Random Matrix Theory: Sampled Covariance Matrix

$X_i \in \mathbb{R}^p \sim N(\mu, \sigma^2 I_p)$

$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\hat{\Sigma}_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)(X_i - \hat{\mu}_n)^T$

M.L.E.
p fixed
$n \to \infty$

$p \nearrow \quad n \nearrow \qquad \frac{p}{n} \to \gamma \neq 0$ 高维
$\to \infty$

Stein's Phenomenon $\quad \hat{\mu}_n^{MLE}$ 不是最优的. inadmissible

$R(\hat{\mu}_n, \mu) \le R(\hat{\mu}_n^{MLE}, \mu) \quad \forall \mu \in \mathbb{R}^p$

$\quad = \quad < \quad " \qquad \exists \mu$

$\hat{\mu}_n^{JS}, \quad \hat{\mu}_n^{ST}.$

Sampled Covariance Matrix $\hat{\Sigma}_n$? $\quad p$ 大

$\lim_{n\to\infty} \frac{p}{n} = \gamma \neq 0.$

Johnstone '2006 $\quad$ RMT & Statistics

$X_i \sim N(0, I_p) \qquad X^{p\times n} = [X_1 \cdots X_n]$

$\hat{\Sigma}_n = \frac{1}{n} X X^T \xrightarrow{\lim_{n\to\infty} \frac{p}{n} = 0} I_p$

Wishart Matrix

$\lambda_i(\hat{\Sigma}_n)$ eigenvalue $\quad$ Marcenko-Pastur Dist.

$$\sim \quad \mu_0^{MP}(t) = \begin{cases} 0, & t \notin [a,b] \\ \frac{\sqrt{(b-t)(t-a)}}{\sqrt{2\pi\gamma t}}\, dt, & t \in [a,b] \end{cases}, \quad \gamma \le 1 \quad \lim_{n\to\infty}\frac{p}{n} \le 1$$

$a = (1-\sqrt{\gamma})^2 \quad \xrightarrow{\gamma \to 0} \quad a = b = 1$
$b = (1+\sqrt{\gamma})^2$

$\mu^{MP}(t) = 1(\gamma > 1)\, \delta_0(t) + \mu_0^{MP}(t), \quad \int \delta_0(t)\, dt\; 1 - \frac{1}{\gamma}$

PCA: $\hat{\Sigma}_n = \frac{1}{n}\tilde{X}\tilde{X}^T \longrightarrow U\Lambda U^T \quad \Lambda = \mathrm{diag}(\lambda_i)$

$\lambda_1 \geq \lambda_2 \cdots \lambda_k \cdots$

gap $\lambda_i - \lambda_{i+1}$? heuristic

Phase Transition in PCA

$$Y = X + \varepsilon = \alpha u + \varepsilon \qquad \|u\|_2 = 1$$

$u \in \mathbb{R}^p$ fixed given. $\alpha \sim N(0, \sigma_X^2)$ signal

noise $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_p)$

SNR: $R = \frac{\sigma_X^2}{\sigma_\varepsilon^2}$

rank-1 signal + noise $Y \sim N(0, \Sigma)$

$$\Sigma = \underset{\substack{\uparrow \\ \text{rank-1 matrix}}}{\sigma_X^2 uu^T} + \underset{\substack{\uparrow \\ \text{sparse matrix}}}{\sigma_\varepsilon^2 I_p}, \quad \frac{\sigma_X^2}{\sigma_\varepsilon^2} = R$$

$$\begin{bmatrix} * & & 0 \\ & \ddots & \\ 0 & & * \end{bmatrix}$$

$\hat{\Sigma}_n = \frac{1}{n}\tilde{Y}\tilde{Y}^T \xrightarrow{PCA}$ primary $\begin{cases} \text{eigenvalue} & \hat{\lambda}_{max}(\hat{\Sigma}_n) \text{ vs } \sigma_X^2 \\ \text{eigenvector} & \hat{v}_{max}(\cdot, \cdot): u \end{cases}$ uu^T

$\sigma_\varepsilon^2 = 1$ MP dist

$$\hat{\lambda}_{max} \to \begin{cases} b = (1+\sqrt{\gamma})^2 & R \leq \sqrt{\gamma} \quad \text{随机谱上界} \\ (1+\sigma_X^2)\left(1+\frac{\gamma}{\sigma_X^2}\right) \neq \sigma_X^2, & R > \sqrt{\gamma} \quad \text{信号谱 biased} \end{cases}$$

$$|\langle \hat{v}_{max}, u\rangle|^2 \to \begin{cases} 0, & R \leq \sqrt{\gamma} \quad \text{PCA 随机} \\ \dfrac{1-\frac{\gamma}{\sigma_X^4}}{1+\frac{\gamma}{\sigma_X^2}} \neq 1 & R > \sqrt{\gamma}, \text{ 锥域 cone (biased)} \end{cases}$$

PT in PCA: SNR 小 是随机谱 Johnstone 2006

Dual ↕ SNR 大 biased 信号谱 Est.

MDS → Kernel (El Karoui, Cheng-Singer)

$Y_i \sim N(0, \Sigma) \qquad i=1,\dots,n$

$\Sigma = \sigma_x^2 u u^T + \sigma_\varepsilon^2 I_p, \qquad \sigma_\varepsilon^2 = 1 \qquad$ SNR $R = \frac{\sigma_x^2}{\sigma_\varepsilon^2} = \sigma_x^2$

$\hat{\Sigma}_n = \frac{1}{n} Y Y^T \qquad Y = [Y_1, \dots, Y_n]_{p\times n}$

Phase Transition

1) $\hat{\lambda}_{max}(\hat{\Sigma}_n) \to \begin{cases} b & R \le \sqrt{\gamma} \\ f(\sigma_x^2, \gamma), & R \ge \sqrt{\gamma} \end{cases}$

$= (1+\sigma_x^2)(1+\frac{\gamma}{\sigma_x^2}) = (1+R)(1+\frac{\gamma}{R})\sigma_\varepsilon^2$

2) $|\langle \hat{v}_{max}, u\rangle| \to \begin{cases} 0 & R \le \sqrt{\gamma} \\ \left(\dfrac{1-\frac{\gamma}{R^2}}{1+\gamma+\frac{2\gamma}{R}}\right)^{\frac{1}{2}} < 1, & R \ge \sqrt{\gamma} \end{cases}$

$\mathrm{eig}(\hat{\Sigma}_n) = \frac{1}{n} Y Y^T \qquad Y_i = \Sigma^{\frac{1}{2}} Z_i \qquad Z_i \sim N(0, I_p)$

$= \frac{1}{n} \Sigma^{\frac{1}{2}} Z Z^T \Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} S_n \Sigma^{\frac{1}{2}}, \qquad S_n = \frac{1}{n} Z Z^T \sim$ MP dist, Wishart

$\sim \Sigma^{\frac{1}{2}} (\Sigma^{\frac{1}{2}} S_n \Sigma^{\frac{1}{2}}) \Sigma^{-\frac{1}{2}} = (S_n \Sigma)$ eigval

(λ, v) $S_n\Sigma$ eigenpair

$(*)$ $S_n \Sigma v = \lambda v \qquad \|v\|_2 = 1$

$\hat{v} = c(\Sigma^{\frac{1}{2}} v) \qquad \|\hat{v}\|_2 = 1 \qquad \hat{\Sigma}_n = \Sigma^{\frac{1}{2}} S_n \Sigma^{\frac{1}{2}}$

$\hat{\Sigma}_n \hat{v} = \lambda \hat{v}$

$1 = \hat{v}^T \hat{v} = c^2 (\Sigma^{\frac{1}{2}} v)^T (\Sigma^{\frac{1}{2}} v) = c^2 v^T \Sigma v = c^2 v^T (\sigma_x^2 u u^T + I_p) v$

$= c^2 \sigma_x^2 (u^T v)^2 + c^2 \Rightarrow c^2 = \dfrac{1}{\sigma_x^2 (u^T v)^2 + 1}$ σ_ε^2

$\Sigma = \sigma_x^2 u u^T + I_p$

$S_n(\sigma_x^2 u u^T + \sigma_\varepsilon^2 I_p) v = \lambda v$

$(\lambda I_p - \sigma_\varepsilon^2 I_p) v = \sigma_x^2 S_n u (u^T v)$

$$(**) \quad \underline{v} = (\lambda I_p - \sigma_\varepsilon^2 S_n)^{-1} \cdot \sigma_x^2 S_n u \underline{(u^T v)}$$

$$\underline{(u^T v)} = u^T (\quad Z \quad)^{-1} \cdot \sigma_x^2 S_n u \underline{(u^T v)}$$

Assume $u^T v \neq 0$

$$1 = \sigma_x^2 \; u^T (\lambda I_p - \sigma_\varepsilon^2 \underline{\underline{S_n}})^{-1} \underline{\underline{S_n}} u$$

Wishart

$$S_n = W^T \Lambda W^T \qquad \Lambda = \mathrm{diag}(\lambda_i)_{i=1,\ldots,p} \sim \text{MP law}$$

$$\alpha_i = W_i^T u \qquad W^T W = W \cdot W^T = I_p.$$

$$\underbrace{1 = \sigma_x^2 \sum_{i=1}^{p} \frac{\lambda_i}{\lambda - \sigma_\varepsilon^2 \lambda_i} \alpha_i^2} \qquad \lambda_i \sim MP$$

Assume rank-1 u. uniformly on S^{p-1}, $\mathbb{E}(\alpha_i^2) \sim \frac{1}{p}$

$$\sum_{i=1}^{p} \alpha_i^2 = 1$$

Monte Carlo Sum.

$$\sim \sigma_x^2 \frac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i}{\lambda - \sigma_\varepsilon^2 \lambda_i} \sim \sigma_x^2 \int \frac{t}{\lambda - \sigma_\varepsilon^2 t} d\mu^{MP}(t)$$

$$\sigma_\varepsilon^2 = 1$$

Stieltjes Transform (Terry Tao 2011, RMT 网络)

$$s(z) = \int_{-\infty}^{+\infty} \frac{1}{t - z} d\mu^{MP}(t), \quad z \in \mathbb{C}$$

Hilbert transform

Lemma

(A) $$\int \frac{t}{\lambda - t} d\mu^{MP}(t) = -\lambda s(\lambda) - 1 \qquad \lambda \in \mathbb{R}^+$$

(B) $$\int \frac{t^2}{(\lambda - t)^2} d\mu^{MP}(t) = \lambda^2 s'(\lambda) + 2\lambda s(\lambda) + 1$$

Proof: (A) $$T(\lambda) = \int \frac{t}{\lambda - t} d\mu^{MP}(t)$$

$$1 + T(\lambda) = \int \frac{\lambda - \cancel{t} + \cancel{t}}{\lambda - t} d\mu^{MP}(t) = \lambda \underbrace{\int \frac{1}{\lambda - t} d\mu^{MP}(t)}_{= -\lambda s(\lambda)}$$

$$T(\lambda) = -\lambda s(\lambda) - 1 \qquad \boxed{\text{证}}$$

(B) $$\int \frac{t^2}{(\lambda - t)^2} d\mu^{MP}(t) = -T(\lambda) - \lambda T'(\lambda) \qquad \boxed{\text{证}}$$

S-transform

$$S(z) = \frac{(1-\gamma) - z + \sqrt{(z - 1 + \gamma)^2 - 4\gamma z}}{2\gamma z}, \qquad z \in \mathbb{C},\ z \neq 0$$

Note: $$S^{-1}(z) + z = \frac{1}{1 + \gamma S(z)} \Rightarrow \text{poly}(S, z) = 0$$ largest root

Lemma (A) $\qquad \lambda \geq b$

$$(**) \Rightarrow 1 = \frac{\sigma_x^2}{4\gamma}\left[2\lambda - (a+b) - 2\sqrt{(\lambda - a)(\lambda - b)}\right]$$

$$1 = \frac{\sigma_x^2}{4\gamma}\left[2\lambda - (a+b) - 2\sqrt{(\lambda-a)(\lambda-b)}\right]$$

$\lambda \geq b$ holds.

$\lambda = b$ noise spectrum upper bnd.

$$1 = \frac{\sigma_x^2}{4\gamma}[b-a] = \frac{\sigma_x^2}{\sqrt{\gamma}}$$

$a = (1-\sqrt{\gamma})^2$

$b = (1+\sqrt{\gamma})^2$

$b - a = 4\sqrt{\gamma}$

$$\sigma_x^2 = \sqrt{\gamma} \quad (\text{SNR}.\ R = \sqrt{\gamma})$$

Phase - Transition boundary

$$\sigma_x^2 \geq \sqrt{\gamma}. \qquad \lambda = (1+\sigma_x^2)\left(1+\frac{\gamma}{\sigma_x^2}\right) = f(\sigma_x^2; \gamma)$$

$$\left(= (1+R)\left(1+\frac{\gamma}{R}\right)\sigma_\varepsilon^2\right)$$

biased of signal strength

(**) 左右同时 $V^T \cdot (**)$

$$1 = V^T V = \| (u^T v) \sigma_X^2 (\lambda I_p - \sigma_\varepsilon^2 S_n)^{-1} S_n u \|_2^2$$

$$= (u^T v)^2 \sigma_X^4 \; u^T S_n (\lambda I_p - \sigma_\varepsilon^2 S_n)^{-2} S_n u.$$

$$(u^T v)^{-2} = \sigma_X^4 \; \underline{u^T S_n (\lambda I_p - \sigma_\varepsilon^2 S_n)^{-2} S_n u}$$

$$S_n = W \Lambda W^T \qquad \alpha_i = W_i^T u$$

$$\underline{\qquad} = \sigma_X^4 \sum_{i=1}^{p} \frac{\lambda_i^2}{(\lambda - \sigma_\varepsilon^2 \lambda_i)^2} \alpha_i^2 \leftarrow \frac{1}{p}$$

$$\sim \underline{\sigma_X^4 \int \frac{t^2}{(\lambda - \sigma_\varepsilon^2 t)^2} d\mu^{MP}(t)}$$

Lemma (B)

$\sigma_\varepsilon^2 = 1$ $\Rightarrow$
$$= \frac{\sigma_X^4}{\gamma\gamma} \left[-4\lambda + (a+b) + 2\sqrt{(\lambda-a)(\lambda-b)} + \frac{\lambda(2\lambda - (a+b))}{\sqrt{(\lambda-a)(\lambda-b)}} \right]$$

$$u^T v^2 = \frac{1 - \frac{\gamma}{R^2}}{1 + \gamma + \frac{2\gamma}{R}}$$

$$\Rightarrow \quad u^T \hat{v}^2 = u^T (c \Sigma^{\frac{1}{2}} v) = \frac{1+R}{1+R(u^T v)^2} (u^T v)^2$$

$$= g(R, \gamma) < 1$$

$$\underline{\qquad\qquad}$$

$$\frac{1 - \frac{\gamma}{\sigma_X^4}}{1 + \frac{\gamma}{\sigma_X^2}}, \qquad R > \sqrt{\gamma}$$

RMT

MP law for Wishart

$$\Sigma = \sigma_x^2 u u^T + \sigma_\varepsilon^2 I_p$$

$$\hat{\lambda}_{max}(\hat{\Sigma}_n) \neq \sigma_x^2$$

$$= \begin{cases} f(\sigma_x^2), & R \geq \sqrt{\gamma} \\ b, & R \leq \sqrt{\gamma} \end{cases}$$

PC, $\hat{v}_{max}$ in conic neighbor of u

$g(R, \gamma)$

Open. $\sigma_\varepsilon^2 I_p$ homogeous.

$diag(\sigma_{\varepsilon_i}^2)$ hetergenous. ? Owen Wang

Rank-Sparsity: Robust PCA

SDP uu^T + sparse

条件? Compressed Sensing

Candes, Ma, Wright, Recht, Joel Tropp: