Probability and Stochastic Process II: Random Matrix Theory and Applications Lecture 3: MP and semicircle laws

Zhenyu Liao, Tiebin Mi, Caiming Qiu

School of Electronic Information and Communications (EIC)
Huazhong University of Science and Technology (HUST)

March 7, 2023

Outline

What we will have today

- » sample covariance matrix and the limiting Marčenko–Pastur law
- » Wigner matrix and the limiting semicircle law
- » proof via Bai and Silverstein approach and/or Gaussian tool

What we will have today

» sample covariance matrix and the limiting Marčenko–Pastur law

» Wigner matrix and the limiting semicircle law

» proof via Bai and Silverstein approach and/or Gaussian tool

What we will have today

» sample covariance matrix and the limiting Marčenko–Pastur law

» Wigner matrix and the limiting semicircle law

» proof via Bai and Silverstein approach and/or Gaussian tool

Outline

SCM and MP law

Proof of Marčenko–Pastur law

Proof of semicircle law

Generalized MP for SCM

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \le \|\mathbf{A}\| \le p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \le \|\mathbf{A}\| \le p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

» **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,

» Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

» In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{eigenvalue mismatch and not consistent!}$$

» due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

» **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,

» Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

» In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

» due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p \|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").

- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \leq \|\mathbf{A}\| \leq p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

Sample covariance matrix in the large n, p regime

- **Problem**: estimate covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ from n data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$,
- Maximum likelihood sample covariance matrix with entry-wise convergence

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \in \mathbb{R}^{p \times p}, \quad [\hat{\mathbf{C}}]_{ij} \to [\mathbf{C}]_{ij}$$

  almost surely as $n \to \infty$: optimal for $n \gg p$ (or, for p "small").
- In the regime $n \sim p$, conventional wisdom breaks down: for $\mathbf{C} = \mathbf{I}_p$ with $n < p$, $\hat{\mathbf{C}}$ has at least $p - n$ zero eigenvalues:

$$\boxed{\|\hat{\mathbf{C}} - \mathbf{C}\| \not\to 0, \quad n, p \to \infty} \Rightarrow \text{ eigenvalue mismatch and not consistent!}$$

- due to $\|\mathbf{A}\|_\infty \le \|\mathbf{A}\| \le p\|\mathbf{A}\|_\infty$ for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\|\mathbf{A}\|_\infty \equiv \max_{ij} |\mathbf{A}_{ij}|$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x} \sqrt{(x - E_-)^+ (E_+ - x)^+} dx$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1-\sqrt{\mathbf{c}})^2, E_+ = (1+\sqrt{\mathbf{c}})^2]$.
- for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm \mathbf{2}\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+ (E_+ - x)^+} dx$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

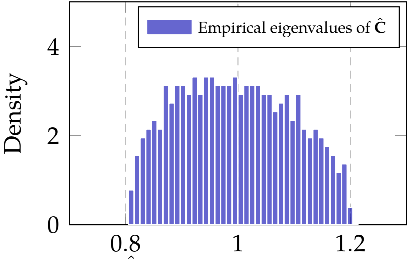

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1 - \sqrt{\mathbf{c}})^2, E_+ = (1 + \sqrt{\mathbf{c}})^2]$.
- for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm \mathbf{2}\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x} \sqrt{(x - E_-)^+ (E_+ - x)^+} dx$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

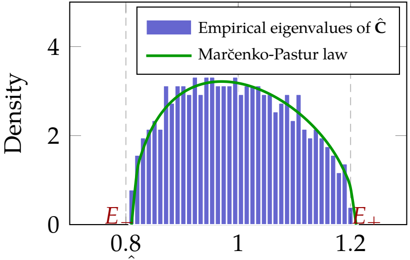

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1-\sqrt{c})^2, E_+ = (1+\sqrt{c})^2]$.
- for $n = 100p$, on a range of $\pm 2\sqrt{c} = \pm 0.2$ around the population eigenvalue 1.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x} \sqrt{(x - E_-)^+ (E_+ - x)^+} dx$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

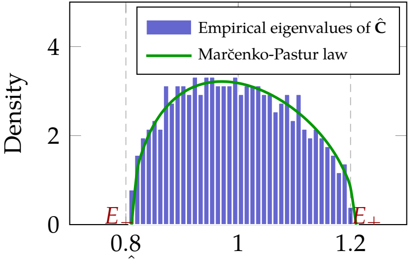

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

» eigenvalues span on $[E_- = (1-\sqrt{c})^2, E_+ = (1+\sqrt{c})^2]$.

» for $n = 100p$, on a range of $\pm 2\sqrt{c} = \pm 0.2$ around the population eigenvalue 1.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x} \sqrt{(x - E_-)^+ (E_+ - x)^+} dx$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

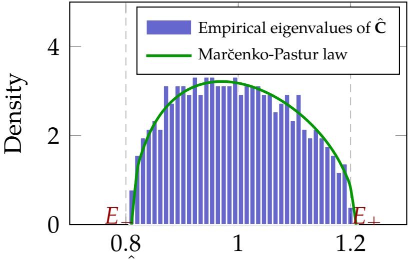

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

» eigenvalues span on $[E_- = (1 - \sqrt{\mathbf{c}})^2, E_+ = (1 + \sqrt{\mathbf{c}})^2]$.

» for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm 2\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

What about $n = 100p$? For $\mathbf{C} = \mathbf{I}_p$, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$: MP law

$$\mu(dx) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+ (E_+ - x)^+} dx$$

where $E_- = (1 - \sqrt{c})^2$, $E_+ = (1 + \sqrt{c})^2$ and $(x)^+ \equiv \max(x, 0)$. Close match!

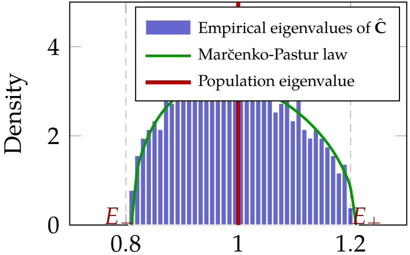

Figure: Eigenvalue distribution of $\hat{\mathbf{C}}$ versus Marčenko-Pastur law, $p = 500$, $n = 50\,000$.

- eigenvalues span on $[E_- = (1-\sqrt{\mathbf{c}})^2, E_+ = (1+\sqrt{\mathbf{c}})^2]$.
- for $\mathbf{n} = \mathbf{100p}$, on a range of $\pm 2\sqrt{\mathbf{c}} = \pm \mathbf{0.2}$ around the population eigenvalue $\mathbf{1}$.

Marčenko–Pastur law

Let $\mathbf{X} \in \mathbb{R}^{p\times n}$ be a random matrix with i.i.d. entries of zero mean and σ^2 variance. Then, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, with probability one, the empirical spectral measure $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}}$ of $\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}$ converges weakly to the probability measure μ

$$\mu(dx) = (1 - c^{-1})^+ \delta_0(x) + \frac{1}{2\pi c\sigma^2 x}\sqrt{(x - \sigma^2 E_-)^+(\sigma^2 E_+ - x)^+}\, dx, \tag{1}$$

where $E_\pm = (1 \pm \sqrt{c})^2$ and $(x)^+ = \max(0, x)$. In particular, with $\sigma^2 = 1$,

$$\boxed{\mu(dx) = (1 - c^{-1})^+ \delta_0(x) + \frac{1}{2\pi c x}\sqrt{(x - E_-)^+\,(E_+ - x)^+}\, dx,} \tag{2}$$

which is known as the Marčenko-Pastur law.

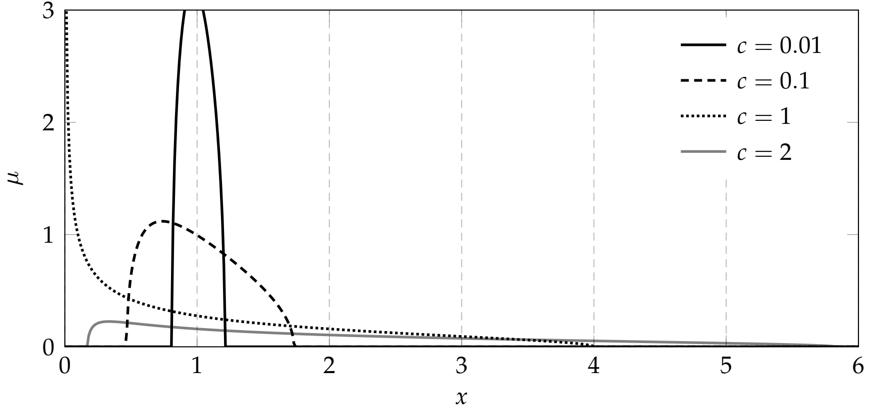

Figure: Marčenko-Pastur distribution for different values of c.

Outline

Proof of Marčenko–Pastur law

Workflow: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ and ST $\frac{1}{p}\operatorname{tr}\mathbf{Q}_{\mathbf{X}}(z) = m_{\mathbf{X}}(z)$ $\Rightarrow$ study the limiting ST $m_{\mathbf{X}}(z) \to m(z) \Rightarrow$ inverse ST to get limiting $\mu_{\mathbf{X}} \to \mu$.

For symmetric $\mathbf{X} \in \mathbb{R}^{p\times p}$, the *empirical spectral distribution* (*ESD*) $\mu_{\mathbf{X}}$ of $\mathbf{X}$ is defined as the normalized counting measure of the eigenvalues $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ of $\mathbf{X}$, i.e., $\mu_{\mathbf{X}} \equiv \frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i(\mathbf{X})}$, where δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

For a real probability measure μ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t-z}. \tag{3}$$

Stieltjes transform

Proof of Marčenko–Pastur law

Workflow: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ and ST $\frac{1}{p}\operatorname{tr}\mathbf{Q}_{\mathbf{X}}(z) = m_{\mathbf{X}}(z)$ $\Rightarrow$ study the limiting ST $m_{\mathbf{X}}(z) \to m(z) \Rightarrow$ inverse ST to get limiting $\mu_{\mathbf{X}} \to \mu$.

For symmetric $\mathbf{X} \in \mathbb{R}^{p\times p}$, the *empirical spectral distribution* (*ESD*) $\mu_{\mathbf{X}}$ of $\mathbf{X}$ is defined as the normalized counting measure of the eigenvalues $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ of $\mathbf{X}$, i.e., $\mu_{\mathbf{X}} \equiv \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\mathbf{X})}$, where δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

For a real probability measure μ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{3}$$

Stieltjes transform

Proof of Marčenko–Pastur law

Workflow: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ and ST $\frac{1}{p}\operatorname{tr}\mathbf{Q}_{\mathbf{X}}(z) = m_{\mathbf{X}}(z)$ $\Rightarrow$ study the limiting ST $m_{\mathbf{X}}(z) \to m(z) \Rightarrow$ inverse ST to get limiting $\mu_{\mathbf{X}} \to \mu$.

For symmetric $\mathbf{X} \in \mathbb{R}^{p\times p}$, the *empirical spectral distribution* (*ESD*) $\mu_{\mathbf{X}}$ of $\mathbf{X}$ is defined as the normalized counting measure of the eigenvalues $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ of $\mathbf{X}$, i.e., $\mu_{\mathbf{X}} \equiv \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\mathbf{X})}$, where δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

For a real probability measure μ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t-z}. \tag{3}$$

Stieltjes transform

Proof of Marčenko–Pastur law

Workflow: random matrix $\mathbf{X}$ of interest $\Rightarrow$ resolvent $\mathbf{Q}_{\mathbf{X}}(z)$ and ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}_{\mathbf{X}}(z) = m_{\mathbf{X}}(z)$ $\Rightarrow$ study the limiting ST $m_{\mathbf{X}}(z) \to m(z)$ $\Rightarrow$ inverse ST to get limiting $\mu_{\mathbf{X}} \to \mu$.

For symmetric $\mathbf{X} \in \mathbb{R}^{p \times p}$, the *empirical spectral distribution* (*ESD*) $\mu_{\mathbf{X}}$ of $\mathbf{X}$ is defined as the normalized counting measure of the eigenvalues $\lambda_1(\mathbf{X}), \ldots, \lambda_p(\mathbf{X})$ of $\mathbf{X}$, i.e., $\mu_{\mathbf{X}} \equiv \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i(\mathbf{X})}$, where δ_x represents the Dirac measure at x.

Empirical Spectral Distribution (ESD)

For a real probability measure μ with support $\operatorname{supp}(\mu)$, the *Stieltjes transform* $m_\mu(z)$ is defined, for all $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) \equiv \int \frac{\mu(dt)}{t - z}. \tag{3}$$

Stieltjes transform

Heuristic proof of MP law via "leave-one-out" approach

- "guess" $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ such that $\mathbb{E}[\mathbf{Q}] \simeq \bar{\mathbf{Q}}$ and $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) = \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\top\right)\bar{\mathbf{Q}}(z)$$
$$= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)\bar{\mathbf{Q}}(z).$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\top\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\top\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out" approach

- "guess" $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ such that $\mathbb{E}[\mathbf{Q}] \simeq \bar{\mathbf{Q}}$ and $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) = \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z)$$
$$= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out" approach

- "guess" $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ such that $\mathbb{E}[\mathbf{Q}] \simeq \bar{\mathbf{Q}}$ and $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) = \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z)$$
$$= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via “leave-one-out” approach

- “guess” $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ such that $\mathbb{E}[\mathbf{Q}] \simeq \bar{\mathbf{Q}}$ and $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) = \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z)$$
$$= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out" approach

- "guess" $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ such that $\mathbb{E}[\mathbf{Q}] \simeq \bar{\mathbf{Q}}$ and $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.
- for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\begin{aligned}\mathbf{Q}(z) - \bar{\mathbf{Q}}(z) &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)\bar{\mathbf{Q}}(z) \\ &= \mathbf{Q}(z)\left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\right)\bar{\mathbf{Q}}(z).\end{aligned}$$

- for $\bar{\mathbf{Q}}(z) \leftrightarrow \mathbf{Q}(z)$ a DE for $\mathbf{Q}(z)$, look for $\frac{1}{p}\operatorname{tr}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \to 0$,

$$\frac{1}{p}\operatorname{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) - \frac{1}{n}\sum_{i=1}^n \frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i \to 0. \tag{4}$$

- $\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i/p$ as a quadratic form close to a trace form independent of $\mathbf{x}_i$.
- cannot be applied directly as $\mathbf{Q}(z)$ depends on $\mathbf{x}_i$.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ so that $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.

- use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\top \mathbf{Q}_{-i}(z)\mathbf{x}_i}$,
- now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\top - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,
- quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\top \mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \quad (5)$$

- So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.
- self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \quad (6)$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ so that $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ so that $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}}(z)$.

- use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,
- now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,
- quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

- So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.
- self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ so that $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ so that $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}}(z)$.

» use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,

» now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,

» quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

» So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.

» self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ so that $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.

- use Sherman–Morrison to write $\mathbf{Q}(z)\mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i}$,
- now $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\mathsf{T} - z\mathbf{I}_p)^{-1}$ is independent of $\mathbf{x}_i$,
- quadratic form close to the trace:

$$\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^\mathsf{T}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)\mathbf{x}_i}{1+\frac{1}{n}\mathbf{x}_i^\mathsf{T}\mathbf{Q}_{-i}(z)\mathbf{x}_i} \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}_{-i}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}_{-i}(z)}. \tag{5}$$

- So $\frac{1}{p}\operatorname{tr}(\mathbf{F}(z)+z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{Q}(z) \simeq \frac{\frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)\mathbf{Q}(z)}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}$, and "guess" $\mathbf{F}(z) \simeq \left(-z+\frac{1}{1+\frac{1}{n}\operatorname{tr}\mathbf{Q}(z)}\right)\mathbf{I}_p$.
- self-consistent equation of limiting ST $m(z)$ as

$$\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq m(z) = \frac{1}{-z+\frac{1}{1+\frac{p}{n}\frac{1}{p}\operatorname{tr}\mathbf{Q}(z)}} \simeq \frac{1}{-z+\frac{1}{1+\frac{p}{n}m(z)}}. \tag{6}$$

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\bar{\mathbf{Q}}(z)$.

» we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n}\operatorname{tr}\bar{\mathbf{Q}}(z)}\right)\mathbf{I}_p$,

» and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1+cm(z)}\right)^{-1}, \text{ or } zcm^2(z) - (1-c-z)m(z) + 1 = 0.$$

» has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{\imath\theta/2}\}$)

$$m(z) = \frac{1-c-z}{2cz} + \frac{\sqrt{((1+\sqrt{c})^2 - z)((1-\sqrt{c})^2 - z)}}{2cz},$$

only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.

» apply inverse Stieltjes transform we conclude the proof.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}}(z)$.

» we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n} \operatorname{tr} \bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,

» and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1 + cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1 - c - z)m(z) + 1 = 0.$$

» has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{\imath\theta/2}\}$)

$$m(z) = \frac{1 - c - z}{2cz} + \frac{\sqrt{((1 + \sqrt{c})^2 - z)((1 - \sqrt{c})^2 - z)}}{2cz},$$

only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.

» apply inverse Stieltjes transform we conclude the proof.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}}(z)$.

- we have $\mathbf{F}(z) = \left(-z + \frac{1}{1+\frac{1}{n} \operatorname{tr} \bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,
- and $\bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1 + cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1 - c - z)m(z) + 1 = 0.$$

- has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho} e^{\imath\theta/2}\}$)

$$m(z) = \frac{1 - c - z}{2cz} + \frac{\sqrt{((1 + \sqrt{c})^2 - z)((1 - \sqrt{c})^2 - z)}}{2cz},$$

  only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.
- apply inverse Stieltjes transform we conclude the proof.

Heuristic proof of MP law via "leave-one-out"

Objective: "guess" the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some $\mathbf{F}(z)$ $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}}(z)$.

- we have $\mathbf{F}(z) = \left(-z + \frac{1}{1 + \frac{1}{n} \operatorname{tr} \bar{\mathbf{Q}}(z)} \right) \mathbf{I}_p$,
- and $\bar{\mathbf{Q}}(z) = m(z) \mathbf{I}_p$ with $m(z)$ unique Stieltjes transform solution to

$$m(z) = \left(-z + \frac{1}{1 + cm(z)} \right)^{-1}, \text{ or } zcm^2(z) - (1 - c - z)m(z) + 1 = 0.$$

- has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{\imath\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho} e^{\imath\theta/2}\}$)

$$m(z) = \frac{1 - c - z}{2cz} + \frac{\sqrt{((1 + \sqrt{c})^2 - z)((1 - \sqrt{c})^2 - z)}}{2cz},$$

  only one of which is such that $\Im[z]\Im[m(z)] > 0$ by definition of Stieltjes transforms.
- apply inverse Stieltjes transform we conclude the proof.

Some thoughts on the “leave-one-out” proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: high-dimensional concentration (around the expectation), anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\stackrel{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: high-dimensional concentration (around the expectation), anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: high-dimensional concentration (around the expectation), anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the "leave-one-out" proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: high-dimensional concentration (around the expectation), anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z) \simeq \frac{1}{p} \operatorname{tr} \mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Some thoughts on the “leave-one-out” proof

- in essence: propose $\bar{\mathbf{Q}}(z)$ as an approximation of $\mathbb{E}[\mathbf{Q}(z)]$, but simple to evaluate (via a quadratic equation)
- quadratic form close to the trace: high-dimensional concentration (around the expectation), anything more than LLN and concentration
- leave-one-out analysis of large-scale system: $\frac{1}{p}\operatorname{tr}\mathbf{Q}(z) \simeq \frac{1}{p}\operatorname{tr}\mathbf{Q}_{-i}(z)$ for n, p large.
- low complexity analysis of large random system: joint behavior of p eigenvalues $\overset{\text{RMT}}{\to}$ a single deterministic (quadratic) equation
- These are the main intuitions and ingredients for almost everything in RMT and high-dimensional statistics!

Proof of MP law with Gaussian method

Let $x \sim \mathcal{N}(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ a continuously differentiable function having at most polynomial growth and such that $\mathbb{E}[f'(x)] < \infty$. Then,

$$\mathbb{E}[xf(x)] = \mathbb{E}[f'(x)]. \tag{7}$$

In particular, for $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p\times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ a continuously differentiable function with derivatives having at most polynomial growth with respect to p,

$$\mathbb{E}[[\mathbf{x}]_i f(\mathbf{x})] = \sum_{j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j}\right], \tag{8}$$

where $\partial/\partial[\mathbf{x}]_i$ indicates differentiation with respect to the i-th entry of $\mathbf{x}$; or, in vector form $\mathbb{E}[\mathbf{x}f(\mathbf{x})] = \mathbf{C}\mathbb{E}[\nabla f(\mathbf{x})]$, with $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Stein's Lemma

Proof of MP law with Gaussian method

First observe that $\mathbf{Q} = \frac{1}{z}\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q} - \frac{1}{z}\mathbf{I}_p$, so that $\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] - \frac{1}{z}\delta_{ij}$, in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}$. Therefore, from Stein's lemma and the fact that $\partial\mathbf{Q} = -\frac{1}{n}\mathbf{Q}\partial(\mathbf{X}\mathbf{X}^\mathsf{T})\mathbf{Q}$,[a]

$$\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}\left[\frac{\partial[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}}{\partial\mathbf{X}_{ik}}\right] = \mathbb{E}[\mathbf{E}_{ik}^\mathsf{T}\mathbf{Q}]_{kj} - \mathbb{E}\left[\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{Q}(\mathbf{E}_{ik}\mathbf{X}^\mathsf{T} + \mathbf{X}\mathbf{E}_{ik}^\mathsf{T})\mathbf{Q}\right]_{kj}$$

$$= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ki}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}\right] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}\mathbf{X}]_{kk}\mathbf{Q}_{ij}\right]$$

for $\mathbf{E}_{ij}$ the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$, so that, summing over k,

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

[a] This is the matrix version of $d(1/x) = -dx/x^2$.

Proof of MP law with Gaussian method

First observe that $\mathbf{Q} = \frac{1}{z}\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q} - \frac{1}{z}\mathbf{I}_p$, so that $\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] - \frac{1}{z}\delta_{ij}$, in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}$. Therefore, from Stein's lemma and the fact that $\partial\mathbf{Q} = -\frac{1}{n}\mathbf{Q}\partial(\mathbf{X}\mathbf{X}^\mathsf{T})\mathbf{Q}$,[a]

$$\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}\left[\frac{\partial[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}}{\partial\mathbf{X}_{ik}}\right] = \mathbb{E}[\mathbf{E}_{ik}^\mathsf{T}\mathbf{Q}]_{kj} - \mathbb{E}\left[\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{Q}(\mathbf{E}_{ik}\mathbf{X}^\mathsf{T} + \mathbf{X}\mathbf{E}_{ik}^\mathsf{T})\mathbf{Q}\right]_{kj}$$
$$= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ki}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}\right] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}\mathbf{X}]_{kk}\mathbf{Q}_{ij}\right]$$

for $\mathbf{E}_{ij}$ the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$, so that, summing over k,

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

[a] This is the matrix version of $d(1/x) = -dx/x^2$.

Proof of MP law with Gaussian method

First observe that $\mathbf{Q} = \frac{1}{z}\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q} - \frac{1}{z}\mathbf{I}_p$, so that $\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] - \frac{1}{z}\delta_{ij}$, in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}$. Therefore, from Stein's lemma and the fact that $\partial\mathbf{Q} = -\frac{1}{n}\mathbf{Q}\partial(\mathbf{X}\mathbf{X}^\mathsf{T})\mathbf{Q}$,[a]

$$\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \mathbb{E}\left[\frac{\partial[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}}{\partial\mathbf{X}_{ik}}\right] = \mathbb{E}[\mathbf{E}_{ik}^\mathsf{T}\mathbf{Q}]_{kj} - \mathbb{E}\left[\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{Q}(\mathbf{E}_{ik}\mathbf{X}^\mathsf{T} + \mathbf{X}\mathbf{E}_{ik}^\mathsf{T})\mathbf{Q}\right]_{kj}$$
$$= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ki}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}\right] - \mathbb{E}\left[\frac{1}{n}[\mathbf{X}^\mathsf{T}\mathbf{Q}\mathbf{X}]_{kk}\mathbf{Q}_{ij}\right]$$

for $\mathbf{E}_{ij}$ the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$, so that, summing over k,

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^\mathsf{T}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Q}]_{ij}.$$

[a] This is the matrix version of $d(1/x) = -dx/x^2$.

Proof of MP law with Gaussian method

We have

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{ij}.$$

The term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \to \infty$. Also, $\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}) = np + zn\operatorname{tr}\mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z}\mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{k\cdot}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}(p + z\operatorname{tr}\mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and $\mathbf{X}_{k\cdot}$ is the k-th column and row of $\mathbf{X}$, respectively. As the random $\frac{1}{p}\operatorname{tr}\mathbf{Q} \to m(z)$ as $n, p \to \infty$, take it out of the expectation in the limit and

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof.

Proof of MP law with Gaussian method

We have

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{ij}.$$

The term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \to \infty$. Also, $\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}) = np + zn\operatorname{tr}\mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z}\mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{k\cdot}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}(p + z\operatorname{tr}\mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and $\mathbf{X}_{k\cdot}$ is the k-th column and row of $\mathbf{X}$, respectively. As the random $\frac{1}{p}\operatorname{tr}\mathbf{Q} \to m(z)$ as $n, p \to \infty$, take it out of the expectation in the limit and

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof.

Proof of MP law with Gaussian method

We have

$$\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{kj}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}_{ij}\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}})] - \frac{1}{z}\frac{1}{n^2}\mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{ij}.$$

The term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \to \infty$. Also, $\operatorname{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^{\mathsf{T}}) = np + zn\operatorname{tr}\mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z}\mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^{\mathsf{T}}\mathbf{Q}]_{k\cdot}] = \frac{1}{z}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}(p + z\operatorname{tr}\mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and $\mathbf{X}_{k\cdot}$ is the k-th column and row of $\mathbf{X}$, respectively. As the random $\frac{1}{p}\operatorname{tr}\mathbf{Q} \to m(z)$ as $n, p \to \infty$, take it out of the expectation in the limit and

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof.

For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p,

$$\operatorname{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right] = \mathbb{E}\left[(\nabla f(\mathbf{x}))^{\mathsf{T}} \mathbf{C} \nabla f(\mathbf{x}) \right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Nash–Poincaré inequality

- allow to bound the "fluctuation" of random functionals, e.g., the ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z)$, etc.
- to further establish stochastic convergence (in probability or almost surely) as $n, p \to \infty$.

For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p,

$$\mathrm{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right] = \mathbb{E}\left[(\nabla f(\mathbf{x}))^{\mathsf{T}} \mathbf{C} \nabla f(\mathbf{x}) \right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Nash–Poincaré inequality

» allow to bound the "fluctuation" of random functionals, e.g., the ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z)$, etc.

» to further establish stochastic convergence (in probability or almost surely) as $n, p \to \infty$.

For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f : \mathbb{R}^p \to \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p,

$$\operatorname{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^{p} [\mathbf{C}]_{ij} \mathbb{E}\left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right] = \mathbb{E}\left[(\nabla f(\mathbf{x}))^{\mathsf{T}} \mathbf{C} \nabla f(\mathbf{x}) \right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$.

Nash–Poincaré inequality

- allow to bound the "fluctuation" of random functionals, e.g., the ST $\frac{1}{p} \operatorname{tr} \mathbf{Q}(z)$, etc.
- to further establish stochastic convergence (in probability or almost surely) as $n, p \to \infty$.

Extension to non-Gaussian case

For $x \in \mathbb{R}$ a random variable with zero mean and unit variance, $y \sim \mathcal{N}(0,1)$, and f a $(k+2)$-times differentiable function with bounded derivatives,

$$\mathbb{E}[f(x)] - \mathbb{E}[f(y)] = \sum_{\ell=2}^{k} \frac{\kappa_{\ell+1}}{2\ell!} \int_0^1 \mathbb{E}[f^{(\ell+1)} x(t)] t^{(\ell-1)/2} dt + \epsilon_k,$$

where κ_ℓ is the ℓ^{th} cumulant of x, $x(t) = \sqrt{t}x + (1-\sqrt{t})y$, and $|\epsilon_k| \leq C_k \mathbb{E}[|x|^{k+2}] \cdot \sup_t |f^{(k+2)}(t)|$ for some constant C_k only dependent on k.

Interpolation trick

Outline

Wigner semicircle law

Let $\mathbf{X} \in \mathbb{R}^{n \times n}$ be symmetric and such that the $\mathbf{X}_{ij} \in \mathbb{R}$, $j \geq i$, are independent zero mean and unit variance random variables. Then, for $\mathbf{Q}(z) = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$, as $n \to \infty$,

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z), \quad \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_n, \tag{9}$$

with $m(z)$ the unique ST solution to

$$m^2(z) + zm(z) + 1 = 0. \tag{10}$$

The function $m(z)$ is the Stieltjes transform of the probability measure

$$\mu(dx) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}\,dx, \tag{11}$$

known as the *Wigner semicircle law*.

Proof of semicircle law: leave one out heuristic

Let $\mathbf{Q} = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$ be the resolvent, by diagonal entries of matrix inverse lemma,

$$\mathbf{Q}_{ii} = \left(\mathbf{X}_{ii}/\sqrt{n} - z - \mathbf{x}_i^\top \mathbf{Q}_{-i}\mathbf{x}_i/n\right)^{-1},$$

with $[\mathbf{Q}]_{-i} = (\mathbf{X}_{-i}/\sqrt{n} - z\mathbf{I}_{n-1})^{-1}$, $\mathbf{X}_{-i} \in \mathbb{R}^{(n-1)\times(n-1)}$ the matrix obtained by deleting the i-th row and column from $\mathbf{X}$, and $\mathbf{x}_i \in \mathbb{R}^{n-1}$ the i-th column/row of $\mathbf{X}$ with its i-th entry removed. Summing over i,

$$\frac{1}{n}\operatorname{tr}\mathbf{Q} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\frac{1}{\sqrt{n}}\mathbf{X}_{ii} - z - \frac{1}{n}\mathbf{x}_i^\top \mathbf{Q}_{-i}\mathbf{x}_i} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{-z - \frac{1}{n}\mathbf{x}_i^\top \mathbf{Q}_{-i}\mathbf{x}_i} + o(1),$$

since $\frac{1}{\sqrt{n}}\mathbf{X}_{ii}$ vanishes as $n \to \infty$. By quadratic form close to the trace, for large n,

$$(\operatorname{tr}\mathbf{Q}/n)^2 + z\operatorname{tr}\mathbf{Q}/n + 1 \simeq o(1).$$

This is $m^2(z) + zm(z) + 1 = 0$ and thus the conclusion.

Proof of semicircle law: Gaussian method

Similar to the proof of the Marčenko-Pastur law, for $\mathbf{Q} = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$,

$$\frac{1}{\sqrt{n}}\mathbb{E}[\mathbf{X}\mathbf{Q}] = \mathbf{I}_n + z\mathbb{E}[\mathbf{Q}], \tag{12}$$

so that by integration by parts and the fact that $\partial\mathbf{Q} = -\frac{1}{\sqrt{n}}\mathbf{Q}(\partial\mathbf{X})\mathbf{Q}$,

$$\begin{aligned}
\mathbb{E}[\mathbf{Q}_{ij}] &= \frac{1}{z}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}[\mathbf{X}_{ik}\mathbf{Q}_{kj}] - \frac{1}{z}\delta_{ij} = \frac{1}{z}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\left[\frac{\partial\mathbf{Q}_{kj}}{\partial\mathbf{X}_{ik}}\right] - \frac{1}{z}\delta_{ij} \\
&= -\frac{1}{z}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\mathbf{Q}_{ki}\mathbf{Q}_{kj} + \mathbf{Q}_{kk}\mathbf{Q}_{ij}] - \frac{1}{z}\delta_{ij} = -\frac{1}{z}\frac{1}{n}\mathbb{E}\left[[\mathbf{Q}^2]_{ij} + \mathbf{Q}_{ij}\cdot\operatorname{tr}\mathbf{Q}\right] - \frac{1}{z}\delta_{ij}.
\end{aligned}$$

Proof of semicircle law: Gaussian method

So in matrix form

$$\mathbb{E}[\mathbf{Q}] = -\frac{1}{z}\frac{1}{n}\mathbb{E}[\mathbf{Q}^2] - \frac{1}{z}\mathbb{E}[\mathbf{Q}] \cdot \frac{1}{n}\operatorname{tr}\mathbb{E}[\mathbf{Q}] - \frac{1}{z}\mathbf{I}_n + o_{\|\cdot\|}(1), \tag{13}$$

where we used the fact that $\frac{1}{n}\operatorname{tr}\mathbf{Q} - \frac{1}{n}\operatorname{tr}\mathbb{E}\mathbf{Q} \xrightarrow{a.s.} 0$ as $n \to \infty$ and thus be asymptotically "taken out of the expectation" (again high-dimensional concentration).
First RHS matrix has asymptotically vanishing operator norm as $n, p \to \infty$,

$$\mathbb{E}[\mathbf{Q}] = -\frac{1}{z}\left(1 + \frac{1}{z}\frac{1}{n}\operatorname{tr}\mathbb{E}[\mathbf{Q}]\right)^{-1}\mathbf{I}_n + o_{\|\cdot\|}(1)$$

which, after taking the trace and using $\frac{1}{n}\operatorname{tr}\mathbb{E}[\mathbf{Q}(z)] - m(z) \to 0$, gives the limiting formula

$$m^2(z) + zm(z) + 1 = 0.$$

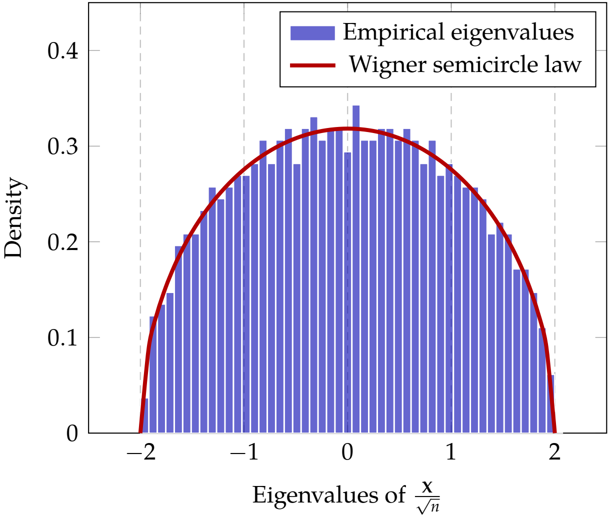

Figure: Histogram of the eigenvalues of $\mathbf{X}/\sqrt{n}$ versus Wigner semicircle law, for standard Gaussian $\mathbf{X}$ and $n = 1\,000$.

Outline

SCM and generalized Marčenko–Pastur law

Let $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p\times n}$ with symmetric nonnegative definite $\mathbf{C} \in \mathbb{R}^{p\times p}$, $\mathbf{Z} \in \mathbb{R}^{p\times n}$ having independent zero mean and unit variance entries. Then, as $n, p \to \infty$ with $p/n \to c \in (0,\infty)$, for $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T} - z\mathbf{I}_p)^{-1}$ and $\tilde{\mathbf{Q}}(z) = (\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X} - z\mathbf{I}_n)^{-1}$,

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = -\frac{1}{z}\left(\mathbf{I}_p + \tilde{m}_p(z)\mathbf{C}\right)^{-1}, \quad \tilde{\mathbf{Q}}(z) \leftrightarrow \bar{\tilde{\mathbf{Q}}}(z) = \tilde{m}_p(z)\mathbf{I}_n,$$

with $\tilde{m}_p(z)$ unique solution to $\tilde{m}_p(z) = \left(-z + \frac{1}{n}\operatorname{tr}\mathbf{C}\left(\mathbf{I}_p + \tilde{m}_p(z)\mathbf{C}\right)^{-1}\right)^{-1}$.
If the empirical spectral measure of $\mathbf{C}$ converges $\mu_{\mathbf{C}} \to \nu$ as $p \to \infty$, then $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}} \to \mu$, $\mu_{\frac{1}{n}\mathbf{X}^\mathsf{T}\mathbf{X}} \to \tilde{\mu}$ where $\mu, \tilde{\mu}$ admitting Stieltjes transforms $m(z)$ and $\tilde{m}(z)$ such that

$$m(z) = \frac{1}{c}\tilde{m}(z) + \frac{1-c}{cz}, \quad \tilde{m}(z) = \left(-z + c\int \frac{t\nu(dt)}{1+\tilde{m}(z)t}\right)^{-1}. \tag{14}$$

A few remarks on the generalized MP law

- different from the explicit MP law, the generalized MP is in general implicit
- we have explicitness in essence due to with $\mathbf{C} = \mathbf{I}_p$, the implicit equation boils down to a quadratic equation that has explicit solution
- if $\mathbf{C}$ has discrete eigenvalues, e.g., $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$, then becomes a (possibly higher-order) polynomial equation, which may admit explicit solution (up to fourth order) using radicals
- the uniqueness of (Stieltjes transform) solution is ensured within a certain region on the complex plane, there may exist solutions $\tilde{m}(z)$ with negative imaginary parts
- **numerical evaluation of $\tilde{m}(z)$**: note that the equation

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \operatorname{tr} \mathbf{C} \left(\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C} \right)^{-1} \right)^{-1} \tag{15}$$

naturally defines a fixed-point equation.

A few remarks on the generalized MP law

- different from the explicit MP law, the generalized MP is in general implicit
- we have explicitness in essence due to with $\mathbf{C} = \mathbf{I}_p$, the implicit equation boils down to a quadratic equation that has explicit solution
- if $\mathbf{C}$ has discrete eigenvalues, e.g., $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$, then becomes a (possibly higher-order) polynomial equation, which may admit explicit solution (up to fourth order) using radicals
- the uniqueness of (Stieltjes transform) solution is ensured within a certain region on the complex plane, there may exist solutions $\tilde{m}(z)$ with negative imaginary parts
- **numerical evaluation of $\tilde{m}(z)$**: note that the equation

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \operatorname{tr} \mathbf{C} \left(\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C} \right)^{-1} \right)^{-1} \tag{15}$$

naturally defines a fixed-point equation.

A few remarks on the generalized MP law

- different from the explicit MP law, the generalized MP is in general implicit
- we have explicitness in essence due to with $\mathbf{C} = \mathbf{I}_p$, the implicit equation boils down to a quadratic equation that has explicit solution
- if $\mathbf{C}$ has discrete eigenvalues, e.g., $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$, then becomes a (possibly higher-order) polynomial equation, which may admit explicit solution (up to fourth order) using radicals
- the uniqueness of (Stieltjes transform) solution is ensured within a certain region on the complex plane, there may exist solutions $\tilde{m}(z)$ with negative imaginary parts
- **numerical evaluation of** $\tilde{m}(z)$: note that the equation

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \operatorname{tr} \mathbf{C} \left(\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C} \right)^{-1} \right)^{-1} \tag{15}$$

naturally defines a fixed-point equation.

A few remarks on the generalized MP law

- different from the explicit MP law, the generalized MP is in general implicit
- we have explicitness in essence due to with $\mathbf{C} = \mathbf{I}_p$, the implicit equation boils down to a quadratic equation that has explicit solution
- if $\mathbf{C}$ has discrete eigenvalues, e.g., $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$, then becomes a (possibly higher-order) polynomial equation, which may admit explicit solution (up to fourth order) using radicals
- the uniqueness of (Stieltjes transform) solution is ensured within a certain region on the complex plane, there may exist solutions $\tilde{m}(z)$ with negative imaginary parts
- **numerical evaluation of $\tilde{m}(z)$**: note that the equation

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \operatorname{tr} \mathbf{C} \left(\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C} \right)^{-1} \right)^{-1} \tag{15}$$

naturally defines a fixed-point equation.

A few remarks on the generalized MP law

- different from the explicit MP law, the generalized MP is in general implicit
- we have explicitness in essence due to with $\mathbf{C} = \mathbf{I}_p$, the implicit equation boils down to a quadratic equation that has explicit solution
- if $\mathbf{C}$ has discrete eigenvalues, e.g., $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$, then becomes a (possibly higher-order) polynomial equation, which may admit explicit solution (up to fourth order) using radicals
- the uniqueness of (Stieltjes transform) solution is ensured within a certain region on the complex plane, there may exist solutions $\tilde{m}(z)$ with negative imaginary parts
- **numerical evaluation of $\tilde{m}(z)$**: note that the equation

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \operatorname{tr} \mathbf{C} \left(\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C} \right)^{-1} \right)^{-1} \tag{15}$$

naturally defines a fixed-point equation.

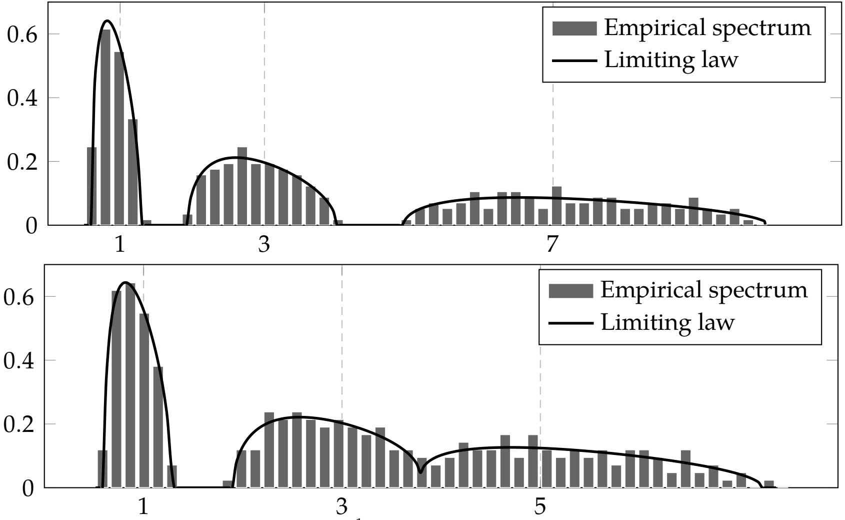

Figure: Histogram of the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{T}}$, $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p\times n}$, $[\mathbf{Z}]_{ij} \sim \mathcal{N}(0,1)$, $n = 3\,000$; for $p = 300$ and $\mathbf{C}$ having spectral measure $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_7)$ (**top**) and $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ (**bottle**).