RMT and it's application in hypothesis testing for high dimensional data

Tiyasa, Samya, Snehashis, Mainak[*]
M.Stat. Course Project

May 17, 2024

1 Introduction

For most of 20^{th} century statisticians have developed mathematically sound and practically feasible large sample testing methodologies for multivariate datasets. Examples of classical multivariate tests include testing equality of means under common variance and testing equality of covariances. However, the solution to these problems were for multivariate datasets whose dimension was considered fixed as sample size increases. From the late 1980's we have seen an exponential increase in high dimensional data whose dimensions are comparable to their sample size. Consequently, the previous results are not applicable in this setup and their application lead to erroneous conclusions. Due to this there was a need to develop new asymptotic results using different techniques where both the dimension p and sample size n increase simultaneously. From early 2000 it was observed that random matrix theory (RMT) provide methods to do so (see (3)).

The field of RMT has significantly advanced over recent decades, driven by its diverse applications in several fields.The origins of RMT can be traced back to Wishart's introduction of the Wishart distribution in 1928, which generalized the gamma distribution to multiple dimensions. This groundwork was further developed by Fisher, Hsu, and Roy in the 1930s, who explored the distribution of eigenvalues for matrices following the Wishart distribution. However, the modern form of RMT emerged in the 1950s with Wigner's pioneering work.

The objective of this project is to obtain a basic understanding of random matrix theory and learn how they are applied to derive asymptotic distribution of test statistics in high dimension setup. While pursuing the project it was collectively felt that there were some gaps between some theoretical results in RMT and statistical application in real datasets which we have tried to fill.

[*] Supervised by Dr. Debashish Paul, Stat-Math Unit, ISI Kolkata

2 Methodology

A **random matrix** is a matrix whose entries are random variables (see (Bose)). One random matrix which is important in statistics is the Wishart matrix. Consider a $m \times n$ matrix X with i.i.d. entries from a normal distribution, then the Wishart matrix W is then defined as $W = XX^T$.

Most multivariate test statistic are functions of eigenvalues of a random matrix. Hence, study of these (random) eigenvalues become very important. We briefly discuss the important ideas in the following subsections.

2.1 Spectral Distribution

We consider A_n to be a n x n random symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ which are **real random** variables. We are interested to quantify the distribution of these random eigenvalues. There are three important notions which are defined as follows :

1. **Empirical Spectral Distribution/Measure (ESD)** assigns mass $\frac{1}{n}$ to each eigenvalue of a matrix A_n. Mathematically, it is defined as:

$$F_{A_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\lambda_i \leqslant x)$$

   Given that λ_i's are random variables, the ESD itself is a random variable.

2. **Expected Empirical Spectral Distribution (EESD)** represents the expected value of the ESD, computed over the joint distribution of eigenvalues. This is expressed as:

$$\mathbb{E}[F_{A_n}(x)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\lambda_i \leqslant x)\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(\lambda_i \leqslant x)$$

3. **Limiting Spectral Distribution (LSD)** of a sequence of square random matrices $\{A_n\}$ is the weak or distributional limit of the sequence $\{\mathbb{E}(F_{A_n})\}$, provided it exists, as $n \to \infty$.

Note that $F_{A_n}(t)$ is a random variable which is a distribution function for every realization. The notion of convergence of these random variables in RMT is defined as either in terms of **almost sure** convergence or **in probability** convergence. However, the existence of LSD is not always guaranteed.

A natural question which arises is how to *find* a LSD for a given sequence of random matrices. There are two popular methods called **moment method** and **stieltjes transform** (see (3)). The LSDs when we have variants of Wishart matrix have already been derived and the study and application of these results have been the primary focal point in this project. Due to this we omit the discussion on method of finding LSD. However, for the sake of completeness we have briefly described them in the supplementary report.

We mentioned that existence and uniqueness results regarding LSD of sequence of Wishart matrix and its variants are already well known. To be specific, there are three results which are relevant for this project, which are as follows ((Bose),(4)) :

1. **Wigner's Semi-Circle law**: A *Wigner* matrix is a type of random matrix where entries are independently and identically distributed (i.i.d) random variables. Let $\{\boldsymbol{W}_n\}_{n=1}^{\infty}$ be a sequence of Wigner matrices. For each n let ν_n denote the empirical spectral measure of $\boldsymbol{A}_n = \frac{\boldsymbol{W}_n}{\sqrt{n}}$. Then, ν_n converges weakly $\mathbb{P}$-almost surely to the semi-circle distribution μ_{sc} with Lebesgue density
$$[\mu_{sc}(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{|x|\leqslant 2}$$
2. **Marchenko-Pastur law**: It states that for covariance type matrix $S_n := n^{-1}AA^T$ under smooth condition on A, then ESD of $S_n \xrightarrow{a.s.} \mathrm{MP}_{y,\sigma^2}$. if $\frac{p}{n} \to y \in (0,\infty)$ and $n,p \to \infty$ again ESD of $\sqrt{\frac{n}{p}}(S_n - \sigma^2 \mathrm{I}_p) \xrightarrow{a.s.} \mathrm{MP}_{y,\sigma^2}$ if $\frac{p}{n} \to 0$
3. LSD of **F Matrices** : Consider $\{X_{ki}, i, k = 1, 2, \ldots\}$ and $\{Y_{ki}, i, k = 1, 2, \ldots\}$ which are independent with mean 0 and variance. We assume the following

   (a) For any fixed $\eta > 0$ and when $n_1, n_2 \to \infty$,
$$\frac{1}{n_1 p}\sum_{j=1}^{p}\sum_{k=1}^{n_1} E\left|X_{jk}\right|^4 \mathbf{1}_{\{|X_{jk}|\geqslant \eta\sqrt{n_1}\}} \to 0, \quad \frac{1}{n_2 p}\sum_{j=1}^{p}\sum_{k=1}^{n_2} E\left|Y_{jk}\right|^4 \mathbf{1}_{\{|Y_{jk}|\geqslant \eta\sqrt{n_2}\}} \to 0.$$

   (b) The sample sizes n_1, n_2 and the dimension p grow to infinity in such a way that
$$y_{n_1} := p/n_1 \to y_1 \in (0, +\infty), \quad y_{n_2} := p/n_2 \to y_2 \in (0, 1).$$

   Under these assumptions, when $\mathbf{n} \to \infty$, almost surely the random ESD $f_{\mathbf{n}}$ of the F-matrix $\mathbf{S}_1\mathbf{S}_2^{-1}$ converges to a distribution $F_{\mathbf{y}}(\mathrm{d}x) = g_{\mathbf{y}}(x)\mathbf{1}_{[a,b]}(x)\mathrm{d}x + (1 - 1/y_1)\,\mathbf{1}_{\{y_1>1\}}\delta_0(\,\mathrm{d}x)$, where
$$h = \sqrt{y_1 + y_2 - y_1 y_2}, \quad a = \frac{(1-h)^2}{(1-y_2)^2}, \quad b = \frac{(1+h)^2}{(1-y_2)^2},$$
$$g_{\mathbf{y}}(x) = \frac{(1-y_2)}{2\pi x\,(y_1 + y_2 x)}\sqrt{(b-x)(x-a)}, \quad a < x < b.$$

2.2 Linear Spectral Statistic

Recall that most LRT statistic are of the form $a_1 Trace(M_1) + a_2 log(det(M_2)) + a_3$, where M_1 and M_2 are random matrices and $a_i's$ are constants. We observe that $Trace(M)$= $\sum_{i=1}^{i=p}\lambda_i$ and $log(det(M))$= $\sum_{i=1}^{i=p} log\lambda_i$, where M is a matrix with the eigenvalues λ_i. Hence, most LRT statistics are linear combination of "additive functions of eigenvalues of random matrices". This property is exploited further to obtain asymptotic distributions.

Let M be a random matrix with eigenvalues λ_i. M is generally functions of (random) observations. A statistic of the form $\sum_{i=1}^{i=p} f(\lambda_i)$ where f is a real valued function and λ_i's are eigen values of sample covariance matrix is called a **linear spectral statistic (LSS).** Hence, both $Trace(M)$ and $log(det(M))$ are LSS for functions $f(x) = x$ and $f(x) = log(x)$ respectively. In RMT there are theorems which help us to obtain CLT for LSS. In the following subsection we discuss a very important result which has been used to derive asymptotic limiting distribution of a finite number of LSS.

A fundamental result Suppose F_y denotes the Marchenko-Pastur law with parameter y. $F^{\mathbf{S}_n}(f) = \sum_{i=1}^{i=p} f(\lambda_i)$ where $\lambda_i's$ are eigenvalues of $\mathbf{S}_n$. $y_n = \frac{p}{n}$ and $y = lim_{n\to\infty}\frac{p}{n}$ κ is 2 if the matrices are real and 1 if the matrices are complex. Assume that the variables $\{x_{ij}\}$ of the data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are independent and identically distributed satisfying $Ex_{ij} = 0, E\left|x_{ij}\right|^2 = 1, E\left|x_{ij}\right|^4 = \beta + 1 + \kappa < \infty$, and in case of complex variables, $Ex_{ij}^2 = 0$.

Let $f_1, \ldots f_k$ be functions analytic on an open region containing the support of F_y. The random vector $\{X_n(f_1), \ldots X_n(f_k)\}$ where $X_n(f) = p\left\{F^{\mathbf{S}_n}(f) - F_{y_n}(f)\right\}$ converges weakly to a Gaussian vector $(X_{f_1}, \ldots X_{f_k})$ with mean function and covariance function:

$$\mathbb{E}\left[X_f\right] = (\kappa - 1)I_1(f) + \beta I_2(f), \operatorname{cov}\left(X_f, X_g\right) = \kappa J_1(f,g) + \beta J_2(f,g)$$

where

$$I_1(f) = -\frac{1}{2\pi i}\oint \frac{y\{\underline{s}/(1+\underline{s})\}^3(z)f(z)}{[1 - y\{\underline{s}/(1+\underline{s})\}^2]^2}dz, I_2(f) = -\frac{1}{2\pi i}\oint \frac{y\{\underline{s}/(1+\underline{s})\}^3(z)f(z)}{1 - y\{\underline{s}/(1+\underline{s})\}^2}dz,$$

$$J_1(f,g) = -\frac{1}{4\pi^2}\oint\oint \frac{f(z_1)\, g(z_2)}{\left(\underline{m}(z_1) - \underline{m}(z_2)\right)^2}\underline{m}'(z_1)\, \underline{m}'(z_2)\, dz_1 dz_2$$

$$J_2(f,g) = \frac{-y}{4\pi^2}\oint f(z_1)\frac{\partial}{\partial z_1}\left\{\frac{\underline{s}}{1+\underline{s}}(z_1)\right\}dz_1, \cdot \oint g(z_2)\frac{\partial}{\partial z_2}\left\{\frac{\underline{s}}{1+\underline{s}}(z_2)\right\}dz_2,$$

where the integrals are along contours (non-overlapping in J_1) enclosing the support of F_y.

Using these fundamental result, the join asymptotic distribution have been already derived (see (4) and (5)), as well as the joint asymptotic distribution of LSS for beta matrices $\boldsymbol{\beta}_{\mathbf{n}} = \mathbf{S}_2\left(\mathbf{S}_2 + d\cdot\mathbf{S}_1\right)^{-1}$.

Apart from LSS, another type of statistic which occur frequently are based on extreme eigenvalues which we discuss in the next section.

2.3 Extreme eigenvalues and Tracy-Widom distributions

Let $\mathbf{A} \sim W_p(\mathbf{I}, m)$ be independent of $\mathbf{B} \sim W_p(\mathbf{I}, n)$, where $m \geqslant p$. Then the largest eigenvalue θ of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$ is called the *greatest root statistic* and its distribution is denoted $\theta(p, m, n)$.

In several classical multivariate tests Roy's greatest root statistic plays a central role. For example in *test for Independence of two sets of variables*, The union-intersection is based on the largest eigenvalue of $S_{22}^{-1}S_{21}S_{11}^{-1}S_{12}$ which under H_0 follows $\theta(p_2, n-1-p_1, p_1)$. Similarly, *test for equality of covariance matrices* is based on the largest eigenvalue θ of $(n_1S_1+n_2S_2)^{-1}n_2S_2$, which under H_0 has the $\theta(p, n1, n2)$ distribution. Hence, to find the asymptotic distribution we need the limit laws for extreme eigenvalues of high-dimensional random matrices.

The joint distribution of eigenvalues of a random matrix are called ensembles (see (3)). If the matrix is orthogonal (or unitary), they are called orthogonal(or unitary) ensemble. These ensembles, as well as the associated Jacobi ensembles described below, are expressed as specific cases of the more general form of the joint density of the eigenvalues $x_1, \ldots, x_N$ of H which is given by

$$f_{N,\beta,w}\left(x_1,\ldots,x_N\right) = c_{N,\beta,w}\prod_{j<k}\left|x_j-x_k\right|^{\beta}\prod_{j}\left(w\left(x_j\right)\right)^{\beta/2},$$

where $c_{N,\beta,w}$ is a proportionality constant and $w(x)$ is a non-negative weight function, and $\beta=1$ and $\beta=2$ correspond to the symmetric (orthogonal) and Hermitian (unitary) matrix ensembles, respectively. For GOE (corresponding to $\beta=1$) and GUE (corresponding to $\beta=2$), the weight function $w(x)=\exp\left(-x^2/2\right)$.

The *Jacobi Orthogonal Ensembl*e (JOE) and *Jacobi Unitary Ensemble* (JUE) refer to the joint density of eigenvalues of the matrix $\mathbf{T}=\mathbf{U}(\mathbf{U}+\mathbf{V})^{-1}$ where $\mathbf{U}=\mathbf{X}\mathbf{X}^*$ and $\mathbf{V}=\mathbf{Y}\mathbf{Y}^*$, where $\mathbf{X}$ and $\mathbf{Y}$ are independent$p\times m$ and $p\times n$ matrices, with i.i.d. real (corresponding to orthogonal) or complex (corresponding to unitary) standard Gaussian entries. The asymptotic distribution of such extreme eigenvalues belong to a family of distributions called Tracy-Widom distributions which are briefly explained below

2.3.1 The Tracy-Widom distributions

These are the distribution of the normalized largest eigenvalue of a random Hermitian matrix. Let $F_\beta(.)$ denote the cumulative distribution function of the Tracy–Widom distribution with given β.The c.d.f. of the Tracy-Widom distribution corresponding to the GOE and LOE, denoted by F_1, is given by (3)

$$F_1(s)=\exp\left(-\frac{1}{2}\int_s^\infty\left(q(x)+(x-s)q^2(x)\right)dx\right), \quad s\in\mathbb{R}$$

where $q(x)$ satisfies the Painleve differential equation $q''(x)= xq(x)+2q^3(x)$ such that $q(x)-A(x)\to 0$ as $x\to\infty$, where $A(x)$ is the Airy function.

Now we mention two important limit theorems (see (2), (4)) which help us obtain asymptotic distribution of Roy's greatest root statistic.

Theorem 2.1. *1. If the the entries of* $\mathbf{X}$ *are i.i.d. real Gaussian with mean* 0 *and variance* 1*, and* $l_{1,p}$ *denotes the largest eigenvalue of* $\mathbf{X}\mathbf{X}^T$*, then as* $n\to\infty$*, so that* $p/n\to(0,1]$,

$\frac{l_{1,p}-\mu'_{n,p}}{\sigma'_{n,p}} \Longrightarrow W_1$, *where* $\mu'_{n,p} = (\sqrt{n-1}+\sqrt{p})^2, \quad \sigma'_{n,p} = (\sqrt{n-1}+\sqrt{p})\left(\frac{1}{\sqrt{n-1}}+\frac{1}{\sqrt{p}}\right)^{1/3}.$
Here the random variables W_1 *are distributed with c.d.f.* F_1.

2. *Suppose* $W(p,m,n) = \operatorname{logit}\theta(p,m,n) = \log\left(\frac{\theta(p,m,n)}{1-\theta(p,m,n)}\right)$. $\frac{W(p,m,n)-\mu(p,m,n)}{\sigma(p,m,n)} \overset{D}{\Rightarrow} F_1$ *The centering and scaling parameters are defined by* $\mu(p,m,n) = 2\log\tan\left(\frac{\varphi+\gamma}{2}\right)$, $\sigma^3(p,m,n) = \frac{16}{(m+n-1)^2}\frac{1}{\sin^2(\varphi+\gamma)\sin\varphi\sin\gamma}$ *where the angle parameters* γ, φ *are defined by*

$$\sin^2\left(\frac{\gamma}{2}\right) = \frac{\min(p,n)-1/2}{m+n-1}, \sin^2\left(\frac{\varphi}{2}\right) = \frac{\max(p,n)-1/2}{m+n-1}$$

The more formal statement of part (2) goes as follows. Assume p, m and $n \to \infty$ together in such a way that $\lim \frac{p\wedge n}{m+n} > 0, \quad \lim \frac{m}{p} > 1$..For each $s_0 \in \mathbb{R}$, there exist $c, C > 0$ such that for $s \geqslant s_0$,

$$|P\{W(p,m,n) \leqslant \mu(p,m,n) + \sigma(p,m,n)s\} - F_1(s)| \leqslant Cp^{-2/3}e^{-cs}.$$

Now we present some theoretical results derived in this project and some simulation studies.

3 Results

In this section we first discuss how to apply the tests based on Roy's greatest root statistic on real data followed by some theoretical results we derived regarding LSS for some classical tests

3.1 Practical considerations regarding Roy's greatest root statistic

- Let f_α denote the α^{th} percentile of F_1. Then the α^{th} percentile of $\theta(p,m,n)$ is given approximately given by

$$\theta_\alpha = e^{\mu+f_\alpha\sigma}/\left(1+e^{\mu+f_\alpha\sigma}\right),$$

  where $\mu = \mu(p,m,n), \sigma = \sigma(p,m,n)$.

- When we have a real life data and we are interested to perform any of the tests based on the Roy's greatest root statistic mentioned above we need to find the cut off point θ_α at α level of significance.Prior to that we need the value of f_α which has to be approximated by using the *qtw* function available in the *RMTstat* pacakge in R. in R. Using this we compute θ_α which is of the form

$$\theta_\alpha = \frac{1}{1+cot^2(\frac{\theta+\gamma}{2})\exp\{-f_\alpha\sigma\}}$$

- So from the data we can compute the largest eigen value and evaluate the Critical Region $\{\alpha : \theta(p,m,n) \geqslant \theta_\alpha\}$.

3.2 Asymptotic rejection regions for two classical multivariate tests in high-dimensional setup

We have the following two results. The proofs are presented in the supplementary report.

Theorem 3.1. *$(\tilde{L}_{id} - p(\mu - 1)) \xrightarrow{\mathcal{D}} N(\mu_1, \sigma^2)$ where $\mu = 2 - \frac{y_n - 1}{y_n} \log(1 - y_n)$, $y_n = p/n$ and $\mu_1 = -\frac{1}{2}\log(1 - y_n)$, $\sigma^2 = -2(\log(1 - y_n) + y_n)$*

Accept H_0 if L_{sp} lies outside the interval $(p\mu + \mu_1 - \sigma, p\mu + \mu_1 + \sigma)$ and reject otherwise

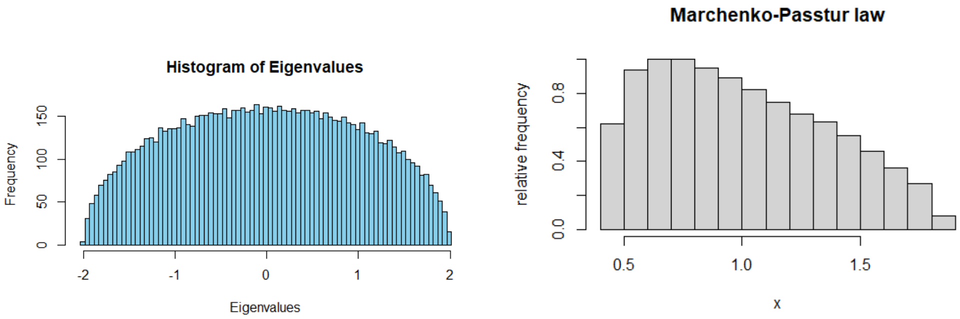
Theorem 3.2. *$(L_{eq} - p\mu') \xrightarrow{\mathcal{D}} N(\mu'', \sigma^2)$ where $\mu' = -\frac{N_1}{N} + F_{y_{2N}} \log(1 + \frac{N_1}{N}x)$, $\mu'' = -\frac{N_1}{N}\mu_1 + \mu_2$ and $\sigma^2 = \left(\frac{N_1}{N}\right)^2 \{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}\}$*

accept H_0 if L_{eq} lies outside the interval $(p\mu + \mu_1 - \sigma, p\mu + \mu_1 + \sigma)$ and reject otherwise.

3.3 Simulations

Now we present some simulations which we have performed. Details regarding each parameter in the simulations are present in the supplementary report along with the R codes.

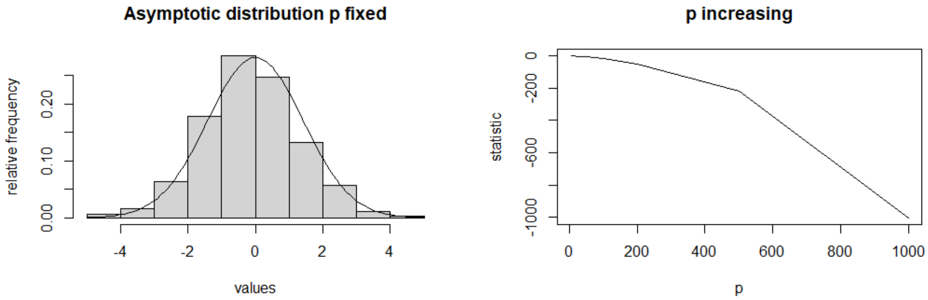
3.3.1 Limiting Spectral Distribution



(a) Semi-Circle Law

(b) Marchenko-Pastur

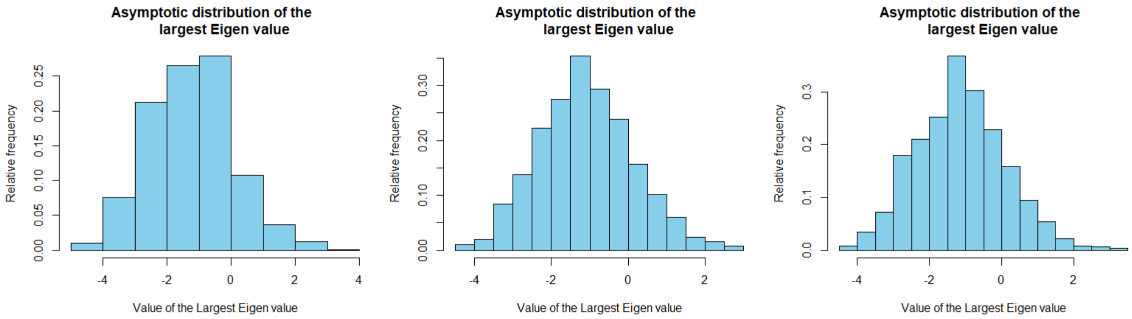
3.3.2 An illustration where classical asymptotic result do not hold



(a) p fixed

(b) p varying

3.3.3 Asymptotic distribution of the largest eigenvalue as the ratio $\frac{p}{n}$ changes



(a) $\frac{p}{n} = 0.2$ (b) $\frac{p}{n} = 0.8$ (c) $\frac{p}{n} = 1$

4 Acknowledgement

We would like to thank our supervisor and course instructor professor Debashish Paul for suggesting this topic and his helpful comments. We would also like to thank professor Arup Bose, Soumendu Sundar Mukherjee and senior research fellows for their lectures and discussions in the reading group on random matrices at SMU, ISI.

References

[Bose] Bose, A. *Random Matrices and Non-Commutative probability.*

[2] Johnstone, I. M. (2009). Approximate null distribution of the largest root in multivariate analysis. *The Annals of Applied Statistics*, 3(4).

[3] Namdari, J., Paul, D., and Wang, L. (2020). High-dimensional linear models: A random matrix perspective. *Sankhya A*, 83:645 – 695.

[4] Yao, J., Zheng, S., and Bai, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis.*

[5] Zheng, S., Bai, Z., Yao, J., and Zhu, H. (2017). Clt for linear spectral statistics of large dimensional sample covariance matrices with dependent data.