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?

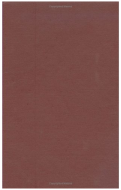
$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & \cdots & a_{n-2} & a_{n-1} \\ a_{-1} & a_0 & a_1 & a_2 & \cdots & \cdots & a_{n-2} \\ a_{-2} & a_{-1} & a_0 & a_1 & \cdots & \cdots & a_{n-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & a_{-1} & a_0 & a_1 &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

Symmetric Toeplitz matrix: $a_{-k} = a_k$ for all k.

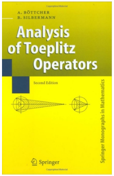
Named after Otto Toeplitz (1881 - 1940).

Deterministic Toeplitz operators

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



Toeplitz Forms and Their Applications by Grenander and Szegö (1958)



Analysis of Toeplitz operators by Böttcher and Silbermann (1990).

- 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.

Usefulness: Toeplitz determinants and Szegö formula

- $\hat{a} : S^1 \to \mathbb{C}$ 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
  $A = \exp\left(\sum_{k=1}^{\infty} k(\log \hat{a})_{-k}(\log \hat{a})_k\right)$ and $\theta = \exp\Big((\log \hat{a})_0\Big)$.
  This is known as strong Szegö 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).

Random (symmetric) Toeplitz matrices

Model

$$\mathbf{T}_n = ((a_{|i-j|}))_{n\times n}$$

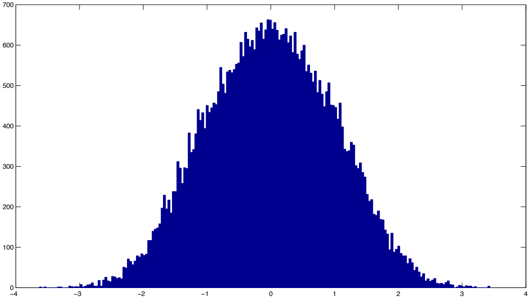
where $\{a_i\}$ is an i.i.d. sequence of random variables with $\mathbb{E}[a_i] = 0, \mathbb{E}[a_i^2] = 1$.

- 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

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

Scaling by $\sqrt{n}$ is necessary to ensure
$\mathbb{E}[\int x^2 \mu_n(dx)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\lambda_i^2(n^{-1/2}\mathbf{T}_n)] = n^{-2}\mathbb{E}[\mathrm{tr}(\mathbf{T}_n^2)] = 1.$



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

Existence of μ_∞

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

μ_∞ exists. μ_∞ does not depend on the distribution of a_0. μ_∞ is nonrandom, symmetric and has unbounded support.

- The proof is based on method of moments.

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

  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.

- $\mathbf{W}_n = n \times n$ Wigner matrix. $(w_{ij})_{i \leq j}$ i.i.d. with mean 0 and variance 1. Then μ_∞ exists and has density $\frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}$. This is famous semicircular law.

What else? Not much

- $\gamma_{2k+1} = 0$.
  γ_{2k} = sum of $\frac{(2k)!}{2^k k!}$ of $(k+1)$-dimensional integrals. But no closed form expression for γ_{2k} and hence for μ_∞.

- $\gamma_{2k} \leq \frac{(2k)!}{2^k k!} \Rightarrow$ subgaussian tail of μ_∞.

- There is no alternative method known to prove convergence of μ_n other than the method of moments.

- As of now, the toolbox to deal with random Toeplitz matrix is pretty limited.

Maximum eigenvalue of random Toeplitz matrices

- 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
$$\mathbb{E}[\lambda_1(\mathbf{T}_n)] \asymp \sqrt{n \log n}.$$

- Adamczak (2010): $\{a_i\}$ i.i.d. with $\mathbb{E}[a_i^2] = 1$.
$$\frac{\|\mathbf{T}_n\|}{\mathbb{E}\|\mathbf{T}_n\|} \to 1.$$

- Bose, Hazra, Saha (2010): $\mathbf{T}_n$ 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
$$\|\mathbf{T}_n\| \asymp n^{1/\alpha}.$$

Convergence of Maximum eigenvalue

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

$$n^{-1/2}\lambda_1(\mathbf{W}_n) \to 2.$$

- For Toeplitz matrix, μ_∞ has unbounded support and hence there is no natural guess for the limit of $\frac{\lambda_1(\mathbf{T}_n)}{\sqrt{n\log n}}$.

- The asymptotics of $\mathrm{tr}(\mathbf{T}_n^{k_n}) = \sum_{i=1}^n \lambda_i^{k_n}(\mathbf{T}_n)$ is not known when $k_n \to \infty$.

First Result: Maximum eigenvalue

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

$$\mathbb{E}a_i = 0, \quad \mathbb{E}a_i^2 = 1, \quad \textit{and} \quad \mathbb{E}|a_i|^\gamma < C.$$

Theorem (Virag, S.)

$$\frac{\lambda_1(\mathbf{T}_n)}{\sqrt{2n\log n}} \xrightarrow{L^\gamma} \|\mathtt{Sin}\|_{2\to 4}^2 = 0.8288\ldots \qquad \textit{as } n\to\infty.$$

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

and its $2\to 4$ operator norm is

$$\|\mathtt{Sin}\|_{2\to 4} := \sup_{\|f\|_2\le 1} \|\mathtt{Sin}(f)\|_4$$

Open problem: limiting behavior of $\lambda_1(\mathbf{T}_n)$

Guess

$\lambda_1(\mathbf{T}_n)$, suitably normalized, converges to Gumbel (double exponential) distribution.

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

$$\frac{\max_i x_i - c_n}{d_n} \to \text{Gumbel}.$$

Second Result: Absolute continuity

- Bryc, Dembo, Jiang (2006) conjectured that μ_∞ (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

$$
\mathbf{C}_{10} = \left[\begin{array}{ccccc|ccccc}
a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 \\
a_9 & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\
a_8 & a_9 & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \\
a_7 & a_8 & a_9 & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\
a_6 & a_7 & a_8 & a_9 & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \\
\hline
a_5 & a_6 & a_7 & a_8 & a_9 & a_0 & a_1 & a_2 & a_3 & a_4 \\
a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_0 & a_1 & a_2 & a_3 \\
a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_0 & a_1 & a_2 \\
a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_0 & a_1 \\
a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_0
\end{array}\right]
$$

- **Fact:** If $a_j = a_{2n-j}$, then

$$
\begin{bmatrix} \mathbf{T}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} \mathbf{C}_{2n}^{\mathrm{sym}} \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}.
$$

Circulants are easy to understand

- Spectral Decomposition:

$$(m)^{-1/2}\mathbf{C}_m = \mathbf{U}_m^* \mathrm{diag}(d_0, d_1, \ldots, d_{m-1})\mathbf{U}_m,$$

$$\mathbf{U}_m(k,l) = \exp\left(\frac{2\pi ikl}{m}\right), \quad d_k = m^{-1/2}\sum_{l=0}^{m-1} a_l \exp\left(\frac{2\pi ikl}{m}\right).$$

- $\mathbf{U}_m$ = discrete Fourier transform.

- Change of basis for $n^{-1/2}\begin{bmatrix} \mathbf{T}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}$

$$n^{-1/2}\mathbf{U}_{2n}\begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}_{2n}^* = \sqrt{2}\mathbf{U}_{2n}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}_{2n}^*\mathbf{D}_{2n}\mathbf{U}_{2n}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}_{2n}^*$$
$$= \sqrt{2}\mathbf{P}\mathbf{D}\mathbf{P}.$$

PDP decomposition

- $\mathbf{D}$ is a random diagonal matrix whose entries have mean zero, variance σ^2 and are uncorrelated.

- Thus for Gaussian Toeplitz matrices, then entries of $\mathbf{D}$ are just i.i.d. Gaussians.

- $\mathbf{P}_{2n} = \mathbf{U}_{2n} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}_{2n}^*$ is a deterministic Hermitian projection matrix.

- $\mathbf{P}_{2n}(i,j)$ is a function of $|i-j|$ (and n).

- As $n \to \infty$, $\mathbf{P}_{2n}$ 'converges' to $\Pi : \ell^2 \to \ell^2$.

$$\Pi : \ell^2(\mathbb{Z}) \xrightarrow{Fourier\ Transf.} L^2(S^1) \xrightarrow{\mathbf{1}_{[0,1/2]}} L^2(S^1) \xrightarrow{Inverse\ F.T.} \ell^2(\mathbb{Z}).$$

Connection to 1-D random Schrödinger operators

- Model. $H_\omega = \Delta + V_\omega$ acts on $\ell^2(\mathbb{Z})$ by

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

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

- Morally, H_ω = random multiplication operator with a local (additive) perturbation.

- Toeplitz matrix in Fourier basis = **PDP**.

  The projection operator **P** behaves like a "local perturbation".

How 2 → 4 norm arises: Gaussian case

- $\frac{1}{\sqrt{2\log n}}\lambda_1(\mathbf{P}_{2n}\mathbf{D}_{2n}\mathbf{P}_{2n}) \approx \sup_{\Theta_k} \lambda_1(\Pi_k\Theta_k\Pi_k)$.
- Θ_k is admissible if

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

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

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

- For large k, $\lambda_1(\Pi_k\Theta_k\Pi_k) \approx \lambda_1(\Pi\Theta\Pi)$.
- We have a double optimization problem,

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

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

Appearance of Sine kernel

Fact (play with Fourier Transform)

$$\|\Pi\|_{2\to4}^2 = \frac{1}{\sqrt{2}}\|\mathtt{Sin}\|_{2\to4}^2.$$

Key reason :

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

- This optimization problem has been studied by Garsia, Rodemich and Rumsey (1969).
- They computed $\|\mathtt{Sin}\|_{2\to4}^4 = 0.686981293033114600949413...!$

A few more words

- They are many (technical) gaps in the sketch.
- Non-Gaussian case is harder due to lack of independence.

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

- We need normal approximation in the moderate deviation regime,

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

- Note that CLT only gives

$$\mathbb{P}(d_1 > \theta_1, \ldots, d_k > \theta_k) = \Big(1 + o(1)\Big) \mathbb{P}(Z_1 > \theta_1, \ldots, Z_k > \theta_k).$$

Stieltjes transform

Definition

For a measure μ,

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

Key Fact

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

then μ 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.

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

- Enough to show

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

for Gaussian Toeplitz matrices.

-

$$S(z,\mathbb{E}\mu_n) = n^{-1}\mathbb{E}\mathrm{tr}(n^{-1/2}\mathbf{T}_n - z\mathbf{I})^{-1}$$
$$= \frac{\sqrt{2}}{n}\sum_{j=1}^{2n}\mathbb{E}\langle \mathbf{P}e_j, (\mathbf{PDP} - z\mathbf{I})^{-1}\mathbf{P}e_j\rangle$$

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

- Let $\mathbf{D}_\theta = \mathrm{diag}(d_1, d_2, \ldots, d_{j-1}, \theta, d_{j+1}, \ldots, d_{2n})$.

- $\mathbb{E}\left[\langle \mathbf{P}e_j, (\mathbf{PDP} - z\mathbf{I})^{-1}\mathbf{P}e_j\rangle | d_i, i \ne j\right]$

Spectral averaging from Random Schrödinger operator

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 φ,

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

- Easy to check $\frac{d}{d\theta} \mathbf{P} \mathbf{D}_\theta \mathbf{P} = \mathbf{P} e_j e_j' \mathbf{P} \geq 2(\mathbf{P} e_j e_j' \mathbf{P})^2$.

Some heuristics about spectral averaging

- Let λ_i be an eigenvalue of $\mathbf{PDP}$ with eigenvector u_i.
- Let $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_j, \ldots, d_{2n})$.
- Bad case: small perturbations of d_j's do not perturb $\lambda_i(\mathbf{D})$.
- Hadamard first variational formula:

$$\frac{\partial}{\partial d_j}\lambda_i = u_i^* \frac{\partial}{\partial d_j}(\mathbf{PDP})u_i = u_i^* \mathbf{P} e_j e_j' \mathbf{P} u_i.$$

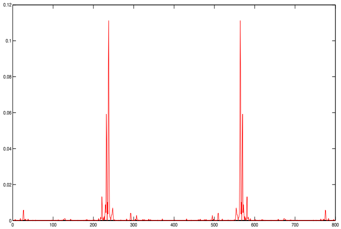
- $u_i^* \mathbf{P} e_j e_j' \mathbf{P} u_i = |e_j' \mathbf{P} u_i|^2 = |u_i(j)|^2 > 0$. Hence,

$$\|\nabla \lambda_i(\mathbf{D})\|_1 = 1 \quad \forall \mathbf{D}.$$

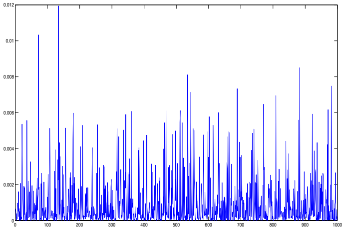
Bad case won't happen.

Question: localized eigenvectors?

Conjecture: With high probability, the eigenvectors of **PDP** are localized (ℓ^2 weight of a generic eigenvector is concentrated on $o(n)$ coordinates).



Eigenvector of **PDP**.
Dominated by a few coordinates.



Eigenvector of Wigner matrix.
None of the coordinates dominates others.

More open problems

- The eigenvalue process of $\mathbf{T}_n$, away from the edge, after suitable normalization, converges to a standard Poisson point process on $\mathbb{R}$.

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

$$\mathbf{V}_n(K_n + i) \to \hat{g}(i),$$

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

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