Random Toeplitz Matrices

Arnab Sen
University of Minnesota

Conference on Limits Theorems in Probability, IISc
January 11, 2013

Joint work with Bálint Virág
What are Toeplitz matrices?

\[
\begin{bmatrix}
  a_0 & a_1 & a_2 & \cdots & \cdots & a_{n-2} & a_{n-1} \\
  a_{-1} & a_0 & a_1 & a_2 & \cdots & \cdots & a_{n-2} \\
  a_{-2} & a_{-1} & a_0 & a_1 & \cdots & \cdots & a_{n-3} \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  \cdots & \cdots & \cdots & a_{-1} & a_0 & a_1 & a_2 \\
  a_{-(n-2)} & \cdots & \cdots & a_{-2} & a_{-1} & a_0 & a_1 \\
  a_{-(n-1)} & a_{-(n-2)} & \cdots & \cdots & a_{-2} & a_{-1} & a_0 \\
\end{bmatrix}
= ((a_{j-i}))_{n \times n}.
\]

Symmetric Toeplitz matrix: \( a_{-k} = a_k \) for all \( k \).

Named after Otto Toeplitz (1881 - 1940).
Deterministic Toeplitz operators

- Toeplitz operator = infinite Toeplitz matrix + \( \sum_{i=-\infty}^{\infty} |a_i|^2 < \infty \).
- It has a vast literature.

Toeplitz forms are ubiquitous. For example, covariance matrix of a stationary time-series or a transition matrix of a random walk on \( \mathbb{Z} \) with absorbing barriers.
â : $S^1 \to \mathbb{C}$ such that $\hat{a}(t) = \sum_{n=-\infty}^{\infty} a_n t^n$. Under certain hypotheses on $\hat{a}$,
$$\det((a_{j-i}))_{n \times n} \sim A \cdot \theta^n,$$
where
$$A = \exp \left( \sum_{k=1}^{\infty} k (\log \hat{a})^{-k} (\log \hat{a})_k \right) \text{ and } \theta = \exp \left( (\log \hat{a})_0 \right).$$
This is known as strong Szegő limit theorem.

The magnetization of Ising model on $n \times n$ Torus can be represented as a Toeplitz determinant: first rigorous proof of Onsagar’s formula and phase transition of Ising model.

Many generating functions in combinatorics can be expressed as Toeplitz determinants. For example, the length of the longest increasing subsequence of a random permutation (Baik, Deift, and Johansson, 1999).
Model

\[ T_n = ((a_{|i-j|}))_{n \times n} \]

where \( \{a_i\} \) is an i.i.d. sequence of random variables with \( \mathbb{E}[a_i] = 0, \mathbb{E}[a_i^2] = 1 \).

- Introduced by Bai (1999).

- Compare to Wigner matrix (matrix with i.i.d. entries modulo symmetry), it has additional structures and much less independence.

- Random Toeplitz matrices have connections to one dimensional random Schrödinger operators.
Eigenvalue distribution of random Toeplitz matrices

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(n^{-1/2}T_n)}.$$  Bai asked:  $\mu_n \to \mu_\infty$?

Scaling by $\sqrt{n}$ is necessary to ensure

$$\mathbb{E}\left[\int x^2 \mu_n(dx)\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\lambda_i^2(n^{-1/2}T_n)] = n^{-2}\mathbb{E}[\text{tr}(T_n^2)] = 1.$$

$\mu_\infty$ is not Gaussian distribution!  $\int x^4 \mu_\infty(dx) = 8/3 < 3.$
Existence of $\mu_\infty$

**Theorem (Bryc, Dembo, Jiang (Ann Probab, 2006))**

$\mu_\infty$ exists. $\mu_\infty$ does not depend on the distribution of $a_0$. $\mu_\infty$ is nonrandom, symmetric and has unbounded support.

- The proof is based on method of moments.

$$\int x^k \mathbb{E} \mu_n(dx) = \mathbb{E} \left[ n^{-1} \text{tr}(n^{-1/2} T_n)^k \right].$$

They show that $\int x^k \mathbb{E} \mu_n(dx) \to \gamma_k$ and $\mu_n - \mathbb{E} \mu_n \to 0$. The proof is combinatorial.

- $W_n = n \times n$ Wigner matrix. $(w_{ij})_{i\leq j}$ i.i.d. with mean 0 and variance 1. Then $\mu_\infty$ exists and has density $\frac{1}{2\pi} \sqrt{4 - x^2} 1_{[-2,2]}$. This is famous semicircular law.
\[ \gamma_{2k+1} = 0. \]
\[ \gamma_{2k} = \text{sum of} \frac{(2k)!}{2^kk!} \text{ of } (k + 1)\text{-dimensional integrals. But no closed form expression for } \gamma_{2k} \text{ and hence for } \mu_\infty. \]

\[ \gamma_{2k} \leq \frac{(2k)!}{2^kk!} \Rightarrow \text{subgaussian tail of } \mu_\infty. \]

There is no alternative method known to prove convergence of \( \mu_n \) other than the method of moments.

As of now, the toolbox to deal with random Toeplitz matrix is pretty limited.
The problem of studying the maximum eigenvalue of random Toeplitz matrices is raised in Bryc, Dembo, Jiang (2006).

Meckes (2007): If the entries have uniformly subgaussian tails, then

$$\mathbb{E}[\lambda_1(T_n)] \asymp \sqrt{n \log n}.$$  

Adamczak (2010): \(\{a_i\}\) i.i.d. with \(\mathbb{E}[a_i^2] = 1\).

$$\frac{\|T_n\|}{\mathbb{E}\|T_n\|} \to 1.$$ 

Bose, Hazra, Saha (2010): \(T_n\) with i.i.d. heavy-tailed entries \(\mathbb{P}(|a_i| > t) \sim t^{-\alpha}L(t)\) as \(t \to \infty, 0 < \alpha < 1\). Then

$$\|T_n\| \asymp n^{1/\alpha}.$$
Convergence of Maximum eigenvalue

Let $W_n = ((w_{ij}))_{n \times n}$ be Wigner matrix. Assume $E[w_{12}^4] < \infty$. Then Bai and Yin (1988) showed that

$$n^{-1/2} \lambda_1(W_n) \to 2.$$

For Toeplitz matrix, $\mu_\infty$ has unbounded support and hence there is no natural guess for the limit of $\frac{\lambda_1(T_n)}{\sqrt{n \log n}}$.

The asymptotics of $\text{tr}(T_n^{k_n}) = \sum_{i=1}^{n} \lambda_i^{k_n}(T_n)$ is not known when $k_n \to \infty$. 

Assumption. \((a_i)_{0 \leq i \leq n-1}\) is a sequence of independent random variables. There exists constants \(\gamma > 2\) and \(C\) finite so that for each variable

\[
\mathbb{E}a_i = 0, \quad \mathbb{E}a_i^2 = 1, \quad \text{and} \quad \mathbb{E}|a_i|^{\gamma} < C.
\]

Theorem (Virag, S.)

\[
\frac{\lambda_1(T_n)}{\sqrt{2n \log n}} \xrightarrow{L^\gamma} \|\operatorname{Sin}\|_{2 \to 4}^2 = 0.8288\ldots \quad \text{as} \ n \to \infty.
\]

\[
\operatorname{Sin}(f)(x) := \int_{\mathbb{R}} \frac{\sin(\pi(x - y))}{\pi(x - y)} f(y) \, dy \quad \text{for} \ f \in L^2(\mathbb{R}),
\]

and its \(2 \to 4\) operator norm is

\[
\|\operatorname{Sin}\|_{2 \to 4} := \sup_{\|f\|_2 \leq 1} \|\operatorname{Sin}(f)\|_4
\]
Open problem: limiting behavior of $\lambda_1(T_n)$

**Guess**

$\lambda_1(T_n)$, suitably normalized, converges to Gumbel (double exponential) distribution.

**Remark.** If $x_1, x_2, \ldots, x_n$ are i.i.d. standard Gaussians, then

$$\frac{\max_i x_i - c_n}{d_n} \to \text{Gumbel}.$$
• Bryc, Dembo, Jiang (2006) conjectured that $\mu_\infty$ (for Toeplitz matrices) has a smooth density w.r.t. Lebesgue measure.

**Theorem (Virag, S.)**

The limiting eigenvalue distribution of random Toeplitz matrices has a bounded density.
Connection between Toeplitz and circulant matrices

\[ C_{10} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 \\ a_9 & a_0 & a_1 & a_2 & a_3 \\ a_8 & a_9 & a_0 & a_1 & a_2 \\ a_7 & a_8 & a_9 & a_0 & a_1 \\ a_6 & a_7 & a_8 & a_9 & a_0 \end{bmatrix} \]

Fact: If \( a_j = a_{2n-j} \), then

\[
\begin{bmatrix} T_n & 0_n \\ 0_n & 0_n \end{bmatrix} = \begin{bmatrix} I_n & 0_n \end{bmatrix} C_{2n}^{\text{sym}} \begin{bmatrix} I_n & 0_n \\ 0_n & 0_n \end{bmatrix}.
\]
Circular matrices are easy to understand

- **Spectral Decomposition:**

\[
(m)^{-1/2}C_m = U_m^* \text{diag}(d_0, d_1, \ldots, d_{m-1})U_m,
\]

\[
U_m(k, l) = \exp\left(\frac{2\pi i kl}{m}\right), \quad d_k = m^{-1/2} \sum_{l=0}^{m-1} a_l \exp\left(\frac{2\pi i kl}{m}\right).
\]

- **U_m =** discrete Fourier transform.

- Change of basis for \(n^{-1/2}\)

\[
\begin{bmatrix}
T_n & 0_n \\
0_n & 0_n
\end{bmatrix}
\]

\[
n^{-1/2}U_{2n} \begin{bmatrix}
T & 0 \\
0 & 0
\end{bmatrix} U_{2n}^* = \sqrt{2}U_{2n} \begin{bmatrix}
I & 0 \\
0 & 0
\end{bmatrix} U_{2n}^* D_{2n} U_{2n} \begin{bmatrix}
I & 0 \\
0 & 0
\end{bmatrix} U_{2n}^* = \sqrt{2}\text{PDP}.
\]
\begin{itemize}
  \item \(D\) is a \textit{random} diagonal matrix whose entries have mean zero, variance \(\sigma^2\) and are \textit{uncorrelated}.
  
  \item Thus for Gaussian Toeplitz matrices, then entries of \(D\) are just i.i.d. Gaussians.
  
  \item \(P_{2n} = U_{2n} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} U_{2n}^*\) is a \textit{deterministic} Hermitian projection matrix.
  
  \item \(P_{2n}(i, j)\) is a function of \(|i - j|\) (and \(n\)).
  
  \item As \(n \to \infty\), \(P_{2n}\) ‘converges’ to \(\Pi : \ell^2 \to \ell^2\).
  
  \begin{align*}
    \Pi : \ell^2(\mathbb{Z}) & \xrightarrow{\text{Fourier Transf.}} L^2(S^1) \xrightarrow{1_{[0,1/2]}} L^2(S^1) \xrightarrow{\text{Inverse F.T.}} \ell^2(\mathbb{Z}).
  \end{align*}
\end{itemize}
Model. $H_\omega = \Delta + V_\omega$ acts on $\ell^2(\mathbb{Z})$ by

$$(H_\omega \phi)(i) = \phi(i - 1) + \phi(i + 1) + v_i(\omega)\phi(i),$$

where $(v_i)_{i \in \mathbb{Z}}$ are i.i.d. random variables.

Morally, $H_\omega = \text{random multiplication operator with a local (additive) perturbation.}$

Toeplitz matrix in Fourier basis $= PD\! P$.

The projection operator $P$ behaves like a “local perturbation”.
How $2 \to 4$ norm arises: Gaussian case

- $\frac{1}{\sqrt{2 \log n}} \lambda_1(P_{2n} D_{2n} P_{2n}) \approx \sup_{\Theta_k} \lambda_1(\Pi_k \Theta_k \Pi_k)$.

- $\Theta_k$ is admissible if

$$\Theta_k = \lim_{n \to \infty} \frac{1}{\sqrt{2 \log n}} (d_{i+1}, d_{i+2}, \ldots, d_{i+k}),$$

for some $i$.

- When is $\Theta_k = \text{diag}(\theta_1, \theta_2, \ldots, \theta_k)$ inadmissible? Ans: $\sum_{i=1}^k \theta_i^2 > 1$.

$$\mathbb{P}(|d_{i+1}| > \theta_1 \sqrt{2 \log n}, \ldots, |d_{i+k}| > \theta_k \sqrt{2 \log n}) \leq n^{-(\theta_1^2 + \ldots + \theta_k^2)}.$$

- For large $k$, $\lambda_1(\Pi_k \Theta_k \Pi_k) \approx \lambda_1(\Pi \Theta \Pi)$.

- We have a double optimization problem,

$$\sup_{\Theta} \lambda_1(\Pi \Theta \Pi) = \sup \left\{ \left\langle v, \Pi \text{diag}(\theta) \Pi v \right\rangle : \|v\|_2 \leq 1, \|\theta\|_2 \leq 1 \right\} = \|\Pi\|_{2\to 4}^2.$$

- Finally, $\frac{\lambda_1(P_{2n} D_{2n} P_{2n})}{\sqrt{2 \log n}} \approx \|\Pi\|_{2\to 4}^2$. 

Arnab Sen
University of Minnesota
Random Toeplitz Matrices
Fact (play with Fourier Transform)

\[ \| \Pi \|_{2 \rightarrow 4}^2 = \frac{1}{\sqrt{2}} \| \text{Sin} \|_{2 \rightarrow 4}^2. \]

Key reason:

\[ \text{F.T. of } (1_{[-1/2,1/2]} \cdot f) = 1_{[-1/2,1/2]} \hat{f} = \frac{\sin(\pi x)}{\pi x} \hat{f} = \text{Sin}(\hat{f}) \]

- This optimization problem has been studied by Garsia, Rodemich and Rumsey (1969).
- They computed \( \| \text{Sin} \|_{2 \rightarrow 4}^4 = 0.686981293033114600949413\ldots \)!
They are many (technical) gaps in the sketch. Non-Gaussian case is harder due to lack of independence.

\[ d_k = n^{-1/2} \sum_{\ell=0}^{n} a_k \cos\left(\frac{2\pi k\ell}{2n}\right). \]

We need normal approximation in the moderate deviation regime,

\[ \mathbb{P}(d_1 > \theta_1 \sqrt{2 \log n}, \ldots, d_k > \theta_k \sqrt{2 \log n}) = \left(1 + o(1)\right) \mathbb{P}(Z_1 > \theta_1 \sqrt{2 \log n}, \ldots, Z_k > \theta_k \sqrt{2 \log n}). \]

Note that CLT only gives

\[ \mathbb{P}(d_1 > \theta_1, \ldots, d_k > \theta_k) = \left(1 + o(1)\right) \mathbb{P}(Z_1 > \theta_1, \ldots, Z_k > \theta_k). \]
Stieltjes transform

Definition

For a measure $\mu$,

$$S(z; \mu) := \int \frac{1}{x - z} \mu(dx), \quad z \in \mathbb{C}, \text{Im}(z) > 0.$$

Key Fact

If

$$\sup_{z: \text{Im}(z) > 0} \text{Im}S(z; \mu) \leq K,$$

then $\mu$ is absolutely continuous w.r.t. the Lebesgue measure and

$$\frac{d\mu}{dx} \leq \frac{K}{\pi}.$$

The proof follows from the inversion formula.

$$\int_x^y \mu(dE) = \lim_{\delta \to 0^+} \frac{1}{\pi} \int_x^y \text{Im}S(E + i\delta; \mu)dE, \quad x < y \in \mathcal{C}(\mu).$$
Enough to show

$$\sup_{z: \text{Im}(z) > 0} S(z, \mathbb{E}_{\mu_n}) \leq C \quad \text{for all} \ n$$

for Gaussian Toeplitz matrices.

$$S(z, \mathbb{E}_{\mu_n}) = n^{-1} \mathbb{E} \text{tr} \left( n^{-1/2} T_n - z I \right)^{-1}$$

$$= \frac{\sqrt{2}}{n} \sum_{j=1}^{2n} \mathbb{E} \langle P e_j, (PDP - z I)^{-1} P e_j \rangle$$

To show that $$\sup_{z: \text{Im}(z) > 0} \mathbb{E} \langle P e_j, (PDP - z I)^{-1} P e_j \rangle \leq C$$ for each $$j$$ uniformly in $$n$$.

Let $$D_{\theta} = \text{diag}(d_1, d_2, \ldots, d_{j-1}, \theta, d_{j+1}, \ldots, d_{2n})$$.

$$\mathbb{E} \left[ \langle P e_j, (PDP - z I)^{-1} P e_j \rangle | d_i, i \neq j \right]$$
Theorem (Combes, Hislop and Mourre, Trans. AMS 1996)

Let $H_\theta, \theta \in \mathbb{R}$ be a family of self-adjoint operators. Assume that there exist a finite positive constant $c_0$, and a positive bounded self-adjoint operator $B$ such that,

I. $\frac{dH_\theta}{d\theta} \geq c_0 B^2$.

II. $\frac{d^2 H_\theta}{d\theta^2} = 0$.

Then for all $g \in C^2(\mathbb{R})$ and for all $\varphi$,

$$
\sup_{\text{Im}(z)>0} \left| \int g(\theta) \langle B\varphi, (H_\theta - z)^{-1} B\varphi \rangle d\theta \right| 
\leq c_0^{-1} (\|g\|_1 + \|g'\|_1 + \|g''\|_1) \|\varphi\|^2.
$$

- Easy to check $\frac{d}{d\theta} PD_\theta P = Pe_j e'_j P \geq 2(P e_j e'_j P)^2$. 

Arnab Sen  University of Minnesota  Random Toeplitz Matrices
Some heuristics about spectral averaging

- Let $\lambda_i$ be an eigenvalue of $PDP$ with eigenvector $u_i$.

- Let $D = \text{diag}(d_1, d_2, \ldots, d_j, \ldots, d_{2n})$.

- **Bad case:** small perturbations of $d_j$'s do not perturb $\lambda_i(D)$.

- **Hadamard first variational formula:**

\[
\frac{\partial}{\partial d_j} \lambda_i = u_i^* \frac{\partial}{\partial d_j} (PDP) u_i = u_i^* Pe_j e'_j Pu_i.
\]

\[
u_i^* Pe_j e'_j Pu_i = \left| e'_j Pu_i \right|^2 = \left| u_i(j) \right|^2 > 0. \text{ Hence,}
\]

\[\| \nabla \lambda_i(D) \|_1 = 1 \quad \forall D.
\]

Bad case won’t happen.
**Conjecture:** With high probability, the eigenvectors of $PDP$ are localized ($\ell^2$ weight of a generic eigenvector is concentrated on $o(n)$ coordinates).

[Graphs showing localized eigenvectors of $PDP$ dominated by a few coordinates and eigenvectors of a Wigner matrix where none of the coordinates dominates others.]
The eigenvalue process of $T_n$, away from the edge, after suitable normalization, converges to a standard Poisson point process on $\mathbb{R}$.

Let $V_n$ be the top eigenvector of $PDP$. Then there exist random integers $K_n$ so that for each $i \in \mathbb{Z}$

$$V_n(K_n + i) \to \hat{g}(i),$$

where $\hat{g}$ is the Fourier transform of the function $g(x) = \sqrt{2}f(2x - 1/2)$ and $f$ is the (unique) optimizer in

$$\sup\{\|f \ast f\|_2 : f(x) = f(-x), \|f\|_2 = 1, \ f \text{ supported on } [-1/2, 1/2]\}.$$