Home work 3 LIU Yunqin id: 21073799.

1. Maximum Likelihood Method

(1) $L_n(u,\Sigma) = \log \prod_{i=1}^{n} f(x_i | u,\Sigma) = \log \prod_{i=1}^{n} \frac{1}{(2\pi|\Sigma|)^{\frac{1}{2}}} \exp[-\frac{1}{2}(x_i-u)^T \Sigma^{-1}(x_i-u)]$

$= -\frac{1}{2}\sum_{i=1}^{n}(x_i-u)^T\Sigma^{-1}(x_i-u) - \frac{n}{2}\log|\Sigma| - \frac{n}{2}\log(2\pi)$

$= -\frac{1}{2}\sum_{i=1}^{n} tr(\Sigma^{-1}(x_i-u)(x_i-u)^T) - \frac{n}{2}\log|\Sigma| + C.$

$= -\frac{n}{2} tr(\Sigma^{-1}\sum_{i=1}^{n}(x_i-u)(x_i-u)^T/n) - \frac{n}{2}\log|\Sigma| + C$

$= -\frac{n}{2} tr(\Sigma^{-1}S_n) - \frac{n}{2}\log|\Sigma| + C.$

(2) $f(x+\Delta) = tr(A(x+\Delta)^{-1}) = tr(AX^{-\frac{1}{2}}(I+X^{-\frac{1}{2}}\Delta X^{\frac{1}{2}})^{-1}X^{-\frac{1}{2}})$

$\approx tr(AX^{-\frac{1}{2}}(I-X^{-\frac{1}{2}}\Delta X^{\frac{1}{2}})X^{-\frac{1}{2}})$

$= tr(AX^{-1} - AX^{-1}\Delta X^{-1})$

$= tr(AX^{-1}) - tr(AX^{-1}\Delta X^{-1})$

$= f(x) - tr(X^{-1}AX^{-1}\Delta)$

(3) $g(x+\Delta) = \log|X+\Delta|$

$= \log|X^{\frac{1}{2}}(I+X^{-\frac{1}{2}}\Delta X^{\frac{1}{2}})X^{\frac{1}{2}}|$

$= \log|X| + \log|I+X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}|$

$= \log|X| + \sum_{i=1}^{n}\log(1+\lambda_i)$, λ_i is eigenvalues of $X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}$

$\approx \log|X| + \sum_{i=1}^{n}\lambda_i$

$= \log|X| + tr(X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}})$

$= \log|X| + tr(X^{-1}\Delta)$

(4) $L_n(u,\Sigma) = -\frac{n}{2}tr(\Sigma^{-1}S_n) - \frac{n}{2}\log|\Sigma| + C.$

$\frac{d(L_n(u,\Sigma))}{d\Sigma} = \frac{n}{2}\Sigma^{-1}S_n\Sigma^{-1} - \frac{n}{2}\Sigma^{-1} = 0 \Rightarrow \hat{\Sigma}_n^{MLE} = S_n.$

2. Shrinkage: Suppose $y \sim N(u, I_p)$.

(a)

$$L(\hat{u}) = \frac{1}{2}\|y-\hat{u}\|_2^2 + \frac{\lambda}{2}\|\hat{u}\|_2^2$$

$$\frac{\partial L(\hat{u})}{\partial \hat{u}_i} = -(y_i - \hat{u}_i) + \lambda \hat{u}_i = 0$$

$$\Rightarrow \hat{u}_i^{ridge} = \frac{1}{1+\lambda} y_i$$

$$E(u, \hat{u}) = E\|u - \hat{u}^{ridge}\|^2 = Var(\hat{u}) + Bias(\hat{u})^2$$

$$= \sigma^2 tr(C^T C) + \|(1-C)u\|_2^2$$

$$= \sum_{i=1}^{p} C_i^2 + \sum_{i=1}^{p} (1-C_i)^2 u_i^2$$

(b) $\partial_{\hat{u}_i} L(\hat{u}) = \hat{u}_i - y_i + \lambda\, sign(\hat{u}_i)$

$$\hat{u}_i^{soft} = sign(y_i)(|y_i| - \lambda)_+$$

Let $y_i = u_i + z_i$, $z_i \sim N(0,1)$.

$$r_i(\lambda, u_i) = E[\hat{u}_i(u_i + z_i) - u_i]^2 = \int [\hat{u}_i(u_i+z_i) - u_i]^2 \phi(z_i)\, dz_i$$

Here $[\hat{u}_i(u_i+z_i) - u_i]^2 = \begin{cases} (z_i+\lambda)^2 & u_i + z_i < -\lambda \\ u_i^2 & -\lambda < u_i + z_i \le \lambda \\ (z_i - \lambda)^2 & u_i + z_i > \lambda \end{cases}$.

$$\frac{\partial r_i(\lambda, u_i)}{\partial u_i} = 2u_i P(|u+z| \le \lambda) \le 2u_i$$

Thus $r_i(\lambda, u_i) - r_i(\lambda, 0) \le u_i^2$

$$r_i(\lambda, 0) = 2\int_{\lambda}^{\infty} (z_i - \lambda)^2 \phi(z_i)\, dz_i = 2(\lambda^2+1)\tilde{\Phi}(\lambda) - 2\lambda\phi(\lambda).$$

Since $\tilde{\Phi}(\lambda) \le \frac{\phi(\lambda)}{\lambda}$

$$r_i(\lambda, 0) \le \frac{2\phi(\lambda)}{\lambda} \le e^{-\frac{\lambda^2}{2}}$$

$$r_i(\lambda, \infty) = 1 + \lambda^2.$$

So $r_i(\lambda, u) \le r_i(\lambda, 0) + \min\{u_i^2, 1+\lambda^2\}$ Let $\lambda = \sqrt{2\log p}$

$$r_i(\lambda, 0) \le e^{-\frac{\lambda^2}{2}} = \frac{1}{p}$$

$$\le \frac{1}{p} + (2\log p + 1)\min\{u_i^2, 1\}$$

Finally $E\|\hat{u}(y) - u\|^2 = \sum_{i=1}^{p} r_i(\lambda, u) \le 1 + (2\log p + 1)\sum_{i=1}^{p} \min\{u_i^2, 1\}$. better than MLE

(c)

$$l(u_i) = (y_i - u_i)^2 + \lambda^2 I(u_i \neq 0)$$

$$= \begin{cases} y_i^2 & , u_i = 0 \\ (y_i - u_i)^2 + \lambda^2 & , u_i \neq 0 . \end{cases}$$

$\min_{u_i=0} l(u_i) = y_i^2$, $\min_{u_i \neq 0} l(u_i) = \lambda^2$, $\arg\min_{u_i \neq 0} l(u_i) = y_i$.

thus $\arg\min_{u_i} l(u_i) = \begin{cases} 0 & y_i^2 > \lambda^2 \\ y_i & y_i^2 \leq \lambda^2 \end{cases} = y_i I(|u_i| > \lambda)$.

$\hat{u}(y) = y + g(y)$. $g(y) = [1 - I(|y_i| > \lambda)] y_i$

which is not weakly differentiable.

(d) $U_a(y) = p + 2\nabla^T g(y) + \|g(y)\|^2$

$$= p + 2\sum_{i=1}^{p} \left[-a \frac{\|y\|^2 - 2y_i}{\|y\|^4} \right] + \sum_{i=1}^{p} \frac{a^2 y_i^2}{\|y\|^4}$$

$$= p - 2a\left[\frac{p}{\|y\|^2} - \frac{2}{\|y\|^2} \right] + \frac{a^2}{\|y\|^2}$$

$$= p - 2a(p-2) - a^2 / \|y\|^2$$

$a^* = p - 2$.

$E\, U_{a^*}(y) = p - E\frac{(p-2)^2}{\|y\|^2} < p \rightarrow$ the risk of MLE.

(e)

All rules above are Shrinkage rules.

3. Necessary Condition for Admissibility of linear Estimators.

(a) ~~[illegible]~~ . if C is a linear estimator.

证: $r(\hat{u}_C(y), u) = \sigma^2 tr(C^T C) + \|(I-C)u\|^2$

define $D = I - |I - C|$. D is Symmetric.

$\|(I-D)u\|^2 = u^T(I-D)^T(I-D)u = \|(I-C)u\|^2$. bias are equal

$tr(D^T D) = tr I - 2tr(I-D) + tr(I-D)^T(I-D)$.

$tr(D^T D) < tr(C^T C) \Leftrightarrow tr(I-D) = tr|I-C| > tr(I-C). \Leftrightarrow C \neq C^T$. Thus C has to be Symmetric.

(b)

C is symmetric. $C = P\Lambda P^T$. P is orthogonal

$\Lambda = diag(\lambda_1, \cdots \lambda_p)$.

$$r(\hat{u}_c(y), u) = \delta^2 tr(C^TC) + \|(I-C)u\|^2$$

$$= \sum_{i=1}^{p} \delta^2 \lambda_i^2 + (1-\lambda_i)^2 \eta_i^2 \quad , P^T u = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_p \end{pmatrix}$$

Let $\lambda_i = 1$ when $\lambda_i > 1$.

$\lambda_i = 0$ when $\lambda_i < 0$.

(c) $E\|Cy-u\|^2 = E\|\Lambda x - \eta\|^2 \qquad x \sim N(\eta, \delta^2 I_p)$.

$$= \sum_{i=1}^{p} \delta^2 \lambda_i^2 + (1-\lambda_i)^2 \eta_i^2$$

If $\lambda_1 = \lambda_2 = \cdots \lambda_d = 1$. $d \geq 3$.

$$E\|\Lambda x - \eta\|^2 = E\|I_d x_d - \eta_d\|^2 + E\|\Lambda_{-d} X - \eta_{-d}\|^2$$

$$< r(\hat{u}_{J-S}(x_d), \eta_d) + E\|\Lambda_{-d} X - \eta_{-d}\|^2.$$

Thus $d \leq 2$.

4. $\hat{u}^{JS}(Y) = (1 - \frac{p-2}{\|Y\|^2}) Y$.

when $p=1 \qquad \hat{u}^{JS}(Y) = (1 + \frac{1}{Y^2}) Y = Y + \frac{1}{Y}$. $g(Y) = \frac{1}{Y}$

$g(Y)$ is not weakly differentiable.

when $p=2 \qquad \hat{u}^{JS}(Y) = Y = MLE$.

$\|Y\| \sim \chi^2_{p+2N} \mid N \sim P(\frac{\|u\|}{2})$

$$E\left[\frac{1}{\chi^2_{p+2N}}\right] = E\left[E\left[\frac{1}{\chi^2_{p+2N}} \middle| N\right]\right] = E\left[E\left[\frac{1}{\chi^2_{p+2\tilde{n}}}\right]\middle|_{\tilde{n}=N}\right] = E\left[\frac{1}{p-2+2N}\right]$$

$$\geq \frac{1}{p-2+2EN}$$

$$= \frac{1}{p-2+\|u\|^2}$$

Thus $r(\hat{u}^{JS}, u) \leq p - \frac{(p-2)^2}{p-2+\|u\|^2}$.

5. Empirical Bayes Approach to James-Stein Estimator and Tweedie Formula.

(1)

$$p(x) = \int_{-\infty}^{\infty} p(x|\theta)\cdot p(\theta)\cdot d\theta .$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \frac{1}{\sqrt{2\pi A}} e^{-\frac{(\theta-M)^2}{2A}} d\theta .$$

$$= \frac{1}{2\pi\sqrt{A}} \int_{-\infty}^{\infty} e^{-\frac{1}{2A}\left[Ax^2-2Ax\theta+A\theta^2+\theta^2-2M\theta+M^2\right]} d\theta .$$

$$= \frac{1}{2\pi\sqrt{A}} \int_{-\infty}^{\infty} e^{-\frac{1}{2A}\left[(A+1)\theta^2-(2Ax+2M)\theta+Ax^2+M^2\right]} d\theta .$$

$$= \frac{1}{2\pi\sqrt{A}} \sqrt{\frac{2A\pi}{A+1}} \cdot e^{-\frac{(x-M)^2}{2(A+1)}}$$

$$= \frac{1}{\sqrt{2\pi(A+1)}} e^{-\frac{(x-M)^2}{2(A+1)}}$$

$$= N(M, A+1) .$$

(2)

$$p(\theta|x) = \frac{p(x|\theta)\cdot p(\theta)}{p(x)}$$

$$= \frac{N(\theta,1)\cdot N(M,A)}{N(M,A+1)}$$

$$= \frac{1}{2\pi\sqrt{A}} e^{-\frac{1}{2A}\left[A(x-\theta)^2+(\theta-M)^2\right]} \Big/ \frac{1}{\sqrt{2\pi(A+1)}} e^{-\frac{(x-M)^2}{2(A+1)}} .$$

$$= \frac{1}{\sqrt{2\pi\frac{A}{A+1}}} e^{-\frac{A+1}{2A}\left[\theta^2-\frac{Ax+M}{A+1}\cdot 2\theta+\left(\frac{Ax+M}{A+1}\right)^2\right]}$$

let $B=\frac{A}{A+1}$, $C=M+B(x-M)$.

$$p(\theta|x) = \frac{1}{\sqrt{2\pi B}} e^{-\frac{(\theta-C)^2}{2B}} = N(M+B(x-M), B) .$$

(3)

$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ . . $S = \sum_{i=1}^{n} x_i^2$.

$$\frac{\sum(x_i-\bar{x})^2}{A+1} = \frac{S}{A+1} \sim \chi^2(N-1).$$

$$\frac{A+1}{S} \sim \text{Inverse-}\chi^2(N-1).$$

$$E\left(\frac{A+1}{S}\right) = \frac{1}{N-3}$$

$$E\left(\frac{n-3}{S}\right) = \frac{1}{A+1}.$$

$$E\left(1-\frac{n-3}{S}\right) = B$$

$$\hat{B} = 1-\frac{n-3}{S}.$$

$$\hat{\mu}_i^{JS} = \bar{x} + \left(1-\frac{n-3}{S}\right)(x_i-\bar{x}).$$

(4) $R(u) = E_u\{L(u,\hat{u})\} = E_u\|\hat{u}^{MLE} - u\|^2$.

$\hat{u}^{MLE} = x$. $\quad \pi \to \log \to \frac{\partial}{\partial u}$.

$$R^{(MLE)}(u) = E_u\|x-u\|^2.$$

$$= E_u\left\{E\sum_{i=1}^{N}(x_i-u_i)^2\right\}.$$

$x_i \sim N(u_i, 1)$.

$x_i - u_i \sim N(0,1)$.

$$\sum_{i=1}^{N}(x_i-u_i)^2 \sim \chi^2(N)$$

Thus $R^{(MLE)}(\chi^2(N)) = N$

$$E\|\hat{u}^{MLE} - u\|^2 = n.$$

$\cancel{E\|\hat{u}^{Bays} - u\|^2}$ $= E_u[(Bx-u)'(Bx-u)]$

$R^{Bayes}(u)$. $= E_u[B^2x'x - 2Bx'u + u'u]$

$= (B^2-2B+1)\|u\|^2 + nB^2$.

$= (1-B)^2\|u\|^2 + nB^2$.

$E\|\hat{u}_{Bays} - u\|^2 = E_A\{R^{Bays}(u)\} = (1-B^2)^2 \cdot nA + nB^2$

$= nB$.

(b) Tweedie formula.

$$p(x|\theta) = N(\theta, \sigma^2).$$

$$E[\theta|x] = \int_{-\infty}^{\infty} \theta\, p(\theta|x)\, d\theta$$

$$= \int_{-\infty}^{\infty} \theta \frac{p(x|\theta)\cdot p(\theta)}{p(x)} \cdot d\theta$$

$$= \int_{-\infty}^{\infty} \frac{\theta\, p(x|\theta)\cdot p(\theta)\cdot d\theta}{p(x)}.$$

$$= \frac{1}{p(x)} \cdot \int_{-\infty}^{\infty} \theta \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} p(\theta)\, d\theta.$$

$$= \frac{1}{p(x)} \int_{-\infty}^{\infty} \left[\sigma^2 \frac{\theta - x}{\sigma^2} e^{-\frac{(x-\theta)}{2\sigma^2}} p(\theta) + x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} p(\theta)\right] d\theta.$$

$$= \frac{1}{p(x)} \left[\sigma^2 \cdot \int_{-\infty}^{\infty} \frac{d\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}\right]}{dx} p(\theta)\cdot d\theta + \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} p(\theta)\, d\theta\right]$$

$$= \frac{1}{p(x)} \left[\sigma^2 \int_{-\infty}^{\infty} \frac{dp(x|\theta)}{dx} \cdot p(\theta)\cdot d\theta + \int_{-\infty}^{\infty} x \cdot p(x|\theta)\cdot p(\theta)\cdot d\theta\right]$$

$$= \frac{1}{p(x)} \left[\sigma^2 \frac{d}{dx} \int_{-\infty}^{\infty} p(x|\theta)\cdot p(\theta)\cdot d\theta + x \int_{-\infty}^{\infty} p(x|\theta)\cdot p(\theta)\cdot d\theta\right]$$

$$= \frac{\sigma^2 \cdot \frac{dp(x)}{dx} + x\, p(x)}{p(x)}$$

$$= x + \sigma^2 \frac{d}{dx} \log p(x).$$