DISTRIBUTION OF EIGENVALUES FOR THE ENSEMBLE OF REAL SYMMETRIC TOEPLITZ MATRICES

CHRISTOPHER HAMMOND AND STEVEN J. MILLER

ABSTRACT. Consider the ensemble of real symmetric Toeplitz matrices whose entries are i.i.d. random variables from a fixed probability distribution p of mean 0, variance 1, and finite moments of all order. The limiting spectral measure (the density of normalized eigenvalues) converges weakly to a new universal distribution with unbounded support, independent of p. This distribution's moments are almost those of the Gaussian's, and the deficit may be interpreted in terms of obstructions to Diophantine equations; the unbounded support follows from a nice application of the Central Limit Theorem. With a little more work, we obtain almost sure convergence. An investigation of spacings between adjacent normalized eigenvalues looks Poissonian, and not GOE. A related ensemble (real symmetric palindromic Toeplitz matrices) appears to have no Diophantine obstructions, and the limiting spectral measure's first nine moments can be shown to agree with those of the Gaussian; this will be considered in greater detail in a future paper.

1. INTRODUCTION

One of the central problems in Random Matrix Theory is as follows: consider some ensemble of matrices A with probabilities $p(A)$. As $N \to \infty$, what can one say about the density of normalized eigenvalues? For real symmetric matrices, where the entries are i.i.d.r.v. from suitably restricted probability distributions, the limiting distribution is the semi-circle (see [Wig, Meh]). Note this ensemble has $\frac{N(N+1)}{2}$ independent parameters ($a_{ij}, i \leq j$). For matrix ensembles with fewer degrees of freedom, different limiting distributions arise (for example, McKay [McK] proved d-regular graphs are given by Kesten's Measure). By examining ensembles with fewer than N^2 degrees of freedom, one has the exciting potential of seeing new, universal distributions. In this paper we study ensembles of real symmetric Toeplitz matrices.

Definition 1.1. *A Toeplitz matrix is A is of the form*

$$A = \begin{pmatrix} b_0 & b_1 & b_2 & \cdots & b_{N-1} \\ b_{-1} & b_0 & b_1 & \cdots & b_{N-2} \\ b_{-2} & b_{-1} & b_0 & \cdots & b_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{1-N} & b_{2-N} & b_{3-N} & \cdots & b_0 \end{pmatrix}, \quad a_{ij} = b_{j-i}. \tag{1}$$

We investigate real symmetric Toeplitz matrices whose entries are i.i.d. random variables from a fixed probability distribution p of mean 0, variance 1, and finite moments of all order. The probability density at A is $p(A) = \prod_{i=0}^{N-1} p(b_i)$.

From $\mathrm{Trace}(A^2) = \sum_i \lambda_i^2(A)$ and the Central Limit Theorem, we see that the eigenvalues of A are of order $\sqrt{N}$. This suggests the appropriate scale for normalizing the eigenvalues. As the main diagonal is constant, all b_0 does is shift each eigenvalue by b_0. Therefore, we only consider the case where the main diagonal vanishes.

To each $N \times N$ Toeplitz matrix A we attach a spacing measure by placing a point mass of size $\frac{1}{N}$ at each normalized eigenvalue $\lambda_i(A)$:

$$\mu_{A,N}(x)dx \;=\; \frac{1}{N} \sum_{i=1}^{N} \delta\left(x - \frac{\lambda_i(A)}{\sqrt{N}}\right) dx. \tag{2}$$

Thus, for any Borel set $G \subset \mathbb{R}$, if $I(x \in G) = 1$ for $x \in G$ and 0 otherwise, then

$$\mu_{A,N}(G) \;=\; \frac{1}{N} \sum_{i=1}^{N} I\left(\frac{\lambda_i(A)}{\sqrt{N}} \in G\right). \tag{3}$$

The kth moment of $\mu_{A,N}(x)$ is

$$M_k(A, N) \;=\; \frac{1}{N^{\frac{k}{2}+1}} \sum_{i=1}^{N} \lambda_i^k(A). \tag{4}$$

Definition 1.2. *Let $M_k(N)$ be the average of $M_k(A, N)$ over the ensemble, with each A weighted by its distribution. If p is continuous, we weight A by $\prod_{k=1}^{N-1} p(b_k)db_k$. Let $M_k = \lim_{N \to \infty} M_k(N)$.*

Our main result is the following: for each k, $M_k(N)$ converges to the moments of a new universal distribution, independent of p. This distribution is denoted the limiting spectral measure (the density of normalized eigenvalues). In Theorem 5.4 we prove the limiting spectral measure converges weakly to a new distribution of unbounded support; in Theorem 6.1 we show that if p is an even function, then we have almost sure convergence.

The new distribution looks Gaussian, and numerical simulations and heuristics seemed to support such a conjecture. A more detailed analysis, however, reveals that while M_k agrees with the Gaussian moments for odd k and $k = 0, 2$, the other even moments are less than the Gaussian's. We calculate the odd and first few moments in §2; this involves proving some combinatorial identities and analyzing some Diophantine equations. Using simple combinatorics we give some upper bounds for the higher moments in §3; using the Central Limit Theorem we prove strong enough lower bounds in §4 to show the limiting spectral measure has unbounded support.

We now highlight the proof. By the Eigenvalue Trace Lemma,

$$\sum_{i=1}^{N} \lambda_i^k(A) \;=\; \mathrm{Trace}(A^k) \;=\; \sum_{1 \leq i_1, \ldots, i_k \leq N} a_{i_1,i_2} a_{i_2,i_3} \cdots a_{i_k,i_1}. \tag{5}$$

As our Toeplitz matrices are constant along diagonals, depending only on $|i_m - i_n|$, we have

$$M_k(N) \;=\; \mathbb{E}[M_k(A, N)] \;=\; \frac{1}{N^{\frac{k}{2}+1}} \sum_{1 \leq i_1, \ldots, i_k \leq N} \mathbb{E}(b_{|i_1-i_2|} b_{|i_2-i_3|} \cdots b_{|i_k-i_1|}), \tag{6}$$

where by $\mathbb{E}(\cdots)$ we mean averaging over the Toeplitz ensemble with each matrix A weighted by its probability of occurring, and the b_j are i.i.d.r.v. drawn from $p(x)$.

We show that as $N \to \infty$, the above sums vanish for k odd, and converge independent of p for k even to numbers M_k bounded by the moments of the Gaussian. By showing $\mathbb{E}[|M_k(A, N) - M_k(N)|^m]$ is small for $m = 2$ ($m = 4$), we obtain weak (almost sure) convergence.

Remark 1.3. *This problem was first posed by Bai [Bai], where he also asked similar questions about Hankel and Markov matrices. The methods of this paper should be applicable to at least the Hankel case as well. Bose-Chatterjee-Gangopadhyay [BCG] and Bryc-Dembo-Jiang [BDJ] have independently observed that the limiting distribution is not Gaussian. Using a more probabilistic formulation, [BDJ] have calculated the moments using uniform variables and interpreting results as volumes of solids related to Eulerian numbers. We have independently found the same numbers, but through Diophantine analysis. The novelty of our approach is that we can interpret the deviations from the Gaussian in terms of obstructions to Diophantine equations, and obtain*

significantly sharper estimates on the growth of the moments. Further, this framework seems ideally suited for investigating related Toeplitz ensembles (see §7).

2. Determination of the Moments

2.1. $k = 0, 2$ and k odd.

Theorem 2.1. *Assume p has mean zero, variance one and finite higher moments. Then $M_0 = 1$, $M_2 = 1$ and $M_{2m+1} = 0$.*

Proof. For all N, $M_0(A, N) = M_0(N) = 1$. For $k = 2$, we have

$$M_2(N) \;=\; \frac{1}{N^2} \sum_{1 \leq i_1, i_2 \leq N} \mathbb{E}(b_{|i_1-i_2|} b_{|i_2-i_1|}) \;=\; \frac{1}{N^2} \sum_{1 \leq i_1, i_2 \leq N} \mathbb{E}(b_{|i_1-i_2|}^2). \tag{7}$$

As we have drawn the bs from a variance one distribution, the expected value above is 1 if $i_1 \neq i_2$ and 0 otherwise. Thus $M_2(N) = \frac{N^2-N}{N^2} = 1 - \frac{1}{N}$, so $M_2 = 1$. Note there are two degrees of freedom. We can choose $b_{|i_1-i_2|}$ to be on any diagonal. Once we have specified the diagonal, we can then choose i_1 freely, which now determines i_2.

For k odd, we must have at least one b_j occurring to an odd power. If one occurs to the first power, as the expected value of a product of independent variables is the product of the expected values, these terms contribute zero. Thus the only contribution to an odd moment come when each b_j in the expansion occurs at least twice, and at least one occurs three times. Hence, if $k = 2m + 1$, we see we have at most $m + 1$ degrees of freedom, this coming from the case $b_{j_1}^3 b_{j_2}^2 \cdots b_{j_m}^2$. There are m different factors of b, and then we can choose any one subscript. Once we have specified a subscript and which diagonals we are on, the remaining subscripts are determined. As all moments are finite, we find

$$M_{2m+1}(N) \;\ll_m\; \frac{1}{N^{\frac{2m+1}{2}+1}} N^{m+1} \;\ll_m\; \frac{1}{\sqrt{N}}, \tag{8}$$

where $f(x) \ll_m g(x)$ means for all x sufficiently large, there is a constant depending only on m such that $|f(x)| \leq C_m g(x)$. As $N \to \infty$ we see $M_{2m+1}(N) \to 0$, completing the proof. $\qquad\square$

2.2. Bounds for the Even Moments.

Theorem 2.2. *Assume p has mean zero, variance one and finite higher moments. Then*

$$M_{2m}(N) \;\leq\; (2m-1)!! + O_m\left(\frac{1}{N}\right). \tag{9}$$

Proof. We proceed in stages in calculating $M_{2m}(N)$. First we bound $M_{2m}(N)$ by $2^m \cdot (2m-1)!! + O_m\left(\frac{1}{N}\right)$, where $(2m-1)!!$ is the $2m^{th}$ moment of the Gaussian. We then show that the factor of 2^m can be removed. Later in Theorem 3.1 we show a strict inequality holds. While it is possible to give a more concise proof, the arguments below are useful in bounding the size of the moments and in studying generalized Toeplitz ensembles (see §7 and [MS]).

By (6),

$$M_{2m}(N) \;=\; \frac{1}{N^{m+1}} \sum_{1 \leq i_1, \cdots, i_{2m} \leq N} \mathbb{E}(b_{|i_1-i_2|} b_{|i_2-i_3|} \cdots b_{|i_{2m}-i_1|}). \tag{10}$$

If any b_j occurs to the first power, its expected value is zero and there is no contribution. Thus the b_js must be matched at least in pairs. If any b_j occurs to the third or higher power, there are less than $m+1$ degrees of freedom, and there will be no contribution in the limit. In the main term the b_js are therefore matched in pairs, say $b_{|i_m-i_{m+1}|} = b_{|i_n-i_{n+1}|}$. Let $x_m = |i_m-i_{m+1}| = |i_n-i_{n+1}|$. There are two possibilities:

$$i_m - i_{m+1} \;=\; i_n - i_{n+1} \quad \text{or} \quad i_m - i_{m+1} \;=\; -(i_n - i_{n+1}). \tag{11}$$

Let $x_1, \ldots, x_m$ be the values of the $|i_j - i_{j+1}|$s, and let $\epsilon_1, \ldots, \epsilon_m$ be the choices of sign in (11). Define $\widetilde{x}_1 = i_1 - i_2$, $\widetilde{x}_2 = i_2 - i_3, \ldots, \widetilde{x}_{2m} = i_{2m} - i_1$. Note for each j exactly one $\widetilde{x}_m$ is $\eta_j x_j$ and

exactly one $\widetilde{x}_m$ is $\epsilon_j \eta_j x_j$, where $\eta_j = \pm 1$. We have

$$
\begin{aligned}
i_2 &= i_1 - \widetilde{x}_1 \\
i_3 &= i_1 - \widetilde{x}_1 - \widetilde{x}_2 \\
&\;\;\vdots \\
i_1 &= i_1 - \widetilde{x}_1 - \cdots - \widetilde{x}_{2m}.
\end{aligned}
\tag{12}
$$

By the final relation for i_1, we find

$$
\widetilde{x}_1 + \cdots + \widetilde{x}_{2m} = \sum_{j=1}^{m} (1 + \epsilon_j)\eta_j x_j = 0.
\tag{13}
$$

If any $\epsilon_j = 1$, then the x_j are not linearly independent, and we have fewer than $m + 1$ degrees of freedom; these terms contribute $O_m\left(\frac{1}{N}\right)$ in the limit. Therefore the only valid assignment is to have all $\epsilon_j = -1$. There are now 2^m possible choices (from the η_j), and $m + 1$ degrees of freedom. Hence $M_{2m}(N) \le 2^m \cdot N^{m+1} + O_m\left(\frac{1}{N}\right)$ as the b_js are matched in pairs and the second moment of p is 1; the error term is from the matchings that aren't all pairs. We eliminate 2^m by changing our viewpoint.

We have $m + 1$ degrees of freedom. We match our differences in m pairs. Choose i_1 and i_2. We now look at the freedom to choose the remaining indices i_j. Once i_1 and i_2 are specified, we have $i_1 - i_2$, and a later difference must be the negative of that. If $i_2 - i_3$ is matched with $i_1 - i_2$, then i_3 is uniquely determined (because it must give the opposite of the earlier difference). If not, i_3 is a new variable and there are N choices for i_3. Now we look at i_4. If $i_3 - i_4$ is matched with an earlier difference, then the sign of its difference is known, and i_4 is uniquely determined; if this difference belongs to a new pair not previously encountered, then i_4 is a new variable and free. Proceeding in this way, we note that if we encounter i_n such that $i_{n-1} - i_n$ is paired with a previous difference, the sign of its difference is specified and i_n is uniquely determined; otherwise, if this is a difference of a new pair, i_n is a free variable with at most N choices. Thus we see there are at most N^{m+1} choices (note that it is possible that not all choices work, as for example the final difference $i_{2m} - i_1$ is determined before we get there because of earlier choices). □

Remark 2.3. *Having $m + 1$ degrees of freedom does not imply each term contributes fully – we will see there are Diophantine obstructions which bound the moments away from the Gaussian's. However, each matching contributes at most 1, and there are $(2m - 1)!!$ matchings. It is often convenient to switch viewpoints from having $m + 1$ indices to having the m positive subscript differences and one index (say i_d) as our degrees of freedom; note once we specify the positive differences and any index, which difference in a pair is positive and which is negative is determined.*

2.3. **The Fourth Moment.** The fourth moment calculation highlights the Diophantine obstructions encountered, which bound the moments away from the Gaussian (whose fourth moment is 3). Let p_4 be the fourth moment of p.

Theorem 2.4 (Fourth Moment).

$$
M_4(N) = \frac{8}{3} + O\left(\frac{1}{N}\right).
\tag{14}
$$

Proof. By (6),

$$
M_4(N) = \frac{1}{N^3} \sum_{1 \le i_1, i_2, i_3, i_4 \le N} \mathbb{E}(b_{|i_1 - i_2|} b_{|i_2 - i_3|} b_{|i_3 - i_4|} b_{|i_4 - i_1|})
\tag{15}
$$

Let $x_j = |i_j - i_{j+1}|$. If any b_{x_j} occurs to the first power, its expected value is zero. Thus either the x_j are matched in pairs (with different values), or all four are equal (in which case they are still matched in pairs). From Theorem 2.2, the signs ϵ_j are all negative in the pairings. There are three possible matchings, and each matching contributes at most 1; if each matching

contributed 1 then the fourth moment would be 3, the same as the Gaussian's. By symmetry (simply relabel), we see the contribution from $x_1 = x_2$, $x_3 = x_4$ is the same as the contribution from $x_1 = x_4$, $x_2 = x_3$. The reason is both of these cases are adjacent pairings; the neighbors of x_1 are x_2 and x_4, and in each case everything is matched with a neighbor.

If $x_1 = x_2$, $x_3 = x_4$, by Theorem 2.2 we have

$$i_1 - i_2 = -(i_2 - i_3) \quad \text{and} \quad i_3 - i_4 = -(i_4 - i_1). \tag{16}$$

Thus $i_1 = i_3$ and i_2 and i_4 are arbitrary. Using these three variables as our independent degrees of freedom, we see there are N^3 such quadruples. Almost all of these will have $x_1 \neq x_3$ and contribute $\mathbb{E}(b_{x_1}^2 b_{x_3}^2) = 1$. Given i_1 and i_2, $N - 1$ choices of i_4 yield $x_1 \neq x_3$, and one choice yields the two equal. Letting p_4 denote the fourth moment of p, we see this case contributes

$$\frac{1}{N^3} \left(N^2(N - 1) \cdot 1 + N^2 \cdot p_4 \right) = 1 - \frac{1}{N} + \frac{p_4}{N} = 1 + O\left(\frac{1}{N}\right). \tag{17}$$

The third possibility is for $x_1 = x_3$ and $x_2 = x_4$. Non-adjacent pairing leads to Diophantine obstructions, which decrease the contribution to the moment. Again by Theorem 2.2 we have

$$i_1 - i_2 = -(i_3 - i_4) \quad \text{and} \quad i_2 - i_3 = -(i_4 - i_1). \tag{18}$$

This yields

$$i_1 = i_2 + i_4 - i_3, \quad i_1, i_2, i_3, i_4 \in \{1, \ldots, N\}. \tag{19}$$

The fact that each $i_j \in \{1, \ldots, N\}$ is what leads to the Diophantine obstructions. When $x_1 = x_2$ and $x_3 = x_4$, we saw we had three independent variables, and $N^3 + O(N^2)$ choices that were mutually consistent. Now it is possible for choices of i_2, i_3 and i_4 to lead to impossible values for i_1. For example, if $i_2, i_4 \geq \frac{2N}{3}$ and $i_3 < \frac{N}{3}$, we see $i_1 > N$. Thus, there are at most $(1 - \frac{1}{27})N^3$ valid choices; we have lost a positive percent of triples. This is enough to show the Gaussian moment is strictly greater; later in Theorem 3.1 we will see that if there is one moment less than the Gaussian, all larger even moments are also smaller.

Note the number of tuples with $x_1 = x_2 = x_3 = x_4$ is $O(N^2)$, and these contribute $O\left(\frac{1}{N}\right)$. As this is a lower order term, in the arguments below we may assume the two pairs do not have the same difference. The proof is completed by the following lemma, which shows the case when $x_1 = x_3$ and $x_2 = x_4$ contributes $\frac{2}{3}$ in the limit to the fourth moment; this refines our upper bound of $\frac{26}{27}$. $\square$

Lemma 2.5. Let $I_N = \{1, \ldots, N\}$. Then $\#\{x, y, z \in I_N : 1 \leq x + y - z \leq N\} = \frac{2}{3}N^3 + \frac{1}{3}N$.

Proof. Say $x + y = S \in \{2, \ldots, 2N\}$. For $2 \leq S \leq N$, there are $S - 1$ choices of z such that $1 \leq x + y - z \leq N$, and for $S \geq N + 1$, there are $2N - S + 1$ choices. Similarly, the number of $x, y \in I_N$ with $x + y = S$ is $S - 1$ if $S \leq N + 1$ and $2N - S + 1$ otherwise. The number of triples is therefore

$$\sum_{S=2}^{N} (S - 1)^2 + \sum_{S=N+1}^{2N} (2N - S + 1)^2 = \frac{2}{3}N^3 + \frac{1}{3}N. \tag{20}$$

$\square$

2.4. **Sixth and Eight Moments.** Any even moment can be explicitly determined by brute-force calculation, though deriving exact formulas as $k \to \infty$ requires handling involved combinatorics. To calculate the $2k^{\text{th}}$ moment, we consider $2k$ points on the unit circle, and see how many different shapes we get when we match in pairs. Direct computation gave $M_6(N) = 11$ (compared to the Gaussian's 15), and $M_8(N) = 64\frac{4}{15}$ (compared to the Gaussian's 105). For the sixth moment, there are five non-isomorphic configurations:
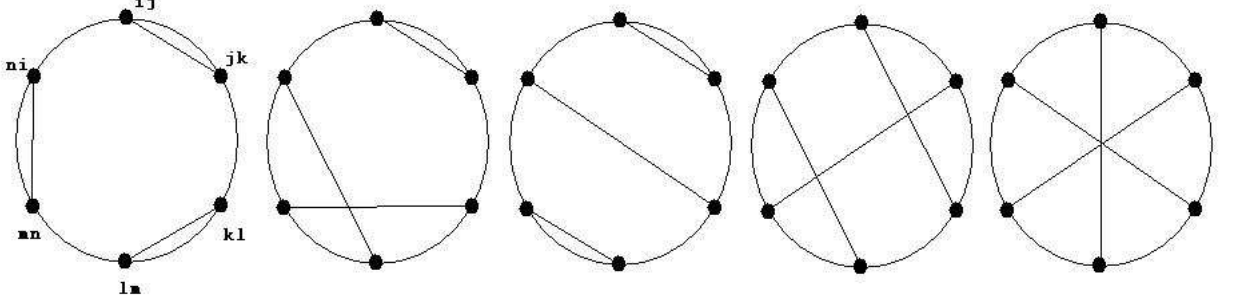
These occur $2, 6, 3, 3$ and 1 time, contributing $1, \frac{2}{3}, 1, \frac{1}{2}$, and $\frac{1}{2}$ (respectively); these correspond to the $15 = (6-1)!!$ pairings. For the eighth moment, the smallest contribution is $\frac{1}{4}$, coming from the matching $x_1 = x_3$, $x_2 = x_4$, $x_5 = x_7$, $x_6 = x_8$. It seems the more crossings (in some sense), the greater the Diophantine obstructions and the smaller the contribution.

To show the M_k converge to a new limiting distribution, we first must show the M_k are finite. The following proof was suggested to us by David Farmer.

Theorem 2.6. *If p has mean zero and variance one then for all k, $M_k = \lim_{N \to \infty} M_k(N)$ exists and is finite.*

Proof. From the arguments in Theorems 2.2 and 2.4, it suffices to investigate the case of even moments; moreover, the main term is when the b_js are matched in pairs with all signs $\epsilon_j = -1$. There are $(2k-1)!!$ matchings. For any matching $\mathcal{M}$, we have a system of linear equations with $i_1, \ldots, i_{2k} \in \{1, \ldots, N\}$ and $k+1$ degrees of freedom. Let $z_\ell = \frac{i_\ell}{N} \in \{\frac{1}{N}, \ldots, 1\}$. The linear equations now determine a nice region in the $(k+1)$-dimensional unit cube. As $N \to \infty$ we obtain to first order the volume of this region, which is finite. Unfolding back to the i_ℓ, we obtain $M_{2k}(\mathcal{M})N^{k+1} + O_k\left(N^k\right)$, where $M_{2k}(\mathcal{M})$ is the volume associated to this matching. Summing over all matchings gives $M_{2k}N^{k+1} + O_k\left(N^k\right)$. $\qquad\square$

The above proof it is similar to the proof in [BDJ]. Arguing along these lines, they interpret the M_k in terms of volumes of Euler solids. We prefer to view the problem in terms of Diophantine obstructions, which is very useful for determining growth rates (see for instance Theorem 3.2), as well as in studying related ensembles (see §7).

In §3 and §4 we obtain upper and lower bounds for the moments. Then in §5 and §6 we prove weak and almost sure convergence.

3. Upper Bounds of High Moments

3.1. **Weak Upper Bound of High Moments.**

Theorem 3.1. *For $2k \geq 4$, $\lim_{N \to \infty} M_{2k}(N) < (2k-1)!!$.*

Proof. We first show that if a matching contributes less than 1 for some moment, it lifts to matchings for higher moments that also contribute less than 1. Say we have such a matching on $b_{|i_1-i_2|} \cdots b_{|i_{2k_0}-i_1|}$ giving less than 1. We extend this to a pairing on $2k > 2k_0$ as follows. We now have

$$b_{|i_1-i_2|} \cdots b_{|i_{2k_0-1}-i_{2k_0}|} b_{|i_{2k_0}-i_{2k_0+1}|} b_{|i_{2k_0+1}-i_{2k_0+2}|} \cdots b_{|i_{2k-1}-i_{2k}|} b_{|i_{2k}-i_1|}. \tag{21}$$

In groups of two, pair adjacent neighbors from $b_{|i_{2k_0+1}-i_{2k_0+2}|}$ to $b_{|i_{2k-1}-i_{2k}|}$. This implies $i_{2k_0} = i_{2k_0+2} = \cdots = i_{2k}$. Looking at the first $2k_0 - 1$ and the last factor gives

$$b_{|i_1-i_2|} \cdots b_{|i_{2k_0-1}-i_{2k_0}|} b_{|i_{2k}-i_i|} \;=\; b_{|i_1-i_2|} \cdots b_{|i_{2k_0-1}-i_{2k_0}|} b_{|i_{2k_0}-i_1|}. \tag{22}$$

Now pair these as in the matching which gave less than 1, and we see this matching contributes less than 1 as well. By Theorem 2.4 we know there exists a matching from the fourth moment which contributes $\frac{2}{3} < 1$, which completes the proof. $\qquad\square$

3.2. **Strong Upper Bound of High Moments.** In general, the further away one moment is from the Gaussian, the more one can say about higher moments. While we do not have exact asymptotics, we can show

Theorem 3.2. $\lim_{k \to \infty} \frac{M_{2k}}{(2k-1)!!} = 0$.

Proof. We show that for any positive integer c and k sufficiently large, as $N \to \infty$ then $M_{2k}(N) \leq (\frac{2}{3})^c (2k-1)!! + O_c \left(\frac{(2k-1)!!}{k} \right)$. There are $(2k-1)!!$ matchings, and each matching contributes at most N^{k+1} (or 1 after division by N^{k+1}; we use both phrasings). We have shown (see Remark 2.3) that we may take as independent variables the k values of the subscripts of the $b_{|j|}$s $(x_1, \ldots, x_k)$ and any index (say i_d). The goal is to show that almost all of the matchings, for k large, have at least c Diophantine obstructions of the type encountered in the fourth moment. If there were no obstructions, these terms would contribute N^3; the obstructions reduce the contribution to $\frac{2}{3} N^3$.

We strategically replace our set of independent variables $i_d, x_1, \ldots, x_k$ with new variables which exhibit the obstructions. We give full details for $c = 1$ (one obstruction), and sketch how to add more. For notational convenience, instead of referring to $i_1, i_2, \ldots, i_{2k}$, we use $i, j, k, \ldots$ and $p, q, r, \ldots$. Thus, in the eigenvalue trace expansion we have terms like $a_{i_1 i_2} = b_{|i_1 - i_2|}$; we refer to this term by $i_1 i_2$ or by ij.
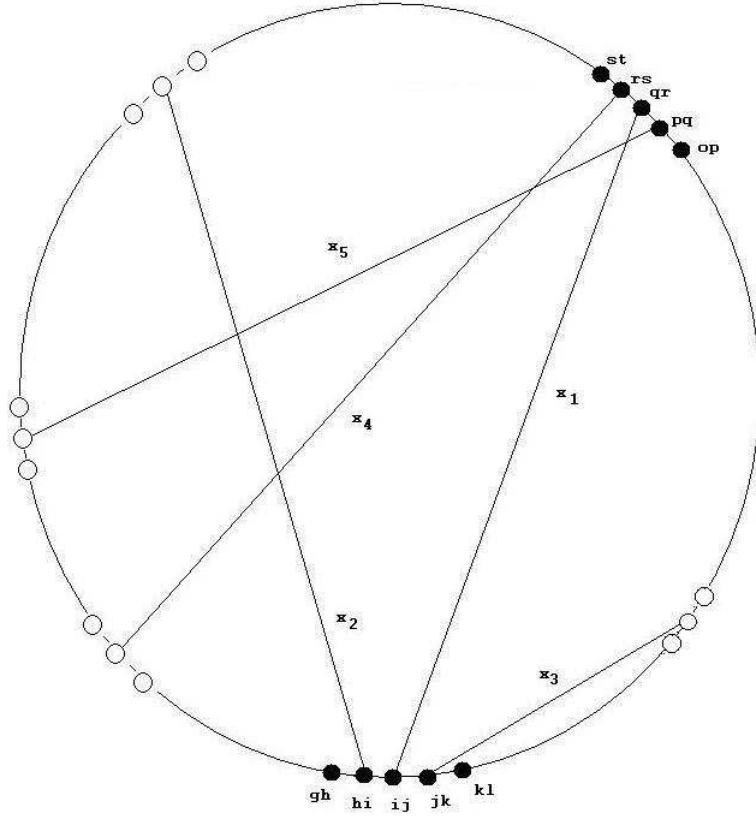
Say we pair $b_{|i-j|}$ with $b_{|q-r|}$. Let $x_1 = i - j = -(q - r)$. If we knew $i = j + r - q$ with j, r and q independent free variables, then our earlier results show there are only $\frac{2}{3} N^3 + O(N^2)$, not $N^3 + O(N^2)$, solutions. Unfortunately, j, r and q need not be independent; however, for almost all of the $(2k-1)!!$ matchings, they will be. We give a good bound on the number of matchings where j, r and q may be taken as independent when $c = 1$; a similar argument works for general c.

Create a buffer zone around ij and qr of two vertices on each side, and assume that neither buffer zone intersects. Given ij, there are $(2k-1) - 8 = 2k - O_c(1)$ possible choices to place qr. Now connect the neighbors of ij and qr such that nothing is connected within one vertex of another. There will be $(2k - O_c(1)) \cdot (2k - O_c(1)) \cdot (2k - O_c(1)) \cdot (2k - O_c(1))$ such matchings. Note that, as we start placing some of these connections, some vertices become unavailable. For example, say there is exactly one vertex between the buffer of ij and the buffer of qr. This vertex is *not* available for use, for if we were to place another vertex there, the indices it gives would not be independent. The same would be true if there were just two vertices between the two buffers, and so on. In each case, however, we only lose $O_c(1)$ vertices. As all these matchings are separated, we may label their differences by x_2, x_3, x_4 and x_5, independent free variables.

The reason for the buffer zones is that the separation allows us to replace some of the independent variables x_ℓ with j, r and q. Note that each index appears in exactly two vertices on the circle, and they are adjacent. Thus, these are the only occurrences of i, j, q and r, and we may replace x_5 as a free parameter with q as a free parameter, x_4 with r, and x_1 with j. We now have the desired situation: $i = j + r - q$, with all three on the right independent free parameters.

There are $(2k-11)!!$ ways to pair the remaining vertices. For those pairs that have j, q, r independent, the contribution is at most $\frac{2}{3}N^3 \cdot N^{k+1-3}$; for the others, we bound the contribution by N^{k+1}. Hence

$$
\begin{aligned}
M_{2k}(N) &\leq \frac{1}{N^{k+1}}\left[(2k)^5 (2k-11)!! \frac{2}{3} N^{k+1} + O_c(k^4) \cdot (2k-11)!! \cdot N^{k+1}\right] \\
&\leq \frac{2}{3}(2k-1)!! + O_c\left(\frac{(2k-1)!!}{k}\right).
\end{aligned}
\tag{23}
$$

Therefore,

$$
\frac{M_{2k}(N)}{(2k-1)!!} \leq \frac{2}{3} + O_c\left(\frac{1}{k}\right).
\tag{24}
$$

There are two ways to handle the general case with c Diophantine obstructions. One may start with enormous buffer zones around the initial pairs. As the construction progresses, we open up more and more portions of the parts of the buffer zones not immediately near the vertices. This keeps all but $O_c(1)$ vertices available for use. Alternatively, along the lines of the first construction, we can just note that by the end of stage c, $O_c(1)$ vertices were unusable. The correction term is smaller than the main term by a factor of $\frac{1}{k}$. $\square$

4. LOWER BOUND OF HIGH MOMENTS

4.1. **Preliminaries.** We know the moments of the limiting spectral measure are bounded by those of the Gaussian, $(2k-1)!!$; the limiting value of the $2k^{\text{th}}$ root of the Gaussian (by Stirling's Formula) is $\frac{k}{e}$. By obtaining a sufficiently large lower bound for the even moments, we show the limiting spectral measure has unbounded support. If it had bounded support, say $[-B, B]$, then the $2k^{\text{th}}$ moment M_{2k} is at most B^{2k}, and $\lim_{k\to\infty} \sqrt[2k]{M_{2k}} < \infty$. We prove

Theorem 4.1. *For any $\epsilon > 0$, for k sufficiently large*

$$
\sqrt[2k]{M_{2k}} \geq k^{\frac{1}{2}-\epsilon}.
\tag{25}
$$

Thus the support of the limiting spectral measure is unbounded.

The construction is as follows: in studying the $2k^{\text{th}}$ moment, we are led to sums of the form

$$
\frac{1}{N^{k+1}}\mathbb{E}\left[\sum_{i_1=1}^{N}\cdots\sum_{i_{2k}=1}^{N} a_{i_1,i_2} a_{i_2,i_3} \cdots a_{i_{2k},i_1}\right] = \frac{1}{N^{k+1}}\mathbb{E}\left[\sum_{i_1=1}^{N}\cdots\sum_{i_{2k}=1}^{N} b_{|i_1-i_2|} b_{|i_2-i_3|} \cdots b_{|i_{2k}-i_1|}\right].
\tag{26}
$$

If any $b_{|i_n - i_{n+1}|}$ occurs only once, as it is drawn from a mean zero distribution, there is no contribution to the expected value. Thus the $2k$ numbers (the bs) are matched in at least pairs, and to obtain a lower bound it suffices to consider the case where the differences are matched in k pairs. Let these positive differences (of $|i_n - i_{n+1}|$) be $x_1, \ldots, x_k$.

In Theorem 2.2 we showed the matchings must occur with negative signs. Thus, if $|i_n - i_{n+1}| = |i_y - i_{y+1}|$, then $(i_n - i_{n+1}) = -(i_y - i_{y+1})$. We let $\widetilde{x}_j = i_j - i_{j+1}$. Thus, for any x_j, there is a unique j_1 such that $\widetilde{x}_{j_1} = x_j$ and a unique j_2 such that $\widetilde{x}_{j_2} = -x_j$. We call the first set of differences positive, and the other set negative; we often denote these by $\widetilde{x}_p$ and $\widetilde{x}_n$, and note that we have k of each.

We have $k + 1$ degrees of freedom. We may take these as the k differences x_k, and then any index, say i_1. We have the relations (see (12))

$$
\begin{aligned}
i_2 &= i_1 - \widetilde{x}_1 \\
i_3 &= i_1 - \widetilde{x}_1 - \widetilde{x}_2 \\
&\vdots \\
i_{2k} &= i_1 - \widetilde{x}_1 - \cdots - \widetilde{x}_{2k}.
\end{aligned}
\tag{27}
$$

Once we specify i_1 and the differences $\widetilde{x}_1$ through $\widetilde{x}_{2k}$, all the indices are determined. If everything is matched in pairs and each $i_j \in \{1, \ldots, N\}$, then we have a valid configuration, which will contribute $+1$ to the $2k^{\text{th}}$ moment. The reason it contributes $+1$ is because, as everything is matched in pairs, we have the expected value of the second moment of $p(x)$, k times.

We need to show the number of valid configurations is sufficiently large. The problem is that, in (27), we need each index $i_j \in \{1, \ldots, N\}$; however, it is possible that a running sum $i_1 - \widetilde{x}_1 - \cdots - \widetilde{x}_m$ is not in this range for some m. Using the Central Limit Theorem, we show that we can keep all these running sums in the desired range sufficiently often.

4.2. **Construction.** Let $\alpha \in (\frac{1}{2}, 1)$; we need $\alpha > \frac{1}{2}$ in order to apply the Central Limit Theorem later. Let $I_A = \{1, \ldots, A\}$, where $A = \frac{N}{k^\alpha}$. Choose each difference x_j from I_A; there are A^k ways to do this. In the end, we want to study k-tuples such that no value is chosen twice. Note such tuples are lower order, namely there are at most $\binom{k}{2} A^{k-1}$ such tuples. This is $O(N^{k-1})$. As i_1 takes on at most N values (not all values will in general lead to valid configurations), we see tuples with repeated values occur at most $O(N^k)$ times; as we divide by N^{k+1}, these terms will not contribute for fixed k as $N \to \infty$. Thus, with probability one (as $N \to \infty$), we may assume the k values x_j are distinct.

Let us consider k distinct positive numbers (the x_js) drawn from I_A, giving rise to k positive differences $\widetilde{x}_p$s and k negative differences $\widetilde{x}_n$s. Let us make half of the numbers $\widetilde{x}_1, \ldots, \widetilde{x}_k$ positive (arising from the $\widetilde{x}_p$s), and half of these numbers negative (arising from the $\widetilde{x}_n$s). Call this the first block (of differences).

Then, in the differences $\widetilde{x}_{k+1}, \ldots, \widetilde{x}_{2k}$ (the second block), we have the remaining differences. Note every positive (negative) difference in $\widetilde{x}_1, \ldots, \widetilde{x}_k$ is paired with a negative (positive) difference in $\widetilde{x}_{k+1}, \ldots, \widetilde{x}_{2k}$. Note we have not specified the order of the differences, just how many positive (negative) are in the first block / second block.

Note two different k-tuples of differences x_j *cannot* give rise to the same configuration (if we assume the differences are distinct). This trivially follows from the fact that the differences specify which diagonal of the Toeplitz matrix the $a_{i_m i_{m+1}}$s are on; if we have different tuples, there is at least one diagonal with an entry on one but not on the other.

Let us assume we have chosen the order of the differences in the first block, $\widetilde{x}_1, \ldots, \widetilde{x}_k$. We look at a subset of possible ways to match these with differences in the second block. In the second block, there are $\frac{k}{2}$ positive (negative) differences $\widetilde{x}_p$ ($\widetilde{x}_n$). There are $(\frac{k}{2})!$ ways to choose the relative order of the positive (negative) differences. Note we are *not* giving a complete ordering of the differences in the second block. There are $k! > (\frac{k}{2})!^2$ ways to completely order. We are merely specifying the relative order among the positive (negative) elements, and not specifying how the positive and negative differences are interspersed.

Thus the number of matchings, each of which contribute 1, obtainable by this method is at most

$$
N \cdot (A^k - O(A^{k-1})) \cdot (k/2)!^2,
\tag{28}
$$

where N is from the possible values for i_1, $A^k - O(A^{k-1})$ is the number of k-tuples of distinct differences $x_j \in I_A$, and $(k/2)!^2$ is the number of relative arrangements of the positive and negative differences in the second block (each of which is matched with an opposite difference in the first block).

Not all of the above will yield a $+1$ contribution to the $2k^{\text{th}}$ moment. Remember, each index i_m must be in $\{1, \ldots, N\}$. We now show that for a large number of the above configurations, we do have all indices appropriately restricted. We call such a configuration *valid*.

4.3. **Number of Valid Configurations.** Most of the time, the sum of the positive differences $\widetilde{x}_p$ in the first block will be close to the negative of the sum of the negative differences $\widetilde{x}_n$ in the first block.

Explicitly, we may regard the $\widetilde{x}_p$s ($\widetilde{x}_n$s) as independent random variables taken from the uniform distribution on I_A ($-I_A$) with mean approximately $\frac{1}{2}A$ ($-\frac{1}{2}A$) and standard deviation approximately $\frac{1}{2\sqrt{3}}A$. By the Central Limit Theorem, for k large the sum of the $\frac{k}{2}$ positive (negative) $\widetilde{x}_p$s ($\widetilde{x}_n$s) in the first block converges to a normal distribution with mean approximately $\frac{kA}{4}$ ($-\frac{kA}{4}$) and standard deviation approximately $\sqrt{\frac{k}{2}} \cdot \frac{A}{2\sqrt{3}}$.

For N and k sufficiently large, the probability that the sum of the positive differences in the first block is in $[\frac{kA}{4} - \frac{\sqrt{k}A}{2\sqrt{6}}, \frac{kA}{4} + \frac{\sqrt{k}A}{2\sqrt{6}}]$ is at least $\frac{1}{2}$ (and a similar statement for the negatives). By the Central Limit Theorem, at least $\frac{1}{4}A^k$ of the A^k tuples will have the sum of the positive (negative) differences lying in this interval (in the negative of this interval). We call such choices *good*.

Remember, in the arguments leading up to (28), we only specified two items. First, the *absolute values* of the k differences (all distinct); second, that half the positive differences are in the first block, and the *relative* orderings of the positive (negative) differences in the second block is given.

We therefore have freedom to choose how to intersperse the positives and negatives in the first and second blocks. Consider a good choice of x_ks. We place these differences in the first block of length k as follows. Choose the first positive difference from our good list, and make the first difference positive. Keep assigning (in order) the positive differences from our good list until the running sum of the differences assigned to the first block exceeds A. Then assign the negative differences from our good list until the running sum of differences in the first block is less than $-A$. We then assign positive differences again until the running sum exceeds A, and so on. If we run out of differences of one sign, we then assign the remaining differences. Note we assigned half of the positive (negative) differences to the first block.

Throughout the process, the largest the running sum can be in absolute value is $\max(2A, 2 \cdot \frac{\sqrt{k}A}{2\sqrt{6}})$. This is because the $\frac{k}{2}$ positive (negative) differences yield sums whose negatives are very close to each other, and each added difference can change the running sum by at most $\pm A$.

We now assign the differences in the second block. We have already chosen the positive and negative differences. There are $(\frac{k}{2})!$ orderings of the positive (negative) differences. Fix a choice for the relative ordering. We intersperse these in a similar manner as in the first block. We put down the differences, again making sure the running sum never exceeds in absolute value $\max(2A, 2 \cdot \frac{\sqrt{k}A}{2\sqrt{6}})$.

Let $i_1 = 0$. From (27), we now see that each index is at most $2\max(2A, 2 \cdot \frac{\sqrt{k}A}{2\sqrt{6}})$. Therefore, each index is in $\left[-\frac{2}{\sqrt{6}}\frac{N}{k^{\alpha - \frac{1}{2}}}, \frac{2}{\sqrt{6}}\frac{N}{k^{\alpha - \frac{1}{2}}}\right]$. If we shift i_1 so that $i_1 \in \left[\frac{7}{8}\frac{N}{k^{\alpha - \frac{1}{2}}}, \frac{N}{k^{\alpha - \frac{1}{2}}}\right]$, as $\alpha > \frac{1}{2}$ for k large all indices will now be in $\{1, \ldots, N\}$. Thus, this is a valid assignment of indices. We now count the number of valid assignments. We see this is at least

$$\left(\frac{1}{8}\frac{N}{k^{\alpha - \frac{1}{2}}}\right) \cdot \left(\frac{1}{4}A^k - \binom{k}{2}A^{k-1}\right) \cdot (k/2)!^2. \tag{29}$$

To calculate the contribution to the $2k^{\text{th}}$ moment from this matching, we divide by N^{k+1}. If any of the differences are the same, there is a slight complication; however, as N is large relative

to k, we may remove the small number of cases (at most $\binom{k}{2}A^k$) when we have repeat differences among the $\widetilde{x}_p$s and $\widetilde{x}_n$s. We divide by N^{k+1}, and by Stirling's Formula the main term is

$$\frac{1}{N^{k+1}}\frac{1}{32}\frac{N^{k+1}}{k^{(k+1)\alpha-\frac{1}{2}}}\left(e^{\frac{k}{2}\log\frac{k}{2}-\frac{k}{2}}\sqrt{2\pi(k/2)}\right)^2 \;=\; \frac{\pi k^{\frac{3}{2}-\alpha}}{32e^{(1+\log 2)k}}\cdot e^{(1-\alpha)k\log k}. \tag{30}$$

The $2k^{\text{th}}$ root is asymptotic to $\frac{e^{(1-\alpha)\log k}}{e^{1+\log 2}} > O(k^{1-\alpha})$, proving the support is unbounded.

5. Weak Convergence

Definition 5.1 (Weak Convergence). *A family of probability distributions μ_n weakly converges to μ if and only if for any bounded, continuous function f we have*

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}f(x)\mu_n(dx) \;=\; \int_{-\infty}^{\infty}f(x)\mu(dx). \tag{31}$$

By Theorem 2.6, we know the moments M_k exist and are finite. To prove we have weak convergence to the limiting spectral measure we need to show that the variances tend to 0. We must show

$$\lim_{N\to\infty}\left(\mathbb{E}[M_m(A,N)^2]-\mathbb{E}[M_m(A,N)]^2\right) \;=\; 0. \tag{32}$$

By (6) we have

$$\mathbb{E}[M_m(A,N)^2] \;=\; \frac{1}{N^{m+2}}\sum_{1\le i_1,\ldots,i_m\le N}\sum_{1\le j_1,\ldots,j_m\le N}\mathbb{E}[b_{|i_1-i_2|}\cdots b_{|i_m-i_1|}b_{|j_1-j_2|}\cdots b_{|j_m-j_1|}]$$

$$\mathbb{E}[M_m(A,N)]^2 \;=\; \frac{1}{N^{m+2}}\sum_{1\le i_1,\ldots,i_m\le N}\mathbb{E}[b_{|i_1-i_2|}\cdots b_{|i_m-i_1|}]\sum_{1\le j_1,\ldots,j_m\le N}\mathbb{E}[b_{|j_1-j_2|}\cdots b_{|j_m-j_1|}]. \tag{33}$$

There are two possibilities: if the absolute values of the differences from the is are completely disjoint from those of the js, then these contribute equally to $\mathbb{E}[M_m(A,N)^2]$ and $\mathbb{E}[M_m(A,N)]^2$. We are left with estimating the difference for the crossover cases, when the value of an $i_\alpha-i_{\alpha+1}=\pm(j_\beta-j_{\beta+1})$. We assume $m=2k$; a similar proof works for odd m. Note $N^{m+2}=N^{2k+2}$, so there are $2k+2$ degrees of freedom. The following two lemmas imply the variance tends to 0.

Lemma 5.2. *The contribution from crossovers in $\mathbb{E}[M_{2k}(A,N)]^2$ is $O_k(\frac{1}{N})$.*

Proof. For $\mathbb{E}[M_{2k}(A,N)]$, the expected value vanishes if anything is unpaired. Thus, in $\mathbb{E}[M_{2k}(A,N)]^2$, in the is and js everything is at least paired, and there is at least one common value from a crossover. The maximum number of such possibilities occurs when everything is paired on each side, and just one set of pairs crosses over; for this crossover there are 2 ways to choose sign. In this case, there are $k+1$ degrees of freedom in the is, and $k+1-1$ degrees of freedom in the js (we lost one degree of freedom from the crossover). Thus, these terms give $O_k(N^{2k+1})$. Considering now matchings on each side with triple or higher pairings, more crossovers, and the two possible assignments of sign to the crossovers, we find that is and js with a crossover contribute $O_k(\frac{1}{N})$ to $\mathbb{E}[M_m(A,N)]^2$. $\square$

Lemma 5.3. *The contribution from crossovers in $\mathbb{E}[M_m(A,N)^2]$ is $O_k(\frac{1}{N})$.*

Proof. If neither the i differences nor the j differences have anything unpaired (i.e., everything is either paired or higher), and there is at least one crossover, it is easy to see these terms are $O_k(\frac{1}{N})$. The difficulty occurs when we have unmatched singletons on either side. Assume there are unmatched differences among the is. We only increase the number of degrees of freedom by replacing triple pairings and higher among the is with pairs and singletons (note we may lose these degrees of freedom as these must be crossed and matched with the js, but we can always cross these over to the js with no net loss of degrees of freedom). Similarly, we can remove triple and higher pairings among the js.

Assume there are $s_i > 0$ singletons and $k - \frac{s_i}{2}$ pairs on the i side, $s_j \geq 0$ singletons on the j side, and $C \geq \max(s_i, s_j)$ crossings. Note s_j equals 0 if we send the singletons on the i side to matched pairs among the js, but C cannot be less than s_i and s_j. Note s_i, s_j are even.

On the i side, there are $1 + (k - \frac{s_i}{2}) + (s_i - 1)$ degrees of freedom; the 1 is from the freedom of assigning any value to one index, then we have $k - \frac{s_i}{2}$ from pairs, and then the last singleton's value is determined, so we have just $s_i - 1$ additional degrees of freedom from singletons.

Assume $s_j > 0$. On the j side, there could have been $1 + (k - \frac{s_j}{2}) + (s_j - 1)$ degrees of freedom, but we know we have C crossings. This loses at least $C - 1$ degrees of freedom (it's possible the last, forced j difference already equalled an i difference). Thus the number of degrees of freedom is

$$\left[1 + \left(k - \frac{s_i}{2} \right) + (s_i - 1) \right] + \left[1 + \left(k - \frac{s_j}{2} \right) + (s_j - 1) - (C - 1) \right] \; = \; 2k + 1 - \frac{1}{2}(2C - s_i - s_j). \tag{34}$$

This is at most $2k + 1$, which is less than $2k + 2$. Therefore there has been a loss of at least one degree of freedom, and these terms contribute $O_k\left(\frac{1}{N}\right)$.

If $s_j = 0$, then there are $1 + k - C$ degrees of freedom on the j side, and we get $2k + 1 - (C - \frac{s_i}{2})$ degrees of freedom, again losing at least one degree of freedom.

Thus there are at most $2k + 1$ degrees of freedom. Doing the combinatorics for choices of sign and number of triples and higher shows these terms also contribute $O_k(\frac{1}{N})$. $\square$

Theorem 5.4. *Let p have mean zero, variance one and finite higher moments. The measures $\mu_{A,N}(x)$ weakly converge to a universal measure of unbounded support, independent of p.*

Proof. By Theorem 2.6 the moments M_k exist and are finite. As $\mathbb{E}[M_k(A, N)] \to M_k$ and the variances tend to zero, standard arguments give weak convergence. As M_k is less than the Gaussian's moments, the M_ks uniquely determine a probability measure, which by §4 has unbounded support. $\square$

6. Almost Sure Convergence

For convenience in presentation, we assume $p(x)$ is even (i.e., the odd moments vanish); we remark later on the modifications to handle the additional book-keeping for general $p(x)$. In Theorem 6.3 we show for all m that

$$\lim_{N \to \infty} \mathbb{E}\left[|M_m(A, N) - \mathbb{E}[M_m(A, N)]|^4 \right] \; = \; O_m \left(\frac{1}{N^2} \right). \tag{35}$$

We first show how (35) (plus Chebychev and Borel-Cantelli) yields almost sure convergence, and then prove Theorem 6.3.

6.1. **Almost Sure Convergence.**

Theorem 6.1. *Let p have mean zero, variance one and finite higher moments. If p is even, as $N \to \infty$ we have almost sure convergence to the limiting spectral distribution determined by the M_ms.*

We first introduce some notation. Fix $p(x)$ as before. Let Ω_N be the outcome space $(T_N, \prod_{i=1}^{N-1} p(b_i) db_i)$, where T_N is the space of all $N \times N$ real symmetric Toeplitz matrices. Let Ω be the outcome space $(T_{\mathbb{N}}, \prod p)$, where $T_{\mathbb{N}}$ is the set of all $\mathbb{N} \times \mathbb{N}$ real symmetric Toeplitz matrices and $\prod p$ is the product measure built from having the entries i.i.d.r.v. from $p(x)$. For each N we have projection maps from Ω to Ω_N. Thus, if $A \in T_{\mathbb{N}}$ is a real symmetric Toeplitz matrices, then A_N is the restriction obtained by looking at the upper left $N \times N$ block of A.

We slightly adjust some notation from before. Let $\mu_{A_N}(x)dx$ be the probability measure associated to the Toeplitz $N \times N$ matrix A_N. Then

$$
\begin{aligned}
\mu_{A_N}(x)dx &= \frac{1}{N}\sum_{i=1}^{N}\delta\left(x - \frac{\lambda_i(A_N)}{\sqrt{N}}\right) \\
M_m(A_N) &= \int_{\mathbb{R}} x^m \mu_{A_N}(x)dx \\
M_m(N) &= \mathbb{E}[M_m(A_N)] \\
M_m &= \lim_{N\to\infty} M_m(N).
\end{aligned}
\tag{36}
$$

As $N \to \infty$, we have shown $M_m(N)$ converges to M_m, and the convergence for each m is at the rate of $\frac{1}{N}$. The expectation above is with respect to the product measure on T_N built from $p(x)$.

We want to show that for all m, as $N \to \infty$,

$$
M_m(A_N) \longrightarrow M_m \quad \text{almost surely.}
\tag{37}
$$

By the triangle inequality,

$$
|M_m(A_N) - M_m| \leq |M_m(A_N) - M_m(N)| + |M_m(N) - M_m|.
\tag{38}
$$

As the second term tends to zero, it suffices to show the first tends to zero for almost all A.

Chebychev's Inequality states that for any random variable X with mean zero and finite ℓ^{th} moment that

$$
\text{Prob}(|X| \geq \epsilon) \leq \frac{\mathbb{E}[|X|^{\ell}]}{\epsilon^{\ell}}.
\tag{39}
$$

Note $\mathbb{E}[M_m(A_N) - M_m(N)] = 0$, and in Theorem 6.3 we prove the fourth moment of $M_m(A_N) - M_m(N)$ is $O_m\left(\frac{1}{N^2}\right)$. Then Chebychev's Inequality (with $\ell = 4$) and Theorem 6.3 yield

$$
\text{Prob}(|M_m(A_N) - M_m(N)| \geq \epsilon) \leq \frac{\mathbb{E}[|M_m(A_N) - M_m(N)|^4]}{\epsilon^4} \leq \frac{C_m}{N^2\epsilon^4}.
\tag{40}
$$

The proof of almost sure convergence (Theorem 6.1) is completed by applying the following:

Lemma 6.2 (Borel-Cantelli). *Let B_i be a sequence of events with $\sum_i Prob(B_i) < \infty$. Let*

$$
B = \left\{\omega : \omega \in \bigcap_{j=1}^{\infty}\bigcup_{k=j}^{\infty} B_i\right\}.
\tag{41}
$$

Then the probability of B is zero.

In other words, an ω is in B if and only if that ω is in infinitely many B_i, and the probability of events ω which occur infinitely often is zero.

Fix a large k and let

$$
B_N^{(k,m)} = \{A \in T_{\mathbb{N}} : |M_m(A_N) - M_m(N)| \geq \frac{1}{k}.\}
\tag{42}
$$

We have seen that $\text{Prob}(B_N^{(k,m)}) \leq \frac{C_m k^4}{N^2}$. Thus, for fixed m and k, the conditions of the Borel-Cantelli Lemma are met, and we deduce that the probability of $A \in T_{\mathbb{N}}$ that occur in infinitely many $B_N^{(k,m)}$ is zero. Letting $k \to \infty$ we find that for any fixed m, as $N \to \infty$, $M_m(A_N) \to M_m$ with probability one. Let $B_m^{i.o.}$ be the probability zero sets where we do not have such convergence.

Let $B^{i.o.} = \bigcup_{m=1}^{\infty} B_m^{i.o.}$. As a countable union of probability zero sets has probability zero, we see that $\text{Prob}(B^{i.o.}) = 0$; however, this is precisely the set where for some m we do not have pointwise convergence. Thus, except for a set of probability zero, we find $M_m(A_N) \to M_m$ for all m. Subject to proving Theorem 6.3, this completes the proof of Theorem 6.1. $\square$

6.2. Proof of Fourth Moment Bounds.

Theorem 6.3. *Let p be an even distribution with mean zero, variance one and finite higher moments. Then*

$$\lim_{N\to\infty} \mathbb{E}\left[|M_m(A,N) - \mathbb{E}[M_m(A,N)]|^4\right] = O_m\left(\frac{1}{N^2}\right). \tag{43}$$

We prove Theorem 6.3 for p even; more involved counting arguments would prove the claim for arbitrary p. We must show

$$\lim_{N\to\infty} \mathbb{E}\left[|M_m(A,N) - \mathbb{E}[M_m(A,N)]|^4\right] = O_m\left(\frac{1}{N^2}\right). \tag{44}$$

We analyze $\mathbb{E}\left[|M_m(A,N) - \mathbb{E}[M_m(A,N)]|^4\right]$, which leads to the proof. Expanding this out, it suffices to study

$$\mathbb{E}[M_m(A,N)^4] - 4\mathbb{E}[M_m(A,N)^3]\mathbb{E}[M_m(A,N)] + 6\mathbb{E}[M_m(A,N)^2]\mathbb{E}[M_m(A,N)]^2 - 3\mathbb{E}[M_m(A,N)]\mathbb{E}[M_m(A,N)]^3. \tag{45}$$

For even moments, we may write the pieces as

$$\mathbb{E}[M_{2m}(A,N)^4] = \frac{1}{N^{4m+4}}\sum_i\sum_j\sum_k\sum_l \mathbb{E}[b_{is}b_{js}b_{ks}b_{ls}]$$

$$\mathbb{E}[M_{2m}(A,N)^3]\mathbb{E}[M_{2m}(A,N)] = \frac{1}{N^{4m+4}}\sum_i\sum_j\sum_k\sum_l \mathbb{E}[b_{is}b_{js}b_{ks}]\mathbb{E}[b_{ls}], \tag{46}$$

(note we combined the $\binom{4}{3}$ and $\binom{4}{4}$ terms) and so on, where for instance

$$E_1 = \mathbb{E}[b_{is}b_{js}b_{ks}b_{ls}] = \mathbb{E}\left[b_{|i_1-i_2|}\cdots b_{|j_{2m}-j_1|}b_{|k_1-k_2|}\cdots b_{|k_{2m}-k_1|}b_{|l_1-l_2|}\cdots b_{|l_{2m}-l_1|}\right]. \tag{47}$$

We analyze the even moments first in Theorems 6.4. The odd moments are handled analogously in Theorem 6.18.

We fix some notation. Denote the expected value sums above by E_1, E_2, E_3 and E_4 (which occur with factors of $1, -4, 6$ and -3 respectively). For $h \in \{i, j, k, l\}$, let b_h refer to the differences in $b_{|h_1-h_2|}\cdots b_{|h_{2m}-h_1|}$. If a difference in a b_h is matched with another difference in b_h, we say this is an *internal* matching; otherwise, it is an *external* matching. By a singleton, pair, triple, quadruple and so on, we refer to matchings within a b_h (i.e., an internal matching). For example, a triple occurs when exactly three of the differences in a b_h are equal.

Let p_a denote the ath moment of $p(x)$. Note $p_1 = 0$ and $p_2 = 1$. For example, in $\sum \mathbb{E}[b_ib_jb_kb_l]$, if we have all differences occurring twice except for two different differences occurring four times (two quadruples) and another different one occurring six times (one sextuple), we would have $1^{2m-7}p_4^2p_6$.

Note there are at most $4m + 4$ degrees of freedom – everything must be matched in at least pairs (we have $8m$ total differences, as we are looking at the fourth power of the $2m^{\text{th}}$ moment), and then each b_h has at most one more degree of freedom (we can choose any index). Thus any terms with a loss of at least two degrees of freedom contribute at most $O_m(\frac{1}{N^2})$.

6.2.1. *Even Moments.*

Theorem 6.4. *Let p be an even distribution with mean zero, variance one and finite higher moments. Then for m even,*

$$\lim_{N\to\infty} \mathbb{E}\left[|M_m(A,N) - \mathbb{E}[M_m(A,N)]|^4\right] = O_m\left(\frac{1}{N^2}\right). \tag{48}$$

We first show in Theorem 6.13 that there is negligible contribution if p is even and there is an internal triple or higher; thus it suffices to consider the case where there are no internal triples or higher. We then show in Theorem 6.15 that if there are no singletons the contribution is negligible, and then we complete the proof by showing in Theorem 6.16 that there is a negligible contribution from singletons.

The following three lemmas are the cornerstone of the later combinatorics:

Lemma 6.5. *If there is a singleton in b_h paired with something in b_g, then there is a loss of at least one degree of freedom.*

Note if every difference in a b_h (all singletons) is paired with a difference in b_g (all singletons), we have a loss of exactly one degree of freedom, so the lemma is sharp. We can choose any index and $2m - 1$ differences in b_h; the last difference is now determined. Once we choose one index in b_g, all other indices are determined, for a total of $1 + (2m - 1) + 1$ (instead of $2m + 2$) degrees of freedom. Thus, instead of being able to choose $2m$ differences freely, we can only choose $2m - 1$. Note the above argument holds if, instead of all singletons, we have elements of b_g and b_h only matched internally and externally with each other.

Proof. As we can cycle the labels, we may assume that $b_{|h_{2m} - h_1|}$ is the singleton. Note that once any index and the values of the other differences in b_h are given, then $|h_{2m} - h_1|$ is determined. We need to conclude we have lost a degree of freedom. It suffices to consider the case where every difference is be paired with another difference; this is because p has mean zero, and any unpaired differences thus contribute zero. Thus, to have the maximum number of degrees of freedom, the difference in each pairing must be free.

We know $b_{|h_{2m} - h_1|}$ must equal the difference from another b_g ($h \neq g \in \{i, j, k, l\}$), say $b_{|g_a - g_{a+1}|}$. We must show the difference $|g_a - g_{a+1}| = |h_{2m} - h_1|$ is not a free parameter. This is obvious because if it were free, that would contradict $|h_{2m} - h_1|$ being determined by the other b_h differences. $\square$

Remark 6.6. *In the above, we did not need the matching to be with a singleton – a pair, triple or higher would also have worked.*

Lemma 6.7. *If at least three of the b_hs have a singleton, there is a loss of at least two degrees of freedom.*

Proof. If there is a matching of singletons from say b_i and b_j, and another matching from b_k and b_l, the lemma is clear from above. Without loss of generality, the remaining case is when a singleton from b_i is matched with one from b_j, and another singleton from b_i is matched with one from b_l. We then apply the previous lemma to (b_j, b_i) and (b_k, b_i). $\square$

Lemma 6.8. *For even moments, if there are no crossovers, there is no net contribution.*

Proof. If there are no crossovers, the expected value of the products are the products of the expected values. Thus, each term becomes $\mathbb{E}[M_{2m}(A, N)]^4$, and $1 - 4 + 6 - 3 = 0$. $\square$

Lemma 6.9. *If $p(x)$ is even and there are at least two internal triples among all of the b_hs, the contribution is $O_m\left(\frac{1}{N^2}\right)$.*

Proof. Everything must be matched in at least pairs (or its expected value vanishes). If there are only two values among six differences, then instead of getting 3 degrees of freedom, we get 2. This is enough to see decay like $O_m\left(\frac{1}{N}\right)$. If we didn't assume $p(x)$ were even, we would have more work to do; as the odd moments vanish, however, the two triples must be paired with other differences, or with each other. In either case, we lose at least one degree of freedom from each, completing the proof. $\square$

Remark 6.10. *Similarly, one can show there cannot be a triple and anything higher than a triple. Further, we cannot have two quadruples or more, as a quadruple or more loses one degree of freedom (a quadruple is two pairs that are equal – instead of having two degrees of freedom, we now have one).*

Lemma 6.11. *If there is an internal quadruple, quintuple, or higher matching within a b_h, the contribution is $O_m(\frac{1}{N^2})$.*

Proof. There can be no sextuple or higher, as this gives at least three pairs matched, yielding one degree of freedom (instead of three). If there is a quadruple or quintuple, everything else must be pairs or singletons or we have lost two degrees of freedom. As the odd moments vanish,

a quintuple must be matched with at least a singleton, again giving six points matched, but only one degree of freedom.

We are left with one quadruple (which gives a loss of one degree of freedom) and all else pairs and singletons. No pairs can be matched to the quadruple or each other, as we would then lose at least two degrees of freedom. If there are any singletons, by Lemma 6.5 there is a loss of a degree of freedom. If we have a quintuple or higher, this is enough to lose two degrees of freedom. Thus, we need only study the case of all pairs and one quadruple, with no external matchings.

As everything is independent, we find a contribution of

$$1 \cdot p_4 - 4 \cdot p_4 + 6 \cdot p_4 - 3p_4 \;=\; 0, \tag{49}$$

where p_4 is the fourth moment of p. $\square$

Lemma 6.12. *If there is only one internal triple (say in b_h) and p is even, then the contribution is $O_m(\frac{1}{N^2})$.*

Proof. As odd moments vanish, the triple must be paired with a singleton from another b_h; further, there must be at least one singleton in the same b_h as the triple (as there are an even number of terms). We thus lose a degree of freedom from the triple matched with a singleton (four points, but one instead of two free differences), and we lose a degree of freedom from the singleton in the same b_h as the triple (Lemma 6.5). Thus we have lost two degrees of freedom. $\square$

We have proved

Theorem 6.13. *The contribution from having an internal triple or higher matching is $O_m(\frac{1}{N^2})$ if p is even.*

Remark 6.14. *Similar arguments work for general $p(x)$, but become more involved.*

We have shown there is no net contribution if there are triples or higher pairings. We now consider the case of singletons and pairs.

Theorem 6.15. *Assume there are no singletons or triples or higher pairings and p is even. Then the contribution is $O_m(\frac{1}{N^2})$.*

Proof. If there are no matchings between b_hs, then everything is independent and we get $1 - 4 + 6 - 3 = 0$. If two pairs are matched, we lose one degree of freedom. There are $\binom{4}{2} = 6$ ways to choose two out of i, j, k, l to be paired.

For the four expected value sums, we get the following contributions: $\binom{4}{2}p_4$ from E_1; $\binom{3}{2}p_4 + (6 - \binom{3}{2})$ from E_2 (three times the two pairs are in the expected value of a product together, giving p_4; the other three times they are separated, giving $p_2 = 1$); $\binom{2}{2}p_4 + (6 - \binom{2}{2})$ from E_3 (only once are the matched pairs together); $\binom{4}{2}$ from E_4. Combining yields

$$1 \cdot 6p_4 - 4(3p_4 + 3) + 6(p_4 + 5) - 3(6) \;=\; 0. \tag{50}$$

If at least three pairs are matched together, or two sets of two pairs are matched together, we lose at least 2 degrees of freedom, giving a contribution of size $O_m(\frac{1}{N^2})$. $\square$

We can now prove

Theorem 6.16. *The contribution when there are no triple or higher internal pairings is at most $O_m(\frac{1}{N^2})$ if p is even.*

Proof. It is sufficient to show the non-zero contributions all lost at least two degrees of freedom. We have already handled the case when there are no singletons. If three or four b_hs have a singleton, we are done by Lemma 6.7. If exactly two have singletons, then there is no contribution in the E_1 through E_4, except for the cases when they are under the expected value together (remember the mean of p vanishes).

We have already lost a degree of freedom in this case; if any pair in any b_h is matched with a pair in a b_g, we lose another degree of freedom. We may therefore assume that there are no matches with four or more elements. Thus every difference that occurs, occurs exactly twice.

There are $\binom{4}{2} = 6$ ways to choose which two of the four b_hs have singletons paired. The contribution from E_1 is 6, from E_2 is 3 (3 of the 6 times they are under the expected value together; the other 3 times they are separated, and the expected value of a difference occurring once is 0), from E_3 is 1 (only 1 of the 6 ways have them under the expected value together), and from E_4 is 0. We thus have a contribution of

$$1 \cdot 6 - 4 \cdot 3 + 6 \cdot 1 - 3 \cdot 0 \ = \ 0. \tag{51}$$

We are left with the case when the only singletons are in one b_h. As we are assuming there are no triple or higher internal matchings, these singletons must then be matched with pairs, giving external triples; as the odd moments of $p(x)$ vanish, there is no net contribution. $\qquad \square$

Remark 6.17. *If we do not assume the odd moments of p vanish, additional book-keeping yields the contribution is of size $\frac{1}{N^2}$. We give an example of the type of argument one needs. If exactly two of the b_hs have singletons, then each has at least two; we've already handled the case when they are matched together. As no difference can be left unmatched, we just need to study the case when we get four triples or two triples and a pair; each clearly loses two degrees of freedom;*

We are left with the case when only one b_h has singletons. We are down one degree of freedom already, so there cannot be another non-forced matching. If there are at least four singletons, we are done. If there are two singletons, we get two triples (either with the same or different b_gs). Similar arguments as before yield the contributions are

$$1 \cdot 6p_3^2 - 4 \cdot 3p_3^2 + 6 \cdot p_3^2 - 3 \cdot 0 \ = \ 0 \tag{52}$$

if the two external triples involve matchings from b_h to the same b_g, and

$$1 \cdot 4p_3^2 - 4 \cdot 3p_3^2 + 6 \cdot 0 - 3 \cdot 0 \ = \ 0. \tag{53}$$

6.2.2. Odd Moments.

Theorem 6.18. *If p is even and has mean zero, variance one and finite higher moments, then*

$$\lim_{N \to \infty} \mathbb{E}\left[|M_{2m+1}(A, N) - \mathbb{E}[M_{2m+1}(A, N)]|^4\right] \ = \ O_m\left(\frac{1}{N^2}\right) \tag{54}$$

Proof. Define E_1, E_2, E_3 and E_4 analogously as in (45). In each b_h, there is at least one odd internal matching (or singleton); thus for p even, only E_1 can be non-zero. If there are four (or more) internal triples (or higher), we lose at least two degrees of freedom.

If there are exactly three internal triples, either two are matched together and one is matched with a singleton, or all three are matched with singletons; in both cases we lose at least two degrees of freedom.

If there are exactly two internal triples, there must be at least two b_hs with singletons. If the triples are matched with singletons, we lose two degrees of freedom; if the triples are matched together we lose one degree from that, and one more degree from the singletons (Lemma 6.5).

If there is exactly one triple, at least three b_hs have singletons, and similar arguments yield a loss of at least two degrees.

If there are no triples, then there are four singletons and by Lemma 6.7 there is a loss of at least two degrees. $\qquad \square$

7. Future Work

As there are only $N - 1$ degrees of freedom for the Toeplitz Ensemble, and not $O(N^2)$, it is reasonable to believe the spacings between adjacent normalized eigenvalues $\left(\frac{\lambda_{i+1}(A)}{\sqrt{N}} - \frac{\lambda_i(A)}{\sqrt{N}}\right)$ may differ from those of full real symmetric matrices. For example, band matrices of width 1 are just diagonal matrices, and there the spacing is Poissonian (e^{-x}); full real symmetric matrices are conjectured to have their normalized spacing given by the GOE distribution (which is well approximated by Axe^{-Bx^2}); however, not all ensembles with $O(N)$ degrees of freedom exhibit Poissonian behavior. For example, there are $\frac{dN}{2}$ degrees of freedom for d-regular graphs, but it

has been numerically observed (see [JMRR] among others) that the spacings between adjacent eigenvalues look GOE.

We chose 1000 Toeplitz matrices (1000×1000), with entries i.i.d.r.v. from the standard normal. We looked at the spacings between the middle 11 normalized eigenvalues for each matrix, giving us 10 spacings. A plot of the spacings between normalized eigenvalues looks Poissonian.
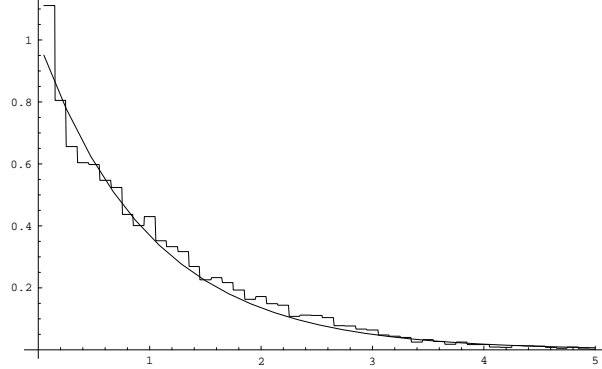
We conjecture that in the limit as $N \to \infty$, the local spacings between adjacent normalized eigenvalues will be Poissonian. While random d-regular graphs have a comparable number of degrees of freedom, there is significantly more independence in the a_{ij} in their adjacency matrices – for the Toeplitz ensemble, we have a strict structure, namely a_{ij} depends only on $|i - j|$.

Additional topics for investigation are to obtain sharper bounds for the growth of the moments M_{2k}. Does

$$\lim_{k \to \infty} \frac{\sqrt[2k]{M_{2k}}}{\sqrt[2k]{G_{2k}}} = 0, \tag{55}$$

where G_{2k} is the $2k^{\text{th}}$ moment of the Gaussian? By Theorem 3.2 all we know is that

$$\lim_{k \to \infty} \frac{M_{2k}}{G_{2k}} = 0, \tag{56}$$

and our method of proof does not yield sharp enough bounds to investigate the $2k^{\text{th}}$ roots.

Finally, one may investigate Toeplitz matrices with additional symmetry. Consider the ensemble of real symmetric palindromic Toeplitz matrices, where in addition to $b_{j-i} = b_{i-j}$ we have $b_{j-i} = b_{N-1-(j-i)}$ (note the first row is a palindrome). Arguing similarly as in Theorem 2.2, we find the only matchings that contribute are those where all signs are negative. The extra symmetry beautifully fixes the Diophantine obstructions. What happens is we have many systems of equations which can be pieced together into one system. For example, (19) becomes

$$i_1 = i_2 + i_4 - i_3 + A_N, \quad A_N \in \{-(N-1), \, 0, \, N-1\}, \tag{57}$$

and for each triple (i_2, i_3, i_4) there is a choice of A_N such that $i_1 \in \{1, \dots, N\}$. To date the first nine moments have been shown to agree with the Gaussian moments. For more details see [MS].

Acknowledgements

References

[Bai]   Z. Bai, *Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review*, Statistica Sinica **9**, 1999, 611-677.

[BCG]   A. Bose, S. Chatterjee and S. Gangopadhyay, *Limiting spectral distributions of large dimensional random matrices*, 2003.

[BM]   A. Bose and J. Mitra, *Limiting spectral distribution of a special circulant*, Statist. Probab. Lett. **60** (2002), no. 1, 111-120.

[BDJ]   W. Bryc, A. Dembo, T. Jiang, *Spectral Measure of Large Randm Hankel, Markov and Toeplitz Matrices*, preprint.

[JMRR]   Jakobson, D., Miller, S.D., Rivin, I., and Rudnick, Z., *Eigenvalue spacings for regular graphs*, Emerging applications of number theory (Minneapolis, MN, 1996), 317–327, IMA Vol. Math. Appl., 109, Springer, New York, 1999.

[Lo]   M. Loéve, *Probability Theory*, D. Van Nostrand Company, Inc., Princeton, 1963.

[McK]   B. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra Appl. **40** (1981), 203-216.

[Meh]   M. Mehta, *Random Matrices, 2nd edition*, Academic Press Inc., Boston, 1991.

[MS]   S. J. Miller and J. Sinsheimer, *Distribution of Eigenvalues for the Ensemble of Real Symmetric Palindromic Toeplitz Matrices*, preprint.

[Wig]   E. Wigner, *On the Distribution of the Roots of Certain Symmetric Matrices*, Ann. of Math., **67**, 1957, 325-327.

E-mail address: `hammond.162@osu.edu`

Department of Mathematics, The Ohio State University, Columbus, OH 43210
Current address: Department of Mathematics, University of Michigan, Ann Arbor, MI 48109

E-mail address: `sjmiller@math.ohio-state.edu`

Department of Mathematics, The Ohio State University, Columbus, Ohio, 43210
Current address: Department of Mathematics, Brown University, Providence, RI 02912