STAT7017 Big Data Statistics	**Semester 2 2018**

Lecture 3: 6 August

Lecturer: Dr Dale Roberts *Scribes: Rui Qiu*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Limiting Spectral Distribution

$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ n random samples of dimension p (vectors). Sample covariance matrix:

$$S_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^*$$

which gives you a $p \times p$ matrix.

$\bar{\mathbf{x}}$ is sample mean given by $\bar{\mathbf{x}} := \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$.

Many traditional multivariate statistics are functions of the eigenvalues (λ_i) of S_n.

In the most basic form, $T_n = \frac{1}{p}\sum_{k=1}^{p}\phi(\lambda_k)$. This is just a generalized form, because we don't know what ϕ is. But it takes an eigenvalue and gives back a number. $\phi : \mathbb{C} \to \mathbb{R}$.

Example: The generalized variance (last week) can be written

$$T_n = \frac{1}{p}\log|S_n| = \frac{1}{p}\sum_{k=1}^{p}\log(\lambda_k)$$

T_n is a "linear spectral statistic of the sample covariance matrix S_n with test function $\phi(x) = \log(x)$".

First order Random matrix limits are concerned with when and how shall $T_n \to c$ (converges to some constant c) as $p, n \to \infty$.

It concerns the "joint limit" of the p eigenvalues. $(\lambda_k)_{k=1}^{p}$.

Empirical distributions and their limits

Let $\mathbb{M}_p(\mathbb{C})$ be $p \times p$ matrices with $\mathbb{C}$-valued entries and let $(\lambda_k)_{k=1}^{p}$ be the eigenvalues of $A \in \mathbb{M}_p(\mathbb{C})$.

Let $\delta_a(x) = \begin{cases} 1 \text{ if } x = a \\ 0 \text{ otherwise} \end{cases}$

The empirical spectral distribution (ESD) of A is given by

$$F^A_{(x)} := \frac{1}{p}\sum_{k=1}^{p}\delta_{\lambda_k}(x)$$

Generally, F^A takes $\mathbb{C}$ values. If $A \in \mathbb{H}_p$, then $F^A(x) \in \mathbb{R}$.

Example: $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ eigenvalues are $-i, +i$.

$$F^A = \frac{1}{2}(\delta_i + \delta_{-1})$$

(empirical density "histogram" vs limiting density)

Take a sequence of matrices $(A_n)_{n\geq 1} \in \mathbb{M}_p(\mathbb{C})$, if the sequence of corresponding ESD F^{A_n} vaguely converges to a (possibly defective) measure F, we call F the limiting spectral distribution (LSD) of $(A_n)_{n\geq 1}$.

Vague convergence means that for any continuous function that is compactly supported, called ϕ,

$$F^{A_n}(\phi) \to F(\phi) \text{ as } n \to \infty.$$

Here, we use the notation

$$F(\phi) := \int_{\mathbb{R}^p} \phi(x)F(dx).$$

(**compact supported**: zero outside the range $[a, b]$.)

If the distribution F is **non-defective** (i.e. $\int F(dx) = 1$.) then vague convergence becomes **weak convergence**, that is,

$$F^{A_n}(\phi) \to F(\phi) \text{ as } n \to \infty$$

for all ϕ continuous and **bounded** (below value a, say).

In our situation, we shall be dealing with sample covariance matrices (S_n). This means that:

- support of F^{S_n} is $\mathbb{R}_+$ since S_n are Hermitian and non-negative definite.
- $F^{S_n}(x) = \frac{1}{p}\sum_{k=1}^{p} \mathbf{1}_{(\lambda_k \leq x)}$ ESD.
- Eigenvalues are random variable and ESDs (F^{S_n}) are random probability distributions on $\mathbb{R}_+$.

The fundamental question is: Does the limit of (F^{S_n}) exist?

How can we show this?

The eigenvalues of a matrix are continuous functions of the entries of the matrix.

There is no closed-form solution for eigenvalues when dimension of a square matrix is greater than 4.

There are three main techniques used in RMT:

- Method of moments.
- Orthogonal polynomial decomposition.
- Stieltjes transform. (ST)

We shall focus on the ST approach.

Stieltjes transfrom (ST)

The ST plays nearly as useful role in RMT as the Moment generating function (MGF) or characteristic function (CF) in classic probability theory.

It is defined for a measure μ as:

$$S_\mu(z) = \int \frac{1}{x-z}\mu(dx), \ z \in \mathbb{C}^+$$

where $\mathbb{C}^+ = \{x + iy : y > 0\}$.

The following lemma allows us to reconstruct the distribution function from its Stieltjes transform.

Lemma 3.1 *(Inversion): Let μ be a probability measure on $\mathbb{R}$. If $a < b$ are points of continuity of the associated distribution, then*

$$\mu((a,b)) = \lim_{\nu \to 0^+} \frac{1}{\pi} \int_a^b Im(S_\mu(x + i\nu))dx.$$

The following lemma gives a necessary and sufficient condition for a sequence of ST to be the ST of a probability measure.

Lemma 3.2 *(Geronimo and Hill, 2003): Suppose that (μ_n) is a sequence of probability measures on $\mathbb{R}$ with Stieltjes transforms (S_{μ_n}). If $\lim_{n\to\infty} S_{\mu_n}(z) = S_\mu(z)$ for all $z \in \mathbb{C}^+$, then there exists a probability measure μ with ST given by S_μ if and only if*

$$\lim_{\nu\to\infty} i\nu S_\mu(i\nu) = -1.$$

In which case, $\mu_n \to \mu$ in distribution.

There are some more technical results that I will now state without proof.

First, we say that a function f is **holomorphic** if it is complex differentiable at every point of its domain, i.e.

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} \text{ exists.}$$

Holomorphic functions are very nice:

- Infinitely differentiable.
- Equals to its Taylor series.

Proposition 3.3 *The Stieltjes transform has the following properties:*

- S_μ *is holomorphic on* $\mathbb{C} \backslash \Gamma_\mu$ *where* $\Gamma_\mu := Supp(\mu)$.
- $z \in \mathbb{C}^+ \iff S_\mu(z) \in \mathbb{C}^+$.
- *If* $\Gamma_\mu \subset \mathbb{R}_+$ *and* $z \in \mathbb{C}^+$, *then* $zS_\mu(z) \in \mathbb{C}^+$.
- $|S_\mu(z)| \leq \frac{\mu(1)}{dist(z,\Gamma_\mu) \vee |Im(z)|}$ *(Distance of z to support and the maximum of imaginary part of z)*

Proposition 3.4 *The mass* $\mu(1)$ *can be recovered through the formula*

$$\mu(1) = \lim_{\nu \to \infty} -i\nu S_\mu(i\nu)$$

Moreover, for all continuous and compactly supported $\phi : \mathbb{R} \to \mathbb{R}$

$$\mu(\phi) = \int \phi(x)\mu(dx) = \lim_{\nu \to 0} \frac{1}{\pi} \int \phi(x) Im[S_\mu(x + i\nu)]dx$$

Proposition 3.5 *Assume that the following conditions hold for a complex-valued* $g(z)$*:*

- g *is holomorphic on* $\mathbb{C}^+$.
- $g(z) \in \mathbb{C}^+$ *for all* $z \in \mathbb{C}^+$.
- $\lim_{\nu \to \infty} \sup |i\nu g(i\nu)| < \infty$.

The g *is a ST of a bounded measure on* $\mathbb{R}$.

Theorem 3.6 *A sequence of measures* (μ_n) *converges vaguely to some positive measure* $\mu \iff (S_{\mu_n})$ *converges to* S_μ *on* $\mathbb{C}^+$.

The idea is that we show $S_{\mu_n} \to S_\mu$ (vague convergence) and then show that μ is a probability measure by checking that $\mu(1) = 1$.

We have A positive semidefinite and symmetric. Then ESD of A is $F^A = \frac{1}{p}\sum_{j=1}^p \delta_{\lambda_j}$, where A is $p \times p$.

$$\begin{aligned} S_A(z) &= \int \frac{1}{x-z} F^A(dx) \\ &= \frac{1}{p}\sum_{k=1}^{p}\int \frac{1}{x-z}\delta_{\lambda_k}(dx) \\ &= \frac{1}{p}\sum_{k=1}^{p}\frac{1}{\lambda_k - z} \\ &= \frac{1}{p}tr[(A-zI)^{-1}] \end{aligned} \tag{3.1}$$

Note: $tr(A+B) = tr(A) + tr(B), tr(A^k) = \sum_{i=1}^{p}\lambda_i^k$.

Trace of an inverse matrix: For $n \times n$ matrix Q, define Q_k to be the submatrix obtained by deleting k-th row and column.

Theorem 3.7 *(Bai and Silvestein, Thm A.U.???): If B and $B_k, k = 1, \ldots, n$, are nonsingular and writing $B^{-1} = [b^{kl}]$, then*

$$tr(B^{-1}) = \sum_{k=1}^{n} \frac{1}{b_{kk} - B_k' B_k^{-1} B_k}$$

b_{kk}: k-th diagonal entry of B.
B_k': vector obtained from k-th row of B by deleting k-th entry.
B_k: vector obtained from k-th column of B by deleting k-th entry.

Applying this theorem:

$$S_A(z) = \frac{1}{p}\sum_{k=1}^{p}\frac{1}{\alpha_{kk} - z - \alpha_k'(A_k - zI)^{-1}\alpha_k} \tag{3.2}$$

We would like to show that denominator is equal to

$$g(z, S_A(z)) + o(1)$$

Then we can solve for $S_A(z) = \frac{1}{g(z,S_A(z))}$ to obtain the ST of the ESD.

Marchenko-Pastur distributions

The Marchenko-Pastur distribution F_{y,σ^2} with index y and scale parameter σ has density

$$P_{y,\sigma^2}(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)'}, & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

Note $a = \sigma^2(1-\sqrt{y})^2,\ b = \sigma^2(1+\sqrt{y})^2$.

If $\sigma^2 = 1$: standard MP distribution.

Special case: $y = 1, \sigma^2 = 1$.

$$P_1(x) = \begin{cases} \frac{1}{2\pi x}\sqrt{x(4-x)}, & 0 < x \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

$\implies$ Hence density is unbounded in region.

As $y \to 0, F_y \to \delta_1$.

MP distribution for independent vectors without cross-correlation

$$\begin{aligned} Sn = &= \frac{1}{n-1}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}^* - \cdots \\ &\approx \frac{1}{n-1}\sum \mathbf{x}_i\mathbf{x}^* \end{aligned} \tag{3.3}$$

We shall sometimes write n sample vectors as $p \times n$ random matrix

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$

So

$$\implies S_n = \frac{1}{n}\mathbf{X}\mathbf{X}^*$$

Marchenko and Pastur found the LSD of the large sample covariance matrix S_n.

Theorem 3.8 *(MP) Suppose that the entries $[x_{ij}]$ of $\mathbf{X}$ are iid complex random variables with mean zero and variance σ^2, and $p/n \to y \in (0, \infty)$. Then, almost surely,*

$$F^{S_n} \to F_{y,\sigma^2}$$

This theorem was shown in a special case in 1960s but its influence in statistics was only recognized recently.

How does the MP dist. appear in the limit?

$\sigma^2 = 1$.

$$\begin{aligned} P_y(x) &= \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}, a \leq x \leq b \\ a &= (1-\sqrt{y})^2, b = (1+\sqrt{y})^2 \end{aligned} \tag{3.4}$$

The Stieltjes transform is

$$\begin{aligned} S(z) &= \int_a^b \frac{1}{x-z} P_y(x)dx \\ &= \frac{(1-y) - z + \sqrt{(z-1-y)^2 - 4y}}{2yz} \end{aligned} \tag{3.5}$$

rearranging notice that $s = S(z)$ satisfies the quadratic equation

$$yzs^2 + (z+y-1)s + 1 = 0$$

The ST of the ESD of S_n is $S_n(z) = \frac{1}{p} tr[(S_n - zI_p)^{-1}]$.

If we can show $S_n(z) \to S(z)$ as $n \to \infty$ for every $z \in \mathbb{C}^+$, then $F^{S_n} \to F_y$.

By (3.2),

$$S_n(z) = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\frac{1}{n}\alpha_k' \bar{\alpha}_k - z - \frac{1}{n^2}\alpha_k' \mathbf{X}_k^* \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - zI_{p-1}\right)^{-1} \mathbf{X}_k \bar{\alpha}_k}$$

$\mathbf{X}_k = \mathbf{X}$ with k-th row removed.

$\alpha_k' = k$-th row of $\mathbf{X}$, size $n \times 1$.

Assume $\mathbb{E}$[“denominator terms with rows removed”]$\to$ $\mathbb{E}$[“terms with rows intact”],

i.e. random error caused by approximation is small for large p and n.

$$\mathbb{E}\left[\frac{1}{n}\alpha_k' \bar{\alpha}_k\right] = \frac{1}{n}\sum_{j=1}^{n} |x_{kj}|^2 = 1.$$

Lemma 3.9 *Let u be a $n \times 1$ random vector with entries u_i that are all independent with mean* 0 *and unit variance. Let* $\mathbf{Q}$ *be a (non-random)* $n \times n$ *complex matrix. Then*

$$\mathbb{E}\left[u^* \mathbf{Q} u\right] = tr\mathbf{Q}.$$

Note $\mathbf{A}, \mathbf{B}$ matrices, $(\mathbf{AB})_{ik} = \sum_{j=1}^m a_{ij} b_{jk}$.

Proof: As $u^*\mathbf{Q}u = \sum_{i=1}^n \sum_{j=1}^n \bar{u}_i \mathbf{Q}_{ij} u_j$.

$$\begin{aligned} \mathbb{E}\left[u^*\mathbf{Q}u\right] &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\left[\mathbf{Q}_{ij}\bar{u}_i u_j\right] \\ &= \sum_{i=1}^n \mathbf{Q}_{ii}\mathbb{E}\left[\bar{u}_i u_i\right] \\ &= tr\mathbf{Q} \text{ as } \mathbb{E}[\bar{u}_i u_i] = 1. \end{aligned} \tag{3.6}$$

■

Corollary 3.10 $\mathbb{E}\left[u^* u\right] = n$.

Proof: Take $\mathbf{Q} = \mathbf{I}_n$, then $tr\mathbf{Q} = tr\mathbf{I}_n = n$. ■

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n^2}\alpha_k' \mathbf{X}_k^* \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k \bar{\alpha}_k\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[tr\left\{\mathbf{X}_k^* \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k \bar{\alpha}_k \alpha_k'\right\}\right] \\
&= \frac{1}{n^2} tr\left\{\mathbb{E}\left[\mathbf{X}_k^* \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1}\right] \mathbb{E}\left[\bar{\alpha}_k \alpha_k'\right]\right\} \\
&= \frac{1}{n^2} tr\left\{\mathbb{E}\left[\mathbf{X}_k^* \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k\right]\right\} \\
&= \frac{1}{n^2}\mathbb{E}\left[tr\left\{\mathbf{X}_k^* \left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k\right\}\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[tr\left\{\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k \mathbf{X}_k^*\right\}\right]
\end{aligned}
\tag{3.7}
$$

We note that $\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* \approx S_n$ (only 1 vector removed).

So

$$
\begin{aligned}
&\frac{1}{n^2}\mathbb{E}\left[tr\left\{\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} \mathbf{X}_k \mathbf{X}_k^*\right\}\right] \\
&\approx \frac{1}{n^2}\mathbb{E}\left[tr\left\{\left(\frac{1}{n}\mathbf{X}\mathbf{X}^* - z\mathbf{I}_{p}\right)^{-1} \mathbf{X}\mathbf{X}^*\right\}\right] \\
&= \frac{1}{n}\mathbb{E}\left[tr\left\{\mathbf{I}_p + z\left(\frac{1}{n}\mathbf{X}\mathbf{X}^* - z\mathbf{I}_{p}\right)^{-1}\right\}\right] \\
&= \frac{p}{n} + z\frac{p}{n}\mathbb{E}\left[S_n(z)\right].
\end{aligned}
\tag{3.8}
$$

So denominator is roughly

$$1 - z - \left\{\frac{p}{n} + z\frac{p}{n}\mathbb{E}\left[S_n(z)\right]\right\}$$

as $p \to \infty, n \to \infty$ and $p/n \to y > 0$.

$$\mathbb{E}\left[S_n(z)\right] \to S(z)$$

So denominator

$$\to 1 - z - (y + zyS(z))$$

and

$$S(z) = \frac{1}{1 - z - (y + yzS(z))}$$

This is ST of MP distribution F_y!