High Dimensional Time Series and Free Probability

Arup Bose
Indian Statistical Institute, Kolkata
(with Monika Bhattacharjee)

Informatics Institute, UFL, Gainesville
March 01, 2018

Model: Linear time series

MA(q):

$$X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j} \;\; t \geq 1. \tag{0.1}$$

ε_t (unobservable), $X_t, t = 1, \ldots, n$ (observable) p-dimensional vectors.

ε_t's are *i.i.d.* with mean 0 and variance-covariance matrix I_p.

ψ_j are $p \times p$ (non-random) *coefficient matrices*. $\psi_0 = I$.

$p = p(n) \to \infty$ such that $\frac{p}{n} \to y \in (0, \infty)$ ($y = 0$ is left out of this talk).

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

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

Goal

Primary goal: large sample properties of sample autocovariance matrices (including their polynomial functions and joint convergence).

Technique: large dimensional IID/Wigner matrices and free variables. (Ideas from Random Matrix Theory (RMT) and Free Probability)

Broadened framework: behavior of matrix polynomials of several sample variance-covariance matrices and deterministic matrices.

Applications: determination of the order q, testing for white noise....

Autocovariance matrix sequence

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

$$\hat{\Gamma}_i := \frac{1}{n} \sum_{t=i+1}^{n} X_t X'_{(t-i)}.$$

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

Non-symmetric random matrices are notoriously hard to analyse. We shall resort to (additive and multiplicative) symmetrisation, such as

$$\hat{\Gamma}_i \hat{\Gamma}_i^*, \quad \hat{\Gamma}_i + \hat{\Gamma}_i^*, \quad \hat{\Gamma}_i \hat{\Gamma}_j \hat{\Gamma}_j^* \hat{\Gamma}_i^* \ldots$$

ESD and LSD

R_p: a $p \times p$ (random) matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$.

Empirical Spectral Distribution (ESD) of R_p is the probability measure

$$\mu_p = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}.$$

Its cumulative form (when all eigenvalues are real)

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

Limiting spectral distribution (LSD): If this ESD converges weakly (for our purposes almost surely) to a probability distribution, then the limit is called the LSD.

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

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

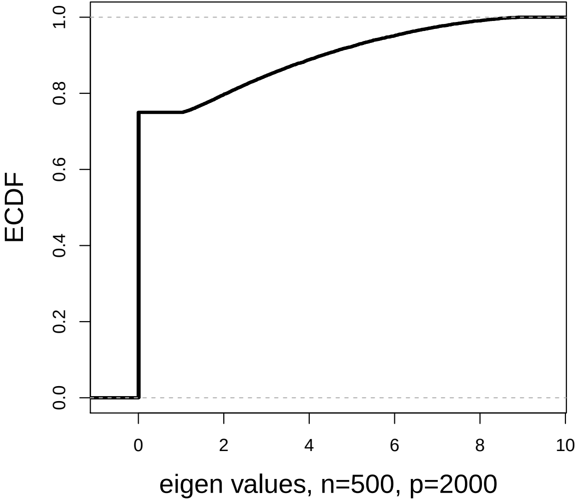

Figure: LSD of $\hat{\Gamma}_0$ is the Marchenko-Pastur law (known from RMT).

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

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

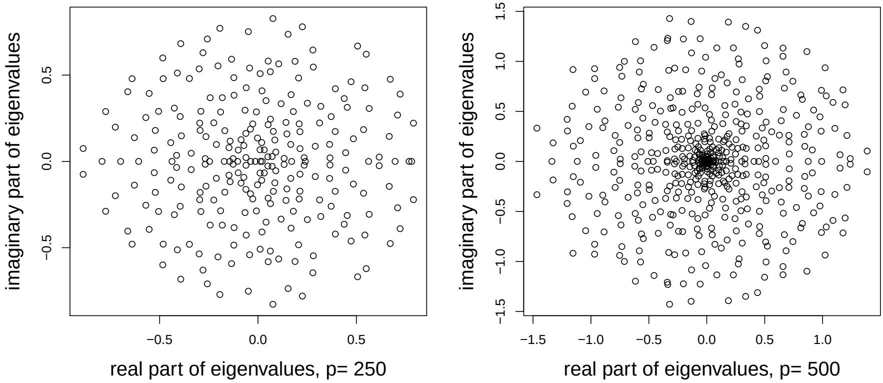

Figure: No theoretical result known.

Model 1 (MA(0)): $X_t = \varepsilon_t$.

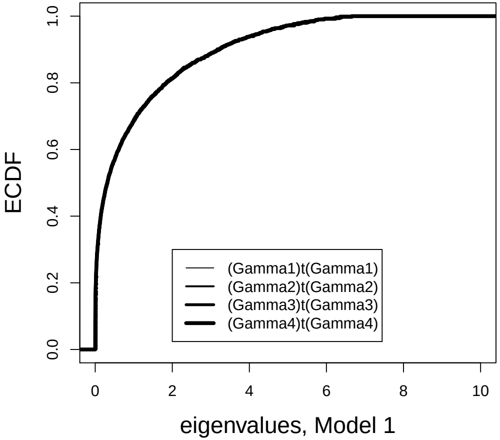

Figure: Identical ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $1 \leq u \leq 4$ for $n = p = 300$.

Model 2 $X_t = \varepsilon_t + A_p \varepsilon_{t-1}, \quad A_p = 0.5 I_p.$

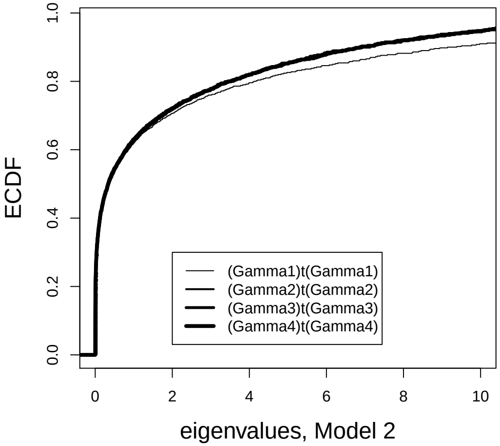

Figure: ECDF of $\hat{\Gamma}_1 \hat{\Gamma}_1^*$ different from ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $2, 3, 4$. $n = p = 300$. Spectral distribution of A_p is degenerate at 0.5.

Model 3 $X_t = \varepsilon_t + B_p \varepsilon_{t-1}$, $B_p = 0.5(I_p + J_p)$.

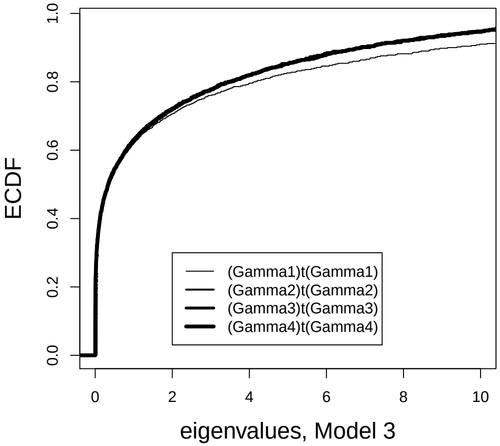

Figure: ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $1 \le u \le 4$ for $n = p = 300$. Spectral distribution of B_p converges to the point mass at 0.5. Graphs are identical with those in Model 2.

Model 4 $X_t = \varepsilon_t + C_p\varepsilon_{t-1} + D_p\varepsilon_{t-2}$,
$C_p = ((c_{i,j}))$, with $c_{i,i} = I(1 \leq i \leq [p/2]) - I([p/2] < i \leq p)$,
$D_p = ((d_{i,j}))$, with $d_{i,p+1-i} = 1$ for all $i \geq 1$ and 0 otherwise.

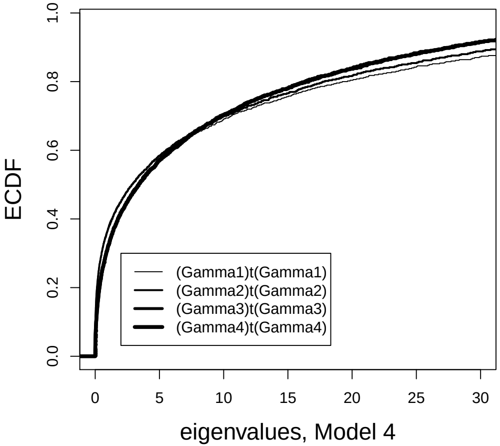

Figure: ECDF of $\hat{\Gamma}_u\hat{\Gamma}_u^*$, $i = 1, 2$ different from $i = 3, 4$. $n = p = 300$. LSD of C_p puts mass 1/2 each at ± 1. LSD of D_p is point mass at 1.

Similar patterns if we use other combinations such as

$$\hat{\Gamma}_u \hat{\Gamma}_u^* + \hat{\Gamma}_{u+1} \hat{\Gamma}_{u+1}^*.$$

Existing results

Symmetrized Autocovariances: $\hat{\Gamma}_0, \quad \hat{\Gamma}_i\hat{\Gamma}_i', \quad \hat{\Gamma}_i + \hat{\Gamma}_i'$ *for all* $i \geq 1$.

Do the LSD exist for the above symmetric matrices?

Answers were known in very specific cases.

Stieltjes transform

Stieltjes transform of any probablity distrbution F on R is

$$m_F(z) = \int \frac{1}{x-z} dF(x), \quad z \in \mathbb{C}^+.$$

It is always defined.

Its moments are defined as (in this talk all moments are finite for all variables we discuss)

$$\beta_h = \int x^h dF(x).$$

Usually LSD are expressed through their Stietljes transform.

Jin, Wang, Bai, Nair and Harding (2014)

Suppose $X_t = \varepsilon_t$ i.i.d. standardized, $E\varepsilon_{1,1}^4 < \infty$.

Then LSD of $\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$, $i > 1$ exists and the Stieltjes transform satisfies

$$(1 - y^2 m^2(z))(yzm(z) + y - 1)^2 = 1 \tag{0.2}$$

Points to note:

Limit does not depend on i.

$p/n \to y \neq 0$.

I.I.D. Model.

Additive symmetrisation.

One autocovariance matrix at a time.

Liu, Aue and Paul (2015), assumptions

$$X_t = \sum_j \psi_j \varepsilon_{t-j}.$$

(a) $\{\varepsilon_{t.i} : t, i = 1, 2, 3, \ldots\}$ i.i.d. standardized, $E|\varepsilon_{1,1}|^{4+\delta} < \infty$.

(b) ψ_j Hermitian, and norm bounded as $p \to \infty$.

(c) (Roughly speaking) $\{\psi_j\}$ are simultaneously diagonalizable....and the joint eigenvalue distribution converges.

Liu, Aue and Paul (2015), result

Then the LSD of $\frac{1}{2}(\hat{\Gamma}_j + \hat{\Gamma}_j^*)$ exists with Stieltjes transform

$$m_j(z) = -\int \frac{dF^{\psi}(\lambda)}{z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(jv)h(\lambda,v)}{1+y\cos(jv)K_j(z,v)}dv} \quad \forall z \in \mathbb{C}^+, \tag{0.3}$$

where,

$$K_j(z,v) = -\int \frac{h(\lambda,v)dF^{\psi}(\lambda)}{z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(jv')h(\lambda,v')}{1+y\cos(jv')K_j(z,v')}dv'}, \tag{0.4}$$

$$h(\lambda,v) = |\sum_{l=0}^{\infty} e^{ilv} f_l(\lambda)|^2. \tag{0.5}$$

Points to note:

Wang and Paul (2015): similar result for the case $p/n \to 0$.

Additive symmetrisation.

One matrix at a time.

Bird's eye view

Drop simultaneously diagonalisable condition.

Other kinds of symmetrisation.

More than one matrix at a time (polynomials).

Combine more than one time series.

How can the limits be described? (Free Probability)

Unified way.

Statistical applications.

Stieltjes transform method

The Stieltjes transform of the ESD of a $p \times p$ real symmetric matrix R_p equals,

$$m_p(z) = \frac{1}{p} \sum \frac{1}{\lambda_i - z} = \frac{1}{p} Tr[(R_p - zI)^{-1}].$$

Express $m_{p+1}(z)$ in terms of $m_p(z)$ and use martingale techniques and derive a limiting functional equation.

Done on a case by case basis.

Moment method

The h-th order moment of the ESD of a $p \times p$ real symmetric matrix R_p equals,

$$\beta_h(R_p) := \frac{1}{p}\operatorname{Tr}(R_p^h) \quad (\textit{Trace-Moment formula}).$$

(M1) (Moment convergence) For every $h \geq 1$,

$$E(\beta_h(R_p)) \to \beta_h$$

, (M4) (Borel-Cantelli, to guarantee almost sure convergence)

$$\sum_{p=1}^{\infty} E(\beta_h(R_p) - E(\beta_h(R_p)))^4 < \infty, \quad \forall h \geq 1$$

, (C) (Unique limit) $\displaystyle\sum_{h=1}^{\infty} \beta_{2h}^{-\frac{1}{2h}} = \infty$ (Carleman's condition).

If (M1), (M4) and (C) hold, then ESD of R_p converges almost surely to the distribution F which is determined uniquely by the moments $\{\beta_h\}$.

Joint convergence

More than one sequence. How does one define convergence?

1. Form polynomials.

2. Establish the (M1) condition for all polynomials. That is convergence of the *average trace*.

3. Identify the *non-commutative* limit variables and their interrelation.

LSD for any specific (symmetric) combination: establish the other conditions (M4) and (C).

Toy example:

$\{X_i\}$ iid F. $\{Y_i\}$ iid G.

$$A_p = Diag(X_1, X_2, \ldots X_p).$$
$$B_p = Diag(Y_1, Y_2, \ldots Y_p).$$

Then ESD of A_p and B_p converge almost surely to F and G respectively.

(M1) condition for all polynomials of A_p, B_p holds if all moments of F and G are finite.

The limit variables, say a, b are *commutative*, *independent* and have distributions F, G.

When we have more general pairs, there is no commutativity (and hence no independence).

Our result for autocovariances

(A1) $\{\varepsilon_{t,j}\}$ are independent with $E(\varepsilon_{t,j}) = 0$ and $E\varepsilon_{t,j}^2 = 1, \ \forall i, j$.

(A2) $\sup_{t,j} E(|\varepsilon_{t,j}|^k) < \infty, \ \forall k \geq 1$.

(A3) $\{\psi_j\}$ are compactly supported and jointly converge: for any polynomial Π in $\{\psi_j, \psi_j^* : j \geq 0\}$, $\lim p^{-1}\mathrm{Tr}(\Pi)$ exists.

Then the LSD exists for **any** symmetric polynomial Π in $\{\Gamma_i, \Gamma_i^*\}$. Can identify the limit as a function of **free variables**.

Similar result with additional scaling and centering when $p/n \to 0$.

Note: For some particular polynomials such as $\hat{\Gamma}_i + \hat{\Gamma}_i^*$ and $\hat{\Gamma}_i\hat{\Gamma}_i^*$, $i \geq 0$, Assumption $(A2)$ can be relaxed.

Our general random matrix result

- $\{Z_u = ((\varepsilon_{u,i,j}))_{p\times n}\},\ 1 \le u \le U$
- $\{\varepsilon_{u,i,j}\}$ are independently distributed with mean 0, variance 1.
- $\{A_i\}$: class of $p \times p$ matrices, $\{B_i\}$: class of $n \times n$ matrices
- $\mathbb{P} = \left(\prod_{i=1}^{k_l} \frac{1}{n} A_{t_i}\ Z_{j_i} B_{s_i} Z_{j_i}^*\right) A_{t_{k_l+1}}$,
- $\mathbb{G} = \left(\prod_{i=1}^{k_l} \frac{1}{n} \mathrm{Tr}\left(B_{s_i}\right)\right) \prod_{i=1}^{k_l+1} A_{t_i}$

Under appropriate assumptions, LSD exists for:

- any symmetric polynomial in $\{\mathbb{P}\}$ when $p/n \to y > 0$
- any symmetric polynomial in $\{\sqrt{np^{-1}}(\mathbb{P} - \mathbb{G})\}$ when $p/n \to 0$.

Moreover,

- $(\mathrm{Span}\{\mathbb{P}\}, p^{-1} E\mathrm{Tr})$ converges when $p/n \to y > 0$
- $(\mathrm{Span}\{\sqrt{np^{-1}}(\mathbb{P} - \mathbb{G})\}, p^{-1} E\mathrm{Tr})$ converges when $p/n \to 0$.

Limits can be expressed in terms of **free variables**.

Remarks

1. The three results, JWBN (2014), LAP (2015), WP (2015) follow as special cases.

2. Results for additive and multiplicative symmetrisation can be proved for $q = \infty$ under additional conditions.

3. The LSD of the additive and multiplicative symmetrisations depend on the order q. They remain same once $i > q$. This can be used for order determination, at least in an exploratory way.

4. Can handle more than one sequence of independent linear processes.

5. For $y = 0$, embedding does not work. Need harder calculations.

Embedding idea

Simple case:

$$X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1},$$

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

ID matrix and "lag" matrices

- Independent (ID) matrix

$$Z = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)_{p \times n}.$$

- Lag matrices: For $i \geq 0$, P_i is the $n \times n$ matrix whose i-th upper diagonal is 1 and 0 otherwise.

$P_0 = I_n$

For $i < 0$, let $P_i = P^*_{-i}$.

First approximation

$$
\begin{aligned}
\hat{\Gamma}_0 &= n^{-1}\sum_{t=1}^{n}(\varepsilon_t + \psi_1\varepsilon_{t-1})(\varepsilon_t + \psi_1\varepsilon_{t-1})^* \\
&= n^{-1}(ZP_0Z^* + \psi_1 ZP_0Z^*\psi_1^* + \psi_1 ZP_1Z^* + ZP_{-1}Z^*\psi_1^*) + R_p \\
&= \Delta_0 + R_p \quad (say).
\end{aligned}
$$

- Fact: LSD of $\hat{\Gamma}_0$ and Δ_0 are identical (R_p is negligible).

Embedding

Need to verify the (M1) condition, $\lim p^{-1} E\mathrm{Tr}(\Delta_0^r) =??$

• Embedding: Wigner matrix, W_k: a symmetric $k \times k$ matrix with iid elements. Nice properties are known. Z is embedded as follows in a Wigner matrix:

$$W_{3,n+p} = \begin{pmatrix} W_{1,p} & Z_{p\times n} \\ Z'_{n\times p} & W_{2,n} \end{pmatrix}.$$

Then use properties of Wigner matrices from RMT and Free Probability:

• LSD of a Wigner matrix exists and is the *semi-circular law*.

• Independent Wigner matrices converge jointly (in the sense of expected trace of polynomials).

• They are *asymptotically free* (needed when we have more than one series to deal with).

• They are also free of any deterministic matrices.

Applications: graphical inference

- White noise testing: already demonstrated.

- Other hypothesis testing for simple cases: similar ideas.

Applications: testing with trace

As already seen, limit distributions are identified by obtaining limits of traces.

As a by-product of the proofs, we prove the asymptotic normality of trace (after centering and scaling) of any polynomial.

This can be used for significance testing of suitable simple versus simple hypothesis involving the coefficient matrices.

Upcomimg book (with Monika): *Large covariance and autocovariance matrices*.

THANK YOU !