Smallest singular value and limit eigenvalue distribution of a class of non-Hermitian random matrices with statistical application

ARUP BOSE
Indian Statistical Institute, Kolkata
WALID HACHEM
Université Paris-Est Marne-la-Vallée, France

January 14, 2019

Motivation: Linear time series

General model, MA(q):

$$Y_t = \sum_{j=0}^{q} \psi_j X_{t-j} \ \ t \geq 1. \tag{0.1}$$

X_t's are *i.i.d.* N-dimensional vectors with mean 0 and variance-covariance matrix I_N.

ψ_j are $N \times N$ (non-random) *coefficient matrices*. $\psi_0 = I$. $N = N(n) \to \infty$ such that $\frac{N}{n} \to \gamma \in [0, \infty)$.

q can be infinite but that needs additional restrictions on $\{\psi_i\}$.

MA(0) is the i.i.d. process.

Autocovariance matrix sequence

The sample autocovariance matrix of order i of $\{Y_t\}$ equals

$$\hat{\Gamma}_i := \frac{1}{n} \sum_{t=i+1}^{n} Y_t Y_{t-i}^* .$$

They are all non-symmetric except $\hat{\Gamma}_0$.

Ambition

Derive the asymptotic behaviour of these sequences of sample autocovariance matrices.

• convergence in the algebraic sense (tracial convergence)–essentially known from Bhattacharjee and Bose (BB) for $y \in [0, \infty)$.

• convergence of the spectral distribution of the autocovariances (and their matrix polynomials)–non-trivial. Results for the symmetrised autocovariances are known from BB.

• Applications: determination of the order q, testing for white noise.... (see book by BB where symmetrised autocovariances have been used).

We have only very limited success so far on the convergence of the spectral distribution in non-symmetric cases.

ESD and LSD

R_N: an $N \times N$ (random) matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$.

Empirical Spectral Distribution (ESD) of R_N is the (random) probability measure

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

Its cumulative form (when all eigenvalues are real)

$$ECDF(x) = \frac{1}{n} \# \text{eigenvalues} \leq x.$$

Limiting spectral distribution (LSD): If this ESD converges weakly (for our purposes, almost surely or in probability) to a probability distribution, then the limit is called the LSD. Our limit measures are non-random.

Simulation 1: ECDF of $\hat{\Gamma}_0$ for MA(0)

$$\hat{\Gamma}_0 = \frac{1}{n}\sum_{t=1}^{n} X_t X_t^* .$$

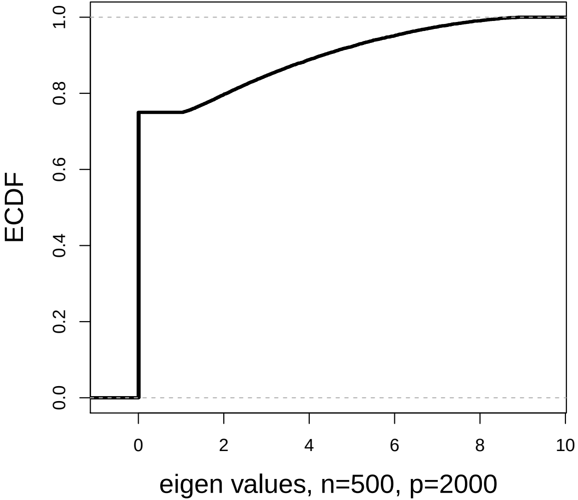


Figure: LSD of $\hat{\Gamma}_0$ is the Marchenko-Pastur law (well known in RMT).

Simulation 2

Model 1 (MA(0)): $Y_t = X_t$.

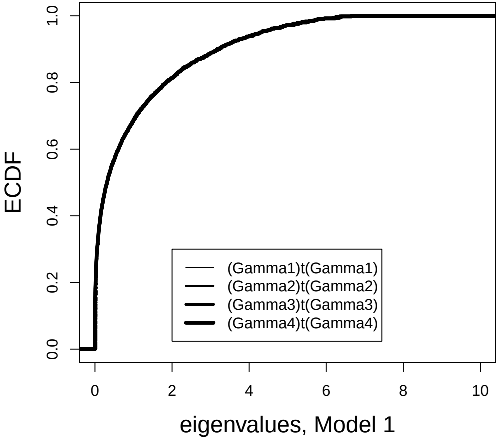


Figure: Identical ECDF of $\hat{\Gamma}_u\hat{\Gamma}_u^*$, $1 \le u \le 4$ for $N = n = 300$. LSD known from BB.

Simulation 3

Model 2 $\quad Y_t = X_t + A_p X_{t-1}, \quad A_N = 0.5 I_N.$

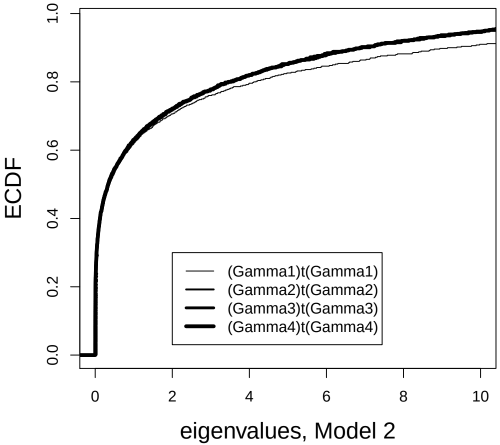


Figure: ECDF of $\hat{\Gamma}_1 \hat{\Gamma}_1^*$ different from ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $2, 3, 4$. $N = n = 300$. Spectral distribution of A_N is degenerate at 0.5. LSD known from BB.

ESD of $\hat{\Gamma}_1$ for MA(0), $n = 500$

$$\hat{\Gamma}_1 = \frac{1}{n}\sum_{t=1}^{n} X_t X_{t-1}^*.$$

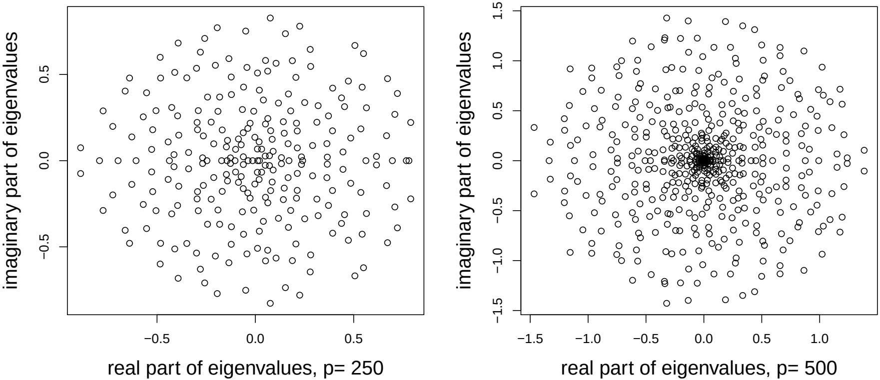


Figure: That this LSD exists is the only (new) LSD result of this talk.

Symmetric polynomials of autocovariances

LSD results for all *symmetric polynomials of autocovariance matrices* have been established by BB. (Minimal assumptions on the ψ matrices and the random matrices have real entries with all moments finite).

These LSD are expressed in terms of *free independent variables* or

in special cases, in terms of functional equations for *Stieltjes transforms*.

Due to symmetry of the matrices, these results can be established by the *method of moments* (showing that the normalised trace of any power converges and the limit sequence determines a unique probability distribution).

Non-symmetric matrices

Tracial moments cannot give the existence of the LSD.

One of the major steps in the proof for any non-symmetric matrix is to establish suitable bounds for the smallest singular value of the matrix.

We shall clarify at the end how this step helps via log-potential.

$(X^{(n)} = [x_{ij}^{(n)}]_{i,j=0}^{N^{(n)}-1,n-1})_{n\geq 1}$ is a sequence of **complex** random matrices such that

Assumption 1. For each $n \geq 1$, $\{x_{ij}^{(n)}\}_{i,j=0}^{N^{(n)}-1,n-1}$ are i.i.d. with

$\mathbb{E}x_{00}^{(n)} = 0$,

$\mathbb{E}|x_{00}^{(n)}|^2 = 1/n$, and

$\sup_n n^2\mathbb{E}|x_{00}^{(n)}|^4 = \boldsymbol{m}_4 < \infty$.

Note the different scaling of the entries.

$(N^{(n)})_{n\geq 1}$, a sequence of positive integers, diverges to ∞ as $n \to \infty$.

$N/n \to \gamma$, $0 < \gamma < \infty$ as $n \to \infty$.

Think of N as the dimension of the vectors and n as the sample size (number of observations).

For any matrix $M \in \mathbb{C}^{n\times n}$, $s_0(M) \geq \cdots \geq s_{n-1}(M)$: the singular values of M.

Assumption 2. $A^{(n)} \in \mathbb{C}^{n\times n}$: sequence of deterministic matrices such that,

$$0 < \inf_n s_{n-1}(A^{(n)}) \leq \sup_n s_0(A^{(n)}) < \infty.$$

The Identity matrix trivially satisfies Assumption 2.

The circulant matrix

$$J^{(n)} = \begin{bmatrix} 0 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{0.2}$$

also satisfies Assumption 2. Its ESD converges to the uniform distribution on the circle of unit radius.

Assumption 3 The random variables $x_{00}^{(n)}$ satisfy $\sup_n |n\mathbb{E}(x_{00}^{(n)})^2| < 1$.

Assumption 3 essentially says that the x_{ij} are *not* real.

This is fine for applications in wireless communications and signal processing.

Not acceptable for time series applications.

This assumption is due to the application of a Berry-Esseen bound as a tool in exactly one of the steps in the proof.

A general smallest singular value result

Let $\|\cdot\|$ denote the spectral norm of a matrix.

Consider the random matrix $X^{(n)}A^{(n)}{X^{(n)}}^* - zI_N$, where z is an arbitrary non-zero complex number and I_N is the identity matrix of dimension N.

Theorem 1 Suppose Assumptions 1, 2 and 3 hold. Let C be a positive constant. Then, there exist $\alpha, \beta > 0$ such that for each $z \in \mathbb{C} \setminus \{0\}$,

$$\mathbb{P}\left[s_{N-1}(X^{(n)}A^{(n)}{X^{(n)}}^* - z) \leq t,\ \|X\| \leq C\right] \leq c\left(n^{\alpha}t^{1/2} + n^{-\beta}\right) + \exp(-c'n),$$

where the constants $c, c' > 0$ depend on C, z, and $\boldsymbol{m}_4$ only. In particular the result is true for XJX^*.

Note that we DO NOT (yet) have the result for matrices with real entries.

We next state our main LSD result. Then we shall explain the connection how Theorem 1 helps in proving Theorem 2.

LSD result for $\hat{\Gamma}_1 = \sum X_t X_{t-1}^*$

To state our result on LSD, let for any $0 < \gamma < \infty$,

$$g(y) = \frac{y}{y+1}(1-\gamma+2y)^2, \ (0 \vee (\gamma-1)) \leq y \leq \gamma. \tag{0.3}$$

Then g^{-1} exists on the interval $[0 \vee ((\gamma-1)^3/\gamma), \gamma(\gamma+1)]$ and maps it to $[0 \vee (\gamma - \gamma^{-1}), \gamma]$. It is an analytic increasing function on the interior of the interval.

Note that g is a cubic polynomial and so a formula can be given for its inverse.

Theorem 2. Suppose Assumptions 1 and 3 hold. Then, the LSD of $\sum X_t X_{t-1}^*$ exists in probability. The limit measure $\boldsymbol{\mu}$ is rotationally invariant on $\mathbb{C}$. Let $F(r) = \boldsymbol{\mu}(\{z \in \mathbb{C} : |z| \leq r\}), 0 \leq r < \infty$ be the distribution function of the radial component.

If $\gamma \leq 1$, then

$$F(r) = \begin{cases} \gamma^{-1} g^{-1}(r^2) & \text{if } 0 \leq r \leq \sqrt{\gamma(\gamma+1)}, \\ 1 & \text{if } r > \sqrt{\gamma(\gamma+1)}. \end{cases}$$

If $\gamma > 1$, then

$$F(r) = \begin{cases} 1 - \gamma^{-1} & \text{if } 0 \leq r \leq (\gamma-1)^{3/2}/\sqrt{\gamma}, \\ \gamma^{-1} g^{-1}(r^2) & \text{if } (\gamma-1)^{3/2}/\sqrt{\gamma} < r \leq \sqrt{\gamma(\gamma+1)}, \\ 1 & \text{if } r > \sqrt{\gamma(\gamma+1)}. \end{cases}$$

The support of $\boldsymbol{\mu}$ is the disc $\{z : |z| \leq \sqrt{\gamma(\gamma+1)}\}$ when $\gamma \leq 1$.

When $\gamma > 1$, the support is the ring $\{z : (\gamma-1)^{3/2}/\sqrt{\gamma} \leq |z| \leq \sqrt{\gamma(\gamma+1)}\}$ together with the point $\{0\}$ where there is a mass $1-\gamma^{-1}$.

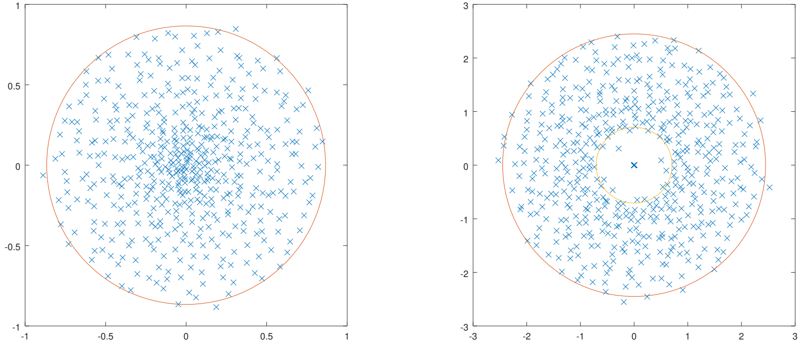
Moreover, $F(r)$ has a positive and analytical density on the open interval $(0 \vee \text{sign}(\gamma-1)|\gamma-1|^{3/2}/\sqrt{\gamma}, \sqrt{\gamma(\gamma+1)})$.

This density is bounded if $\gamma \neq 1$. If $\gamma = 1$, then the density is bounded everywhere except when $r \downarrow 0$.

A cumbersome closed form expression for g^{-1} (and hence for $F(\cdot)$) can be obtained by calculating the root of a third degree polynomial. For the special case $\gamma = 1$, g^{-1} is given by

$$g^{-1}(t) = \frac{t^{1/3}}{2}\left(\left[1+\sqrt{1-\frac{t}{27}}\right]^{1/3} + \left[1-\sqrt{1-\frac{t}{27}}\right]^{1/3}\right), \quad 0 \le t \le 2.$$

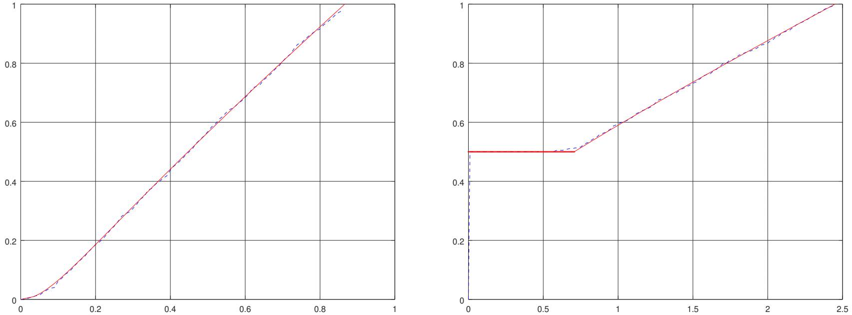
Eigenvalue realizations corresponding to the cases where $\gamma = 0.5$ and $\gamma = 2$ are shown in the next Figure.



: $(N, n) = (500, 1000)$: $(N, n) = (1000, 500)$

Figure: Eigenvalue realizations and LSD support.

Plot of F



: $(N, n) = (500, 1000)$: $(N, n) = (1000, 500)$

Figure: Plots of $F(r)$ (plain curves) and their empirical realizations (dashed curves).

Statistical testing with singular values

LSD results and also normality of trace results are useful for graphical and significance tests of hypothesis.

The book by BB has examples. See also BB (2018). Since they use symmetrization, it amounts to using the singular values. There is loss of information in dealing with singular values rather than eigenvalues.

Statistical applications with eigenvalues

Theorem 2 gives the so-called null distribution of the eigenvalues under white noise (IID) hypothesis. To test other hypothesis, we need LSD results for general MA(q) models. This problem is non-trivial.

log-potential

The *logarithmic potential* of a probability measure μ on $\mathbb{C}$ is the $\mathbb{C} \to (-\infty, \infty]$ superharmonic function defined as

$$U_\mu(z) = -\int_{\mathbb{C}} \log|\lambda - z| \; \mu(d\lambda) \;\; \text{(whenever the integral is finite).}$$

μ can be recovered from $U_\mu(\cdot)$: let $\Delta = \partial_x^2 + \partial_y^2 = 4\partial_z \partial_{\bar{z}}$ for $z = x + \imath y \in \mathbb{C}$ be the Laplace operator defined on

$$C_c^\infty(\mathbb{C}) = \{\varphi : \varphi \text{ is a compactly supported real valued smooth function on } \mathbb{C}\}.$$

Then

$$\mu = -(2\pi)^{-1} \Delta U_\mu \;\; \text{in the sense that} \tag{0.4}$$

$$\int_{\mathbb{C}} \varphi(z) \; \mu(dz) = -\frac{1}{2\pi} \int_{\mathbb{C}} \Delta\varphi(z) \, U_\mu(z) \; dz, \;\; \forall \varphi \in C_c^\infty(\mathbb{C}).$$

Convergence of the logarithmic potentials for Lebesgue almost all $z \in \mathbb{C}$ implies the weak convergence of the underlying measures under a tightness criterion.

Observe that

$\sum X_t X_{t-1}^*$ is essentially the same as the matrix as $XJX^* = Y$ (say).

The logarithmic potential of the spectral measure of Y equals

$$U_{\mu_n}(z) = -\frac{1}{N}\sum \log|\lambda_i - z| = -\frac{1}{N}\log|\det(Y-z)|$$
$$= -\frac{1}{2N}\log\det(Y-z)(Y-z)^* = -\int \log\lambda \, \nu_{n,z}(d\lambda),$$

where the probability measure $\nu_{n,z}$ is the *singular value* distribution of $Y - z$, given as

$$\nu_{n,z} = \frac{1}{N}\sum_{i=0}^{N-1} \delta_{s_i(Y-z)}.$$

Thus we need to study the asymptotic behavior of $U_{\mu_n}(z)$ for Lebesgue almost all $z \in \mathbb{C}$.

Lemma 4.3 of Bordenave and Chafai Let (M_n) be a sequence of random matrices with complex entries. Let ζ_n be its spectral measure and let $\sigma_{n,z}$ be the empirical singular value distribution of $M_n - z$. Assume that

(i) for almost every $z \in \mathbb{C}$, there exists a probability measure $\boldsymbol{\sigma}_z$ such that $\sigma_{n,z} \Rightarrow \boldsymbol{\sigma}_z$ in probability,

(ii) log is uniformly integrable in probability with respect to the sequence $(\sigma_{n,z})$.

Then, there exists a probability measure $\boldsymbol{\zeta}$ such that $\zeta_n \Rightarrow \boldsymbol{\zeta}$ in probability, and furthermore,

$$U_{\boldsymbol{\zeta}}(z) = -\int \log \lambda \; \boldsymbol{\sigma}_z(d\lambda) \quad \mathbb{C} - \text{a.e.}$$

Thus we need to establish that:

Step 1: for almost all $z \in \mathbb{C}$, $\nu_{n,z} \Rightarrow \nu_z$ (a deterministic probability measure) in probability.

Step 2: the function log is uniformly integrable with respect to the measure $\nu_{n,z}$ for almost all $z \in \mathbb{C}$ in probability. That is,

$$\forall \varepsilon > 0, \quad \lim_{T \to \infty} \limsup_{n \geq 1} \mathbb{P}\left[\int_0^{\infty} |\log \lambda| \, 1_{|\log \lambda| \geq T} \; \nu_{n,z} d(\lambda) > \varepsilon\right] = 0. \tag{0.5}$$

Then there exists a probability measure $\boldsymbol{\mu}$ such that $\mu_n \Rightarrow \boldsymbol{\mu}$ in probability, and $U_{\boldsymbol{\mu}}(z) = -\int \log|\lambda| \; \check{\boldsymbol{\nu}}_z(d\lambda)$ $\mathbb{C}$-almost everywhere. It would then remain to identify the measure $\boldsymbol{\mu}$ to complete the proof of Theorem 2.

Steps 1 and 2

Step 1 Proved by covergence of the trace of resolvent. There one first replaces trace by its expectation, then uses Gaussian approximation, PN inequality.

Step 2 requires control of both, the small eigenvalues and the large eigenvalues. The latter is achieved trivially (since spectral norm of X is almost surely bounded). The former is achieved by Theorem 1.

For detailed proof, see Bose and Hachem (2018).

Very selective references

Bhattacharjee and Bose (2016). Large sample behavior of high dimensional autocovariance matrices. Annals of Statistics.

Bhattacharjee and Bose (2018) (Book). Large Covariance and Autocovariance Matrices. Chapman and Hall.

Bhattacharjee and Bose (2018). Joint convergence of sample autocovariance matrices when $p/n \to 0$ with application. To appear in Annals of Statistics.

Bordenave and Chafai (2012). Around the circlar law. Prob Surveys.

Bose and Hachem (2018). Arxiv.

References to other relevant works are available in the above works.

THANK YOU !