Ledoit 2001 well conditioned. $S^{*} = \alpha I + \beta S$.

Background: $\frac{p}{n} < 1$ but not negligible.

in this case: S is not well-conditioned.

inverting S will introduce numerical bias.

Solution: $\Sigma^{*} = \alpha_1 I + \beta_2 S$. ($\alpha_1 \to \mu$)

① finite n. fixed p.

modified Frobenius norm. $\langle A_1, A_2 \rangle = \frac{1}{p} tr(A_1 A_2^T)$. $\|I\| = 1$.

4 important scalars: $\mu = \langle \Sigma, I \rangle$ $\alpha^2 = \|\Sigma - \mu I\|^2$ $\beta^2 = E\|S - \Sigma\|^2$ $\delta^2 = E\|S - \mu I\|^2$.

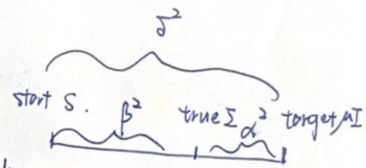
intuition: shrink S to I.

μ: actual shrinkage target ($\to \mu I$)

α^2: difference between true value & target.

β^2: difference between shrinkage start point & true value.

δ^2: difference between shrinkage start & target.



we want to end up somewhere in between. (Σ^{*}). s.t. $\min \|\Sigma^{*} - \Sigma\|^2$

$\alpha^2 + \beta^2 = \delta^2$

shrinkage intensity: do we need shrinkage?

It yields solution $\Sigma^{*} = \frac{\beta^2}{\delta^2}\mu I + \frac{\alpha^2}{\delta^2} S$. $\Rightarrow E\|\Sigma^{*} - \Sigma\|^2 = \frac{\alpha^2 \beta^2}{\delta^2}$ risk.

an interesting insight from eigenvalues:

Σ eigens: $\lambda_1 \cdots \lambda_p$.

S eigens: $l_1 \cdots l_p$. $\Rightarrow \mu = \frac{1}{p}\sum_{i=1}^{p} \lambda_i = \frac{1}{p}\sum_{i=1}^{p} l_i$

re-interpret $\delta^2 = \alpha^2 + \beta^2 \Rightarrow \frac{1}{p} E\left[\sum_{i=1}^{p} (l_i - \mu)^2\right] = \frac{1}{p}\sum_{i=1}^{p} (\lambda_i - \mu)^2 + E[\|S - \Sigma\|^2]$ [$\{l_i\}$ have larger dispersion than true]

Σ^{*} eigens: $\lambda_1^{*} \cdots \lambda_p^{*}$. $\lambda_i^{*} = \frac{\beta^2}{\delta^2}\mu + \frac{\alpha^2}{\delta^2} l_i$

shrink. $\leftarrow$ larger bias.

note: it has smaller dispersion than true value.

why are eigenvalues distorted in sample covariance matrix

thm: the eigenvalues are the most dispersed diagonal elements that can be obtained.

① $\mu = \frac{1}{p} tr(R)$ mean of diagonal elements don't change with: V orthogonal. $\mu = \frac{1}{p} tr(V^T R V)$

② dispersion: $\frac{1}{p}\sum_{i=1}^{p} (V_i^T R V_i - \mu_r)^2$.

① dispersion maximized when: $R = VDV^T$ (eigen decomp) $\Rightarrow D = V^T R V$ diagonal.

true: $\Sigma = T \Lambda T^T \Rightarrow \Lambda = T^T \Sigma T$ true eigen.

sample: $S = GLG^T \Rightarrow L = G^T S G$ sample eigen.

compare dispersions of Λ & L.

$\Lambda = T^T \Sigma T \succ G^T \Sigma G$

$\approx$

$T^T S T$ (unbiased) $\prec G^T S G = L$

② general assumption: (Kolmogorov asymptotics) $\frac{p}{n} > 0$ constant but $\overset{p}{\checkmark}$ growing.

goal: to find a bona fide estimator (consistent)

answer to Q: when is shrinkage important.

- 3 assumptions. [we will work with Y ~~ist~~ instead of X: true Σ_n for X_n, $\Sigma_n = T_n \Lambda_n T_n^T$

$\Rightarrow Y_n = T_n^T X_n \quad S_n = \frac{X_n X_n^T}{n} \Rightarrow S^T_n = \frac{Y_n Y_n^T}{n} = \frac{T_n^T X_n X_n^T T_n}{n} \Rightarrow$ unbiased for Λ_n]

(1). $\frac{p_n}{n}$ bdd : $\exists K_1,\ \forall n,$ s.t. $\frac{p_n}{n} \leq K_1$

(2). Average 8th moment of Y bdd. $\exists K_2$ s.t. $\frac{1}{p_n}\sum_{i=1}^{p_n} E(y_{i1}^n)^8 \leq K_2$.

(3) product of self uncorrelty $y_i y_j$ are average asymptotically uncorrelated.

$$\lim_{n\to\infty} \frac{p_n^2}{n^2} \frac{\sum_{i,j,k,l}\left(Cov[y_{i1}^n y_{j1}^n,\ y_{k1}^n y_{l1}^n]\right)^2}{\# \text{ All combinations}} = 0$$

- define a norm that allow identity to be benchmark:

$\|A\|_n^2 = f(p_n)\,tr(AA^T) \qquad f(p_n) = \frac{1}{p_n}$.

- terms going to work with: $\mu_n = \langle \Sigma_n, I_n\rangle_n = \frac{1}{p_n} tr(\Sigma_n)$. $\alpha_n^2 = \|\Sigma_n - \mu_n I_n\|^2$ $\beta_n^2 = E[\|S_n - \Sigma_n\|_n^2]$

$\delta_n^2 = E[\|S_n - \mu_n I_n\|^2]$

Define $\theta_n^2 = Var[\frac{1}{p_n}\sum_{i=1}^{p_n}(y_{i1}^n)^2]$

And all of them are bdd.

→ 这到底是个什么: $Y_1 = Y_n$ 第1列

$= T_n^T X_1$, X_1 第1列.

$\theta_n^2 = Var[\frac{1}{p_n} Y_1^T Y_1] = Var[\frac{1}{p_n} tr\, Y_1 Y_1^T]$

$= Var[\frac{1}{p_n} tr\ T_n^T X_1 X_1^T T_n]$.

$= Var[\frac{1}{p_n} X_1^T X_1]$ go to 0 if the squared terms are uncorrelated.

unfortunately, can't be guaranteed by Assumption 3.

- Major Thm that sheds light on understanding:

$$\lim_{n\to\infty} \underbrace{E[\|S_n - \Sigma_n\|_n^2]}_{\beta_n^2} - \frac{p_n}{n}(\mu_n^2 + \theta_n^2) = 0.$$

(1). S_n is only consistent when $\overbrace{\boxed{\frac{p_n}{n}}\boxed{(\mu_n^2 + \theta_n^2)}}^{2\text{ cases.}} \to 0$

second ~~term~~ case doesn't usually hold, $\mu_n^2 \to 0$ require: nondegerate r.v.s negligible w.r.t n

(2). if we treat θ_n^2 as ~~negne~~ negligible. $\lim_{n\to\infty}(\beta_n^2 - \frac{p_n}{n}\mu_n^2) = 0$.

By fix p case, shrinkage matters when $\frac{\beta_n^2}{\delta_n^2}$ big $\Leftrightarrow \frac{p_n}{n}\frac{\mu_n^2}{\delta_n^2}$ big $\Leftrightarrow \frac{p_n}{n} / \frac{\delta_n^2}{\mu_n^2}$

this is saying, shrinkage is not important when. $\frac{p_n}{n}$ is negligible compared to $\boxed{\frac{\delta_n^2}{\mu_n^2}}$

scale-free mean of dispersion of eigenvalues.

what bothers me is that this depend on $\frac{p_n}{n}$.

- find bona fide estimator.

$m_n = \langle S_n, I_n\rangle$ consistent for μ_n ~~$E m_n = \mu_n$~~ $m_n^2 - \mu_n^2 \xrightarrow{q.m.} 0 \quad m_n - \mu_n \xrightarrow{q.m.} 0$

$d_n^2 = \|S_n - m_n I_n\|^2$ consistent for $\delta_n^2 \quad d_n^2 - \delta_n^2 \xrightarrow{q.m.} 0$

$\bar{b}_n^2 = \frac{1}{n^2}\sum_{k=1}^{n}\|X_{\cdot k}^n (X_{\cdot k}^n)^T - S_n\|_n^2$ consistent for $\beta_n^2 \quad \bar{b}_n^2 - \beta_n^2 \xrightarrow{q.m.} 0$.

numerically constrain $\bar{b}_n^2 \leq d_n^2$, because $\beta_n^2 \leq \delta_n^2$, by defining: $b_n^2 = \min(\bar{b}_n^2, d_n^2)$

b_n^2 consistent for $\beta_n^2 \quad b_n^2 - \beta_n^2 \xrightarrow{q.m.} 0$.

$a_n^2 = d_n^2 - b_n^2$ consistent for $\alpha_n^2 \quad a_n^2 - \alpha_n^2 \xrightarrow{q.m.} 0$

$\Rightarrow$ final estimator: $S_n^* = \frac{b_n^2}{d_n^2} m_n I_n + \frac{a_n^2}{d_n^2} S_n$

consistent under Frobenius loss & quadratic mean.

S_n^*, Σ_n^* have asypt same risk. $E\left[\left|\frac{a_n^2 b_n^2}{d_n^2} - \frac{\alpha_n^2 \beta_n^2}{\delta_n^2}\right|\right] \to$

- optimality property.

in this problem: $\min_{\rho_1,\rho_2} \|\Sigma_n^{**}-\Sigma_n\|_n^2$, regardless of ρ_1, ρ_2 being r.v. (depends on sample) or being determined ~~with infor~~, S_n^* is optimal (asypt) wrt risk.

$\Sigma_n^{**} = \rho_1 I_n + \rho_2 S_n$

assume Σ_n^{**} is the solution, S_n^* is asypt of Σ_n^{**}.

① $\|S_n^* - \Sigma_n^{**}\|_n \xrightarrow{q.m} 0. \Longrightarrow$ imply same asypt risk.

② S_n^* is best in combination:

best risk: $\hat{\Sigma}_n = \rho_1 m_n I_n + \rho_2 S_n \Rightarrow \lim_{N\to\infty} \inf_{n\geq N} \left(E\|\hat{\Sigma}_n - \Sigma_n\|_n^2 - E\|S_n^* - \Sigma_n\|_n^2 \right) \geq 0$

$\lim_{n\to\infty} \left(E\|\hat{\Sigma}_n - \Sigma_n\|_n^2 - E\|S_n^* - \Sigma_n\|_n^2 \right) = 0 \Rightarrow \|\hat{\Sigma}_n - S_n^*\|_n \xrightarrow{q.m.} 0.$

③ condition number bdd. (need to check)