(\pi)$, which is defined by inequalities:

$$\begin{aligned} \mathbf{0}_{2l} &\leq a\mathbf{1}_{2l} + \widetilde{M}_\pi \overrightarrow{x} \leq c\mathbf{1}_{2l}, \\ -\mathbf{1}_l &\leq \overrightarrow{x} \leq \mathbf{1}_l. \end{aligned}$$

Hence the number of terms inside the polytope $P_N(\pi)$ can be estimated as M^{l+1} multiplied by the volume of $P(\pi)$ and the error of this estimate is $O\left(M^l\right)$. Therefore,

$$E\left[N^{-1}\mathrm{Tr}\left(X^k\right)\right] \to \frac{1}{c}\sum_{\pi\in P_2(2l)} \mathrm{Vol}(P(\pi)).$$

QED.
Proof of Theorem 3.1: Lemma 3.4 shows the convergence of the expectations of the moments of the eigenvalue distributions $F_N(x)$. The limits are smaller or equal than the moments of a Gaussian distribution which we would obtained if we set all $\theta\left(a, \overrightarrow{x}\right)$ equal to 1 in formula (6). This implies the first claim of the theorem. Lemmas 3.2 and 3.3 imply that the kurtosis of the limit distribution is different from 3 for $c \neq 2\frac{1}{4}$. Hence, if $c \neq 2\frac{1}{4}$, then the limiting distribution is not Gaussian. QED.

References

A. Bose and J. Mitra. Limiting spectral distribution of a special circulant. *Statistics and Probability Letters*, 60:111–120, 2002. MR1945684

A. Bose and A. Sen. Spectral norm of random large dimensional noncentral Toeplitz and Hankel matrices. *Electronic Communications in Probability*, 12:29–35, 2007. Paging changed to 21-27 on journal site. MR2284045

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Verlag, second edition, 1991. MR1093459

W. Bryc, A. Dembo, and T. Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Annals of Probability*, 34:1–38, 2006. MR2206341

S. Chatterjee. Fluctuations of eigenvalues and second order Poincare inequalities. *Probability Theory and Related Fields*, 143:1–40, 2009. MR2449121

R. M. Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Transactions on Information Theory*, 18:725–730, 1972. MR0453214

C. Hammond and S. J. Miller. Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *Journal of Theoretical Probability*, 18:537–566, 2005. MR2167641

A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20:37–39, 1953. MR0052379

A. Massey, S. J. Miller, and J. Sinsheimer. Distribution of eigenvalues of real symmetric palindromic T oeplitz matrices and circulant matrices. *Journal of Theoretical Probability*, 20:637–662, 2007.

M. W. Meckes. On the spectral norm of a random Toeplitz matrix. *Electronic Communications in Probability*, 12:315–325, 2007. MR2342710

H. F. Trotter. Eigenvalue distribution of large hermitian matrices; Wigner semi-circle law and a theorem of Kac, Murdock, and Szego. *Advances in Mathematics*, 54:67–82, 1984. MR0761763