The Marcenko-Pastur equation

Let $Y = X\Sigma^{1/2}$, where $X \in \mathbb{R}^{n\times p}$ has independent entries, $\mathbb{E}[X_{ij}]=0$, $\mathbb{E}[X_{ij}^2]=1$.

If $\Sigma = I$, the ESD of $\hat{\Sigma} = \frac{1}{n}Y^TY \to$ Marcenko-Pastur law μ_γ when $\frac{p}{n}\to\gamma$.

What happens when $\Sigma \neq I$?

Let $\lambda_1(\Sigma), \dots, \lambda_p(\Sigma)$ be the eigenvalues of Σ, and

$\frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i(\Sigma)}$ the <u>population spectral distribution</u>

Thm (Marcenko-Pastur '67): Suppose $n, p \to \infty$ such that

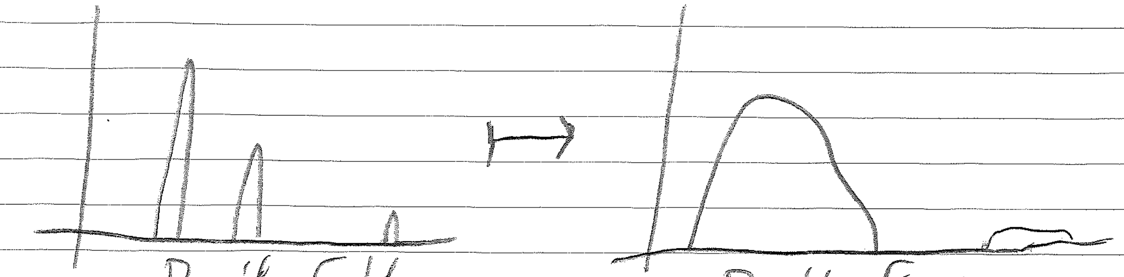
- $\frac{p}{n} \to \gamma \in (0,\infty)$.
- $\frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i(\Sigma)} \to H$ weakly.

Here, H is a probability distribution on $[0,\infty)$. Then the ESD of $\hat{\Sigma}$ converges weakly a.s. to a limit law $\mu_{\gamma,H}$. For each $z \in \mathbb{C}^+$, its Stieltjes transform $m(z) \equiv m_{\gamma,H}(z)$ is the unique solution to

$$m(z) = \int \frac{1}{x(1-\gamma-\gamma z\, m(z)) - z}\, dH(x) \qquad (*)$$

in the domain $\{m \in \mathbb{C} : -\frac{1-\gamma}{z} + \gamma m \in \mathbb{C}^+\}$.

We'll call $(\gamma, H) \mapsto \mu_{\gamma,H}$ the "Marcenko-Pastur mapping".



Density of H

Density of $\mu_{\gamma,H}$.

Note: For $\Sigma \equiv I$, we get $H = \delta_1$. Then $m(z) = (1-\gamma-\gamma z m(z) - z)^{-1}$, which recovers the quadratic equation for the Stieltjes transform of the Marcenko-Pastur law.

Remark: In a n.c.p. space $(\mathcal{A}, \tau)$, let $a \in \mathcal{A}$ have law H and $b \in \mathcal{A}$ have law μ_γ, the Marcenko-Pastur law. I.e.,

$$\tau(a^k) = \int x^k \, dH(x) \text{ and } \tau(b^k) = \int x^k \, d\mu_\gamma(x).$$

Then $\mu_{\gamma,H}$ is the law of $a^{1/2} b a^{1/2}$ when a and b are free in $(\mathcal{A}, \tau)$, where $a^{1/2} \in \mathcal{A}$ satisfies $(a^{1/2})^2 = a$. This is the <u>multiplicative free convolution</u> of H with μ_γ.

Remark: If $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{n \times p}$ are such that $ABv = \lambda v$, so λ is an eigenvalue of AB, then $BA(Bv) = B \cdot \lambda v = \lambda (Bv)$, so λ is also an eigenvalue of BA with eigenvector Bv. The eigenvalues of AB and BA are the same, up to $|p-n|$ zeros.

So $\widehat{\Sigma} = \frac{1}{n} Y^T Y \in \mathbb{R}^{p \times p}$ has the same eigenvalues, up to $|p-n|$ zeros, as the <u>companion matrix</u> $\widetilde{\Sigma} = \frac{1}{n} Y Y^T = \frac{1}{n} X \Sigma X^T \in \mathbb{R}^{n \times n}$. We have

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i(\widetilde{\Sigma}) - z} = \frac{1}{n}\left(\sum_{i=1}^{p} \frac{1}{\lambda_i(\widehat{\Sigma}) - z} + (n-p) \cdot \frac{1}{0-z}\right)$$

$$= \frac{p}{n} \cdot \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i(\widehat{\Sigma}) - z} + \left(1 - \frac{p}{n}\right) \cdot \left(-\frac{1}{z}\right)$$

$$\rightarrow \gamma \cdot m_{\gamma,H}(z) - (1-\gamma) \cdot \frac{1}{z} \equiv \widetilde{m}(z).$$

So the ESD of $\widetilde{\Sigma}$ converges to $\widetilde{\mu}_{\gamma,H}$ with Stieltjes transform $\widetilde{m}(z)$. This is the unique solution in $\mathbb{C}^+$ to

$$\frac{\widetilde{m}(z) + (1-\gamma)\frac{1}{z}}{\gamma} = \int \frac{1}{x\left(1-\gamma-\gamma z\left(\frac{\widetilde{m}(z) + (1-\gamma)\frac{1}{z}}{\gamma}\right)\right) - z} \, dH(x)$$

$$\Leftrightarrow z\widetilde{m}(z) + 1 - \gamma = -\gamma \int \frac{1}{1 + x\widetilde{m}(z)} \, dH(x)$$

$$\Leftrightarrow z = -\frac{1}{\widetilde{m}(z)} + \gamma \int \frac{x}{1 + x\widetilde{m}(z)} \, dH(x) \quad (**).$$

This equation $(**)$ is an equivalent form for the Marcenko-Pastur equation $(*)$, expressed in terms of the companion Stieltjes transform $\widetilde{m}(z)$.

Proof sketch of Thm:

① Suppose first that Σ is diagonal. Define a "linearized resolvent"

$$R(z) = M(z)^{-1} \equiv \begin{pmatrix} -zI & \frac{1}{\sqrt{n}}X \\ \frac{1}{\sqrt{n}}X^T & -\Sigma^{-1} \end{pmatrix}^{-1}. \quad \text{Call this } \begin{pmatrix} R_n(z) & * \\ * & R_p(z) \end{pmatrix}.$$

$$\underbrace{\phantom{\tfrac{1}{\sqrt{n}}X^T}}_{n} \quad \underbrace{\phantom{-\Sigma^{-1}}}_{p}$$

By the Schur-complement identity, the upper-left block $R_n(z)$ is

$$(-zI + \tfrac{1}{n} X \Sigma X^T)^{-1} = (\widetilde{\Sigma} - zI)^{-1},$$ the resolvent of the companion matrix.

Let $M^{(i)}(z)$ be $M(z)$ with row and column i removed, and $R^{(i)}(z) = M^{(i)}(z)^{-1}$.

Applying the Schur complement identity removing row and column i:

- For $i \in \{1, \dots, n\}$, letting x_i^T be row i of X,

$$R_{ii}(z) = \frac{1}{-z - \frac{1}{n}(0, x_i)^T R^{(i)}(z) (0, x_i)}$$

- For $\alpha \in \{1, \dots, p\}$, letting x_α be column α of X and $\Sigma = \operatorname{diag}(\sigma_\alpha)$,

$$R_{n+\alpha, n+\alpha}(z) = \frac{1}{-\sigma_\alpha^{-1} - \frac{1}{n}(x_\alpha, 0)^T R^{(\alpha)}(z) (x_\alpha, 0)}.$$

By concentration of measure, $\frac{1}{n}(0, x_i)^T R^{(i)}(z)(0, x_i) \approx \frac{p}{n} \cdot \frac{1}{p} \operatorname{Tr} R_p^{(i)}(z)$

$$\approx \frac{p}{n} \cdot \frac{1}{p} \operatorname{Tr} R_p(z).$$

Similarly, $\frac{1}{n}(x_\alpha, 0)^T R^{(\alpha)}(z)(x_\alpha, 0) \approx \frac{1}{n}\mathrm{Tr}\, R_n^{(\alpha)}(z) \approx \frac{1}{n}\mathrm{Tr}\, R_n(z)$.

So

$$\tilde{m}_n(z) \equiv \frac{1}{n}\mathrm{Tr}\, R_n(z) \approx \frac{1}{-z - \frac{p}{n}\cdot\frac{1}{p}\mathrm{Tr}\, R_p(z)}$$

$$\approx \frac{1}{-z - \frac{p}{n}\cdot\frac{1}{p}\sum_{\alpha=1}^{p}\frac{1}{-\sigma_\alpha^{-1} - \tilde{m}_n(z)}}.$$

$$\approx \frac{1}{-z - \gamma\cdot\int\frac{1}{-x^{-1} - \tilde{m}_n(z)}\,dH(x)},$$

Rearranging yields

$$-z + \gamma\int\frac{x}{1 + x\tilde{m}_n(z)}\,dH(x) \approx \frac{1}{\tilde{m}_n(z)},$$

that is, $\tilde{m}_n(z)$ approximately satisfies (**). Then $\tilde{m}_n(z) \to \tilde{m}(z)$ for some solution $\tilde{m}(z)$ of (**). Can check $\mathrm{Im}\,\tilde{m}_n(z) \geq c\,\mathrm{Im}\, z$, so in particular (**) has a solution in $\mathbb{C}^+$. To show the solution is unique, suppose $m_1, m_2 \in \mathbb{C}^+$ both satisfy (**). Then

$$0 < \mathrm{Im}\, z = \frac{\mathrm{Im}\, m_1}{|m_1|^2} - \gamma\int\frac{x^2\,\mathrm{Im}\, m_1}{|1 + x m_1|^2}\,dH(x)$$

$$\Rightarrow 1 > \gamma\int\frac{x^2|m_1|^2}{|1 + x m_1|^2}\,dH(x),$$ and similarly for m_2. But

also

$$0 = \left(-\frac{1}{m_1} + \gamma\int\frac{x}{1 + x m_1}\,dH(x)\right) - \left(-\frac{1}{m_2} + \gamma\int\frac{x}{1 + x m_2}\,dH(x)\right)$$

$$= -\left(\frac{m_2 - m_1}{m_1 m_2}\right) + \gamma\int\frac{x^2(m_2 - m_1)}{(1 + x m_1)(1 + x m_2)}\,dH(x)$$

$$\Rightarrow 1 = \gamma\int\frac{x^2 m_1 m_2}{(1 + x m_1)(1 + x m_2)}\,dH(x) \leq \left(\gamma\int\frac{x^2|m_1|^2}{|1 + x m_1|^2}\,dH(x)\right)^{1/2}\left(\gamma\int\frac{x^2|m_2|^2}{|1 + x m_2|^2}\,dH(x)\right)^{1/2}$$

by Cauchy-Schwarz. This is a contradiction, so the solution $\tilde{m}(z)$ in $\mathbb{C}^+$ is unique. Since $\mathrm{Im}\,\tilde{m}_n(z) \geq c\,\mathrm{Im}\, z$ for all n, we must have $\tilde{m}_n(z) \to \tilde{m}(z)$.

② If X is Gaussian, can reduce to ① by diagonalizing Σ, and applying rotational invariance. For non-Gaussian X, apply Lindeberg swapping: Let $\tilde{X}$ differ from X in only the (k,l) entry. Let $\check{X}$ set the (k,l) entry to 0. Let

$$R(z) = \begin{pmatrix} -zI & \frac{1}{\sqrt{n}}X \\ \frac{1}{\sqrt{n}}X^T & -\Sigma^{-1} \end{pmatrix}^{-1},$$ and similarly for $\tilde{R}$, $\check{R}$.

Let Δ_{kl} have 1 in the $(k, l+n)$ and $(l+n, k)$ entries, 0 elsewhere. Then

$$R - \check{R} = -\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)R$$

$$= -\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R} + \check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R}$$

$$- \check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)R,$$

and similarly for $\tilde{R}$. Applying Schur complement,

$$R(z) = \begin{pmatrix} R_n(z) & R_n(z)\left(\frac{1}{\sqrt{n}}X\right)\Sigma \\ \Sigma\left(\frac{1}{\sqrt{n}}X^T\right)R_n(z) & -\Sigma + \Sigma\left(\frac{1}{\sqrt{n}}X^T\right)R_n(z)\left(\frac{1}{\sqrt{n}}X\right)\Sigma \end{pmatrix}.$$

Since $\|R(z)\| \leq C$, $\|\frac{1}{\sqrt{n}}X\| \leq C$, we get $\|R(z)\| \leq C$.

Then, letting $P = \sum_{i=1}^{n} e_i e_i^T$,

$$\frac{1}{n}\sum_{i=1}^{n} e_i^T \check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)\check{R}\left(\frac{1}{\sqrt{n}}X_{kl}\Delta_{kl}\right)Re_i$$

$$\lesssim \frac{1}{n^{5/2}}\left|\mathrm{Tr}\,\Delta_{kl}\check{R}\Delta_{kl}\check{R}\Delta_{kl}(RP\check{R})\right| \leq \frac{C}{n^{5/2}}\|\check{R}\|\cdot\|\check{R}\|\cdot\|\check{R}PR\| \leq \frac{C}{n^{5/2}}.$$

So $\left|\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} R_{ii}(z) - \tilde{R}_{ii}(z)\right]\right| \leq \frac{C}{n^{5/2}}$ by matching of first two moments of X_{kl}. Summing over all (k,l) concludes the proof.

Remark: The convergence $\frac{p}{n} \to \gamma$ and $\frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\Sigma)} \to H$ is a "theoretical construct." Can just as well define $m(z)$ and $\tilde{m}(z)$ as solutions to (*) and (**) with $\frac{p}{n}$ in place of γ, and $\frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\Sigma)}$ in place of H. For instance, (**) then becomes

$$z = -\frac{1}{\tilde{m}(z)} + \frac{p}{n}\cdot\frac{1}{p}\sum_{i=1}^{p} \frac{\lambda_i(\Sigma)}{1+\lambda_i(\Sigma)\tilde{m}(z)}.$$

This equation also has a unique (n,p)-dependent solution $\tilde{m}(z) \in \mathbb{C}^+$, which defines the Stieltjes transform of a deterministic equivalent measure $\tilde{\mu}$. Here $\tilde{\mu}$ depends on (n,p) but has a continuous density (plus a point mass at 0 when $p<n$). It is possible to show:

If $c \le \frac{p}{n} \le C$ and $\|\Sigma\| \le C$ for all n,p as $n,p \to \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(\tilde{\Sigma})} - \tilde{\mu} \to 0 \quad \text{weakly a.s.}$$

Similarly for μ defined by the deterministic equivalent version of (*).

Q: Can we "invert" the Marcenko-Pastur map to estimate the population spectrum of Σ, given sample covariance $\hat{\Sigma}$ or its companion $\tilde{\Sigma}$?

Algorithm of El Karoui '08:

① Take a grid of values $x_1, x_2, \ldots, x_K$. Approximate the population spectrum by a discrete distribution

$$H^{disc} \equiv \sum_{k=1}^{K} w_k \delta_{x_k}.$$

Then the goal is to estimate the weights $w_1, \ldots, w_K$.

② Take a grid of values $z_1, \ldots, z_J \in \mathbb{C}^+$. At each z_j, compute

$$\tilde{m}(z_j) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i(\tilde{\Sigma}) - z_j},$$ where

$\lambda_1(\tilde{\Sigma}), \ldots, \lambda_n(\tilde{\Sigma})$ are the <u>observed</u> sample eigenvalues of $\tilde{\Sigma}$.

③ We expect, for each j,

$$z_j \approx -\frac{1}{\tilde{m}(z_j)} + \frac{p}{n}\int \frac{x}{1+x\tilde{m}(z_j)}\, dH^{disc}(x)$$

$$= -\frac{1}{\tilde{m}(z_j)} + \frac{p}{n}\sum_{k=1}^{K} w_k \cdot \frac{x_k}{1+x_k\tilde{m}(z_j)}$$

Estimate the weights $w_1, \ldots, w_K$ by solving

$$\text{minimize} \quad \max_{j=1}^{J} \left| z_j + \frac{1}{\tilde{m}(z_j)} - \frac{p}{n}\sum_{k=1}^{K} w_k \frac{x_k}{1+x_k\tilde{m}(z_j)} \right|$$

$$\text{subject to} \quad w_1, \ldots, w_K \ge 0, \quad w_1 + \ldots + w_K = 1.$$

This may be written as a linear program in the variables $w_1, \ldots, w_K$.

Thm (El Karoui '08): Suppose $K \equiv K(n)$ and $J \equiv J(n)$ are such that $x_1, \ldots, x_{K(n)}$ form an increasingly fine cover of the support of H, and $z_1, z_2, z_3, \ldots$ have an accumulation point in $\mathbb{C}^+$. Then the estimated law $\hat{H}^{disc} = \sum_{k=1}^{K(n)} \hat{w}_k \delta_{x_k}$ converges weakly a.s. to H, as $n,p \to \infty$ with $\frac{p}{n} \to \gamma$ and $\frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\Sigma)} \to H$.

Remark: In practice, it makes sense to take $z_1, \ldots, z_J$ increasingly close to the real axis, as $n,p \to \infty$.

Eigenvalue shrinkage estimation

Q: Given the data Y (or sample covariance $\hat{\Sigma} = \frac{1}{n} Y^T Y$), can we obtain a better estimate $\tilde{\Sigma}$ of Σ than the "naive" sample covariance $\hat{\Sigma}$? Say, with respect to the squared Frobenius loss

$$\|\tilde{\Sigma} - \Sigma\|_F^2 = \sum_{ij} |\tilde{\Sigma}_{ij} - \Sigma_{ij}|^2.$$

- Without prior information about eigenvectors, we might want an estimator that is rotationally equivariant: Rotating coordinates

$$\Sigma \mapsto O^T \Sigma O \text{ for any orthogonal } O \in \mathbb{R}^{p \times p},$$

we want the estimator $\tilde{\Sigma}$ to satisfy

$$\tilde{\Sigma} \mapsto O^T \tilde{\Sigma} O.$$

(The estimator is invariant to choice of basis for $\mathbb{R}^p$.)

Can show that the optimal estimator $\tilde{\Sigma}$ must have the same eigenvectors as $\hat{\Sigma}$, under this constraint.

- However, the M-P equation tells us that the eigenvalues of $\hat{\Sigma}$ are over-dispersed w.r.t. those of Σ. Can potentially get an improved estimate

$$\tilde{\Sigma} = \sum_{i=1}^p \ell_i \hat{v}_i \hat{v}_i^T$$

where $\hat{\Sigma} = \sum_i \hat{\lambda}_i \hat{v}_i \hat{v}_i^T$ is the spectral decomposition, and $\ell_1, \dots, \ell_p$ "shrink" the sample eigenvalues toward a common center.

Idea goes back to Charles Stein, 1956.

- The optimal estimate is not $\ell_i \approx \lambda_i$, because the eigenvectors $\hat{v}_i$ are inaccurate for v_i. Consider a rank-one case:

$$\|\ell_i \hat{v}_i \hat{v}_i^T - \lambda_i v_i v_i^T\|_F^2 = \ell_i^2 + \lambda_i^2 - 2\ell_i \lambda_i \langle v_i, \hat{v}_i \rangle^2$$

is minimized over ℓ_i at $\ell_i = \lambda_i \langle v_i, \hat{v}_i \rangle^2$, not $\ell_i = \lambda_i$.

$\ell_i(\hat{\lambda}_i)$ should "overshrink" to account for the inaccuracy of $\hat{v}_i$ for v_i, so the solution is not just to apply El Karoui's method to obtain a consistent estimate for the true eigenvalues.

In general, we have

$$\Big\|\sum_{i=1}^p \ell_i \hat{v}_i \hat{v}_i^T - \Sigma\Big\|_F^2 = \Big\|\sum_{i=1}^p \ell_i \hat{v}_i \hat{v}_i^T\Big\|_F^2 + \|\Sigma\|_F^2 - 2\,\mathrm{Tr} \sum_{i=1}^p \ell_i \hat{v}_i \hat{v}_i^T \Sigma$$

$$= \sum_{i=1}^p \ell_i^2 + \|\Sigma\|_F^2 - 2 \sum_{i=1}^p \ell_i \hat{v}_i^T \Sigma \hat{v}_i.$$

This is minimized at $\ell_i^* \equiv \hat{v}_i^T \Sigma \hat{v}_i$. In practice, ℓ_i^* is unknown.

However: Suppose $n, p \to \infty$, $\frac{p}{n} \to \gamma \in (0, \infty)$, $\frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i(\Sigma)} \to H$.

Let $\mu_{\gamma,H}$ be the Marcenko-Pastur forward map, and $m_{\gamma,H}(z)$ its Stieltjes transform. Let $\breve{m}_{\gamma,H}(x) = \lim_{z \in \mathbb{C}^+ \to x} m_{\gamma,H}(z)$, and

$$\ell_i^{or} \equiv \frac{\hat{\lambda}_i}{|1 - \gamma - \gamma \hat{\lambda}_i \breve{m}_{\gamma,H}(\hat{\lambda}_i)|^2}.$$

Thm (Ledoit, Péché 2011): For each fixed $x \in \mathbb{R}$, almost surely

$$\frac{1}{p} \sum_{i=1}^p \ell_i^* \mathbb{I}\{\hat{\lambda}_i \le x\} - \frac{1}{p} \sum_{i=1}^p \ell_i^{or} \mathbb{I}\{\hat{\lambda}_i \le x\} \to 0.$$

Proof sketch: Write

$$\frac{1}{p} \mathrm{Tr} (\hat{\Sigma} - zI)^{-1} \Sigma = \frac{1}{p} \mathrm{Tr} \sum_{i=1}^p \frac{\hat{v}_i \hat{v}_i^T}{\hat{\lambda}_i - z} \Sigma = \frac{1}{p} \sum_{i=1}^p \ell_i^* \frac{1}{\hat{\lambda}_i - z}.$$

Then $\frac{1}{\pi}\operatorname{Im}\frac{1}{p}\operatorname{Tr}(\hat{\Sigma}-(t+i\eta))^{-1}\Sigma = \frac{1}{p}\sum_{i=1}^{p} e_i^{*} \frac{\eta}{\pi[(\hat{\lambda}_i-t)^2+\eta^2]}$

Note that $\int_{-\infty}^{x}\frac{\eta}{\pi[(\hat{\lambda}_i-t)^2+\eta^2]}dt \xrightarrow{\eta\to 0} \begin{cases}0 & \text{if } \hat{\lambda}_i > x\\ 1 & \text{if } \hat{\lambda}_i < x\end{cases}$

So $\frac{1}{p}\sum_{i=1}^{p} e_i^{*}\,\mathbb{1}\{\hat{\lambda}_i\le x\} = \lim_{\eta\to 0}\frac{1}{\pi}\operatorname{Im}\int_{-\infty}^{x}\frac{1}{p}\operatorname{Tr}(\hat{\Sigma}-(t+i\eta))^{-1}\Sigma\, dt.$

Consider $R(z) = \begin{pmatrix}-zI & \frac{1}{\sqrt{n}}X\\ \frac{1}{\sqrt{n}}X^{T} & -\Sigma^{-1}\end{pmatrix}^{-1} \equiv \begin{pmatrix}R_n(z) & *\\ * & R_p(z)\end{pmatrix}.$

By Schur complement, $R_n(z) = (-zI+\frac{1}{n}X\Sigma X^{T})^{-1} = (\tilde{\Sigma}-zI)^{-1}$

So $\frac{1}{n}\operatorname{Tr}R_n(z)\approx\tilde{m}(z)$. But also, for $i\in\{1,\ldots,n\}$,

$$R_{ii}(z) = \frac{1}{-z-\frac{1}{n}x_i^{T}R_p^{(i)}(z)x_i} \approx \frac{1}{-z-\frac{1}{n}\operatorname{Tr}R_p(z)}$$

Here $R_p(z) = (-\Sigma^{-1}+\frac{1}{nz}X^{T}X)^{-1} = \left[\frac{1}{z}\Sigma^{-1/2}(-zI+\Sigma^{1/2}X^{T}X\Sigma^{1/2})\Sigma^{-1/2}\right]^{-1}$

$= z\Sigma^{1/2}(\hat{\Sigma}-z)^{-1}\Sigma^{1/2}.$

$$\tilde{m}(z)\approx\frac{1}{-z-z\cdot\gamma\cdot\frac{1}{p}\operatorname{Tr}(\hat{\Sigma}-z)^{-1}\Sigma}.$$

$$\Rightarrow \frac{1}{p}\operatorname{Tr}(\hat{\Sigma}-z)^{-1}\Sigma\approx-\frac{1}{z\gamma}\left(\frac{1}{\tilde{m}(z)}+z\right)$$

$$= -\frac{1}{z\gamma}\left(\frac{1}{\gamma m_{\gamma,H}(z)-\frac{1-\gamma}{z}}+z\right)$$

$$= \frac{1}{\gamma}\left(\frac{1}{1-\gamma-z\gamma m_{\gamma,H}(z)}-1\right).$$

Combining the above shows

$$\frac{1}{p}\sum_{i=1}^{p}e_i^{*}\,\mathbb{1}\{\hat{\lambda}_i\le x\}\to\lim_{\eta\to 0}\frac{1}{\pi}\operatorname{Im}\int_{-\infty}^{x}\frac{1}{\gamma}\left(\frac{1}{1-\gamma-(t+i\eta)\gamma m_{\gamma,H}(t+i\eta)}-1\right)dt$$

$$=\int_{-\infty}^{x}\lim_{\eta\to 0}\frac{\operatorname{Im}[(t+i\eta)m_{\gamma,H}(t+i\eta)]}{\pi|1-\gamma-(t+i\eta)\gamma m_{\gamma,H}(t+i\eta)|^2}dt$$

$$=\int_{-\infty}^{x}\frac{t\cdot\frac{1}{\pi}\operatorname{Im}\check{m}_{\gamma,H}(t)}{|1-\gamma-\gamma t\check{m}_{\gamma,H}(t)|^2}dt.$$

Using that $\frac{1}{\pi}\operatorname{Im}\check{m}_{\gamma,H}(t)\,dt$ is the density of the limit ESD of $\hat{\Sigma}$, we get also $\frac{1}{p}\sum_{i=1}^{p}e_i^{or}\,\mathbb{1}\{\hat{\lambda}_i\le x\}\to\int_{-\infty}^{x}\frac{t}{|1-\gamma-\gamma t\check{m}_{\gamma,H}(t)|^2}\cdot\frac{1}{\pi}\operatorname{Im}\check{m}_{\gamma,H}(t)\,dt.$

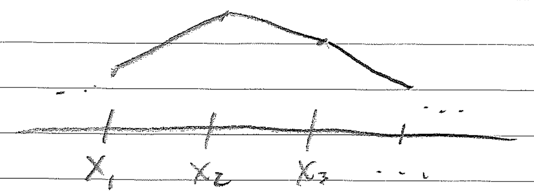
Thus, Ledoit-Wolf 2012 proposes to estimate e_i^{*} by an estimate of e_i^{or}:

① Take a grid $x_1,\ldots,x_K$, and approximate

$$H\approx H^{disc}\equiv\sum_{k=1}^{K}w_k\delta_{x_k}.$$

② For a given approximation H^{disc}:

- Compute $\check{m}_{\gamma,H^{disc}}(x_1),\ldots,\check{m}_{\gamma,H^{disc}}(x_K)$ by solving the Marcenko-Pastur equation.
- Interpolate the density of $\mu_{\gamma,H^{disc}}$ from $\frac{1}{\pi}\operatorname{Im}\check{m}_{\gamma,H^{disc}}(x_1),\ldots,\frac{1}{\pi}\operatorname{Im}\check{m}_{\gamma,H^{disc}}(x_K)$ using the trapezoid rule:



- Use this to compute

$$\sup_{k=1}^{K}\left|\hat{F}(x_k)-F_{\gamma,H^{disc}}(x_k)\right| \qquad (*)$$

where $\hat{F}$ is the empirical CDF of eigenvalues of $\hat{\Sigma}$, and $F_{\gamma,H^{disc}}$ is the interpolated CDF.

③ Optimize $w_1,\ldots,w_K$ by minimizing (*), using a non-convex optimizer. Then estimate e_i^{or} using $\hat{\lambda}_i$ and $\check{m}_{\gamma,H^{disc}}(\hat{\lambda}_i)$.

Problem 8. (a) Implement the spectrum estimation method of El Karoui, taking in the grids $x_1, \dots, x_K$ and $z_1, \dots, z_J$ as "tuning parameters."

(b) Test this algorithm on the examples

- $\Sigma = I$
- $\Sigma = \begin{pmatrix} I & 0 \\ 0 & 2I \end{pmatrix}$
- $\Sigma_{ij} = 0.3^{|i-j|}$
- $\Sigma = \mathrm{diag}\left(\frac{1}{p}, \frac{2}{p}, \frac{3}{p}, \dots, 1\right)$.

How sensitive is the algorithm to the choice of $z_1, \dots, z_J$?

(c) For various problem sizes (n, p), try taking $z_1, \dots, z_J$ to be a grid of values above the real line, with imaginary part $\eta = n^{-\tau}$ for various settings of $\tau \in (0,1)$. Which setting of τ seems to work best? Can you provide some heuristic explanation for this?