Estimation

(1) <u>Definitions:</u>

. Consider $X^i \sim N_p(\mu, \Sigma)$. Data matrix $X = (X^1 \cdots X^n)^T \in M^{n \times p}$. Denote $J_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$.

<u>Def:</u> i) $\bar{X} = \frac{1}{n} \sum_1^n X^i = \frac{X^T J_n}{n}$

ii) $A = \sum_1^n (X^i - \bar{X})(X^i - \bar{X})^T$

$= X^T X - n \bar{X} \bar{X}^T$

$= X^T (I_n - \frac{1}{n} J_n J_n^T) X$

Denote $H = I_n - \frac{1}{n} J_n J_n^T$. $\therefore A = X^T H X$.

iii) $S = \frac{1}{n-1} A$. sample variance matrix.

$D = \text{diag}\{S_{11}, \cdots S_{pp}\}$.

iv) $R = D^{-\frac{1}{2}} S D^{-\frac{1}{2}}$. sample correlated matrix.

<u>Lemma.</u> i) $\text{Cov}(AX, BY) = A\, \text{cov}(X, Y) B^T$.

ii) $E(X^T A Y) = E(X)^T A E(Y) + tr(A\, \text{cov}(X, Y))$

<u>Pf:</u> i) is trivial to see

ii) $E(X^T A Y) = E(\sum x_i a_{ij} \eta_j)$

$= \sum a_{ij} E(x_i \eta_j)$.

Note $\text{cov}(x_i, \eta_j) = E(x_i \eta_j) - E(x_i) E(\eta_j)$

(2) MLE:

① Estimation:

For $X^i \sim N_p(\mu, \Sigma)$, $1 \le i \le n$.

$$L(\mu, \Sigma) = \prod_1^n (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(X^i-\mu)^T \Sigma^{-1}(X^i-\mu)}$$

$$= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_1^n tr(\Sigma^{-1}(X^i-\mu)(X^i-\mu)^T)}$$

Denote: $l(\mu, \Sigma) = \log L(\mu, \Sigma)$.

i) $\sum_{i=1}^n (X^i-\mu)(X^i-\mu)^T = A + n(\bar{X}-\mu)(\bar{X}-\mu)^T$

$$\therefore l = -\frac{np}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma| - \frac{1}{2} tr(\Sigma^{-1}A + n\Sigma^{-1}(\bar{X}-\mu)(\bar{X}-\mu)^T)$$

$$= C - \frac{n}{2}\ln|\Sigma| - \frac{1}{2} tr(\Sigma^{-1}A) - \frac{n}{2}(\bar{X}-\mu)^T \Sigma^{-1}(\bar{X}-\mu)$$

Note: $(\bar{X}-\mu)^T \Sigma^{-1}(\bar{X}-\mu) \ge 0 \Rightarrow \hat{\mu} = \bar{X}$.

ii) Lemma. $B \in M^{p \times p}$, positive definite. Then:

$tr(B) - \ln|B| \ge p$. "=" holds $\Leftrightarrow B = I_p$.

Pf: $\sigma_B = \{\lambda_i\}_1^p$, $\lambda_i > 0$. Then:

$LHS = \sum_1^p \lambda_i - \sum_1^p \ln\lambda_i \ge p$. ($\ln x \le x-1$)

"=" holds $\Leftrightarrow \lambda_i = 1$, $\forall i \Leftrightarrow B = I_p$.

$\Rightarrow$ When $\Sigma = \bar{X}$. Consider:

$$-\frac{1}{2} tr(\Sigma^{-1}A) - \frac{n}{2}\ln|\Sigma| = -\frac{n}{2}\left(tr(\Sigma^{-1}\tfrac{A}{n}) - \ln|\Sigma^{-1}\tfrac{A}{n}| + \ln|\tfrac{A}{n}|\right)$$

$\therefore \hat{\Sigma}^{-1}\frac{A}{n} = I_p$ i.e. $\hat{\Sigma} = \frac{A}{n}$.

So: $(\hat{\mu}, \hat{\Sigma}) = (\bar{X}, A/n)$. $A = \sum_1^n (X^i - \bar{X})(X^i-\bar{X})^T$.

② Properties:

Thm. $X_k \overset{i.i.d}{\sim} N_p(\mu, \Sigma)$. $1 \le k \le n$. Then:

i) $\bar{X} \sim N_p(\mu, \frac{\Sigma}{n})$

ii) $A \sim \sum_1^{n-1} Z_k Z_k^T$. $Z_k \overset{i.i.d}{\sim} N_p(0, \Sigma)$.

iii) $\bar{X}$ is indept with A.

iv) $P(A > 0) = 1 \iff n > p$.

Pf: Let $\Gamma = \begin{pmatrix} & * & \\ \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix} = (r_{ij})_{n\times n}$. orthonormal

So $\sum_{k=1}^{n} r_{ik} = 0$. $\forall i \neq n$. since $r_i \perp r_n$.

$Z = (Z_1 \cdots Z_n)^T = \Gamma (X_1 \cdots X_n)^T = \Gamma X$.

$Z_i \sim N_p(\sqrt{n}\mu\, \delta_{in}, \Sigma)$. $\mathrm{cov}(Z_i, Z_j) = \delta_{ij}\Sigma$.

ii) $A = \sum_1^n X_i X_i^T - n\bar{X}\bar{X}^T = Z^T Z - Z_n Z_n^T$

$= \sum_1^{n-1} Z_k Z_k^T$.

iii) Note $(Z_k)_1^n$ indept. $\bar{X} = \frac{Z_n}{\sqrt{n}}$.

iv) $P(A=0) = P(\exists k.\ Z_k \in \mathrm{span}(Z_i)_{i \neq k})$

$\Rightarrow P(A=0) = 0. \iff n > p.$

Thm. i) Σ is known $\Rightarrow \bar{X}$ is sufficient stat for μ.

ii) μ is known $\Rightarrow A/n$ is sufficient stat for Σ.

iii) $(\bar{X}, A/n)$ is complete sufficient stat for (μ, Σ).

(3) Cramer - Rao Lower Bound:

Denote: $s(X,\theta) = \frac{\partial}{\partial\theta} l(X,\theta) = \frac{1}{L(X,\theta)} \frac{\partial}{\partial\theta} L(X,\theta)$. Score Func.

$\mathcal{J}_n = Var(s(X,\theta))$ Fisher information matrix.

where X is data matrix.

Lemma. If $t = \hat{\theta} = t(X,\theta)$. under regular condition

Then: $\frac{\partial}{\partial\theta} E(t^T) = E(st^T) + E(\frac{\partial t^T}{\partial\theta})$

Pf: $\frac{\partial}{\partial\theta} E(t^T) = \frac{\partial}{\partial\theta} \int t^T(X,\theta) L(X,\theta) dx$

$$= \int (\frac{\partial t^T}{\partial\theta} L + st^T L) dx$$

Cor. If t is unbiased estimator of θ.

Then $E(st^T) = cov(s, t^T) = I$

Pf: $E(s(X,\theta)) = \int \frac{\partial}{\partial\theta} L = \frac{\partial}{\partial\theta} \int L = 0$.

Cor. Set $t = s \Rightarrow \mathcal{J}_n = -E(\frac{\partial^2}{\partial\theta\partial\theta^T} l(X,\theta))$

Thm. $t = \hat{\theta} = t(X)$ is unbiased estimator of θ.

$Var(t) \geq \mathcal{J}_n^{-1}$ under regular condition

Pf: Set $\begin{cases} Y = a^T t \\ Z = c^T s \end{cases}$ $a, c \in \mathbb{R}^p$.

$$\frac{cov(Y,Z)^2}{Var(Y)\,Var(Z)} \leq 1 \text{. Note } cov(s,t) = I.$$

$$\Rightarrow \frac{(a^T c)^2}{a^T \mathrm{Var}(t) a \cdot c^T \mathcal{I}_n c} \le 1 .$$ We optimal it :

$$\max_c \frac{c^T a a^T c}{c^T \mathcal{I}_n c} = a^T \mathcal{I}_n^{-1} a = \max\{\lambda \mid \lambda \in \sigma_{\mathcal{I}_n^{-1} a a^T}\}$$

(since $r(\mathcal{I}_n^{-1} a a^T) = 1 \Rightarrow \lambda^2 - a^T \mathcal{I}_n^{-1} a \lambda$ is eigen. func)

$$\Rightarrow \frac{a^T (\mathcal{I}_n^{-1}) a}{a^T \mathrm{Var}(t) a} \le 1 . \ \forall a . \quad \therefore \mathrm{Var}(t) \ge \mathcal{I}_n^{-1} .$$

<u>Thm.</u> $\{X_i\}$ i.i.d. If $\hat{\theta}$ is MLE of θ. Then

under regular condition : as $n \to \infty$

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N_p(0, \mathcal{I}_1^{-1})$$

<u>Cor.</u> Under condition above. we have :

$$n(\hat{\theta} - \theta)^T \mathcal{I}_1 (\hat{\theta} - \theta) \to_d \chi_p^2 .$$

<u>Rmk:</u> $\{ n(\hat{\theta} - \theta)^T \hat{\mathcal{I}}_1 (\hat{\theta} - \theta) \le \chi^2_{1-\alpha, p} \}$

is $1-\alpha$ confidence region. where

$\hat{\mathcal{I}}_1$ is estimator of $\mathcal{I}_1$.