Empirical Bayes and Shrinkage Estimators

- Empirical Bayes estimation.

$X_i \sim N(\theta_i, 1) \quad i=1,\cdots,p.$

unknown : $\theta_1 \cdots \theta_p$.

Estimate θ_i by $\delta_i(X)$.

compound squared error loss: $L(\theta,\delta) = \sum_{i=1}^{p} (\theta_i - \delta_i(X))^2$

Risk function for δ : $R(\theta,\delta) = E_\theta \| \delta(X) - \theta \|^2$.

$\delta(X) = (\delta_1(X), \cdots \delta_p(X))'$, $\theta = (\theta_1, \cdots \theta_p)$.

Bayesian : $\Theta_1 \cdots \Theta_p \overset{iid}{\sim} N(0, \tau^2)$. prior.

Given $\Theta = \theta$, $X_1 \cdots X_p$ ind. $X_i \sim N(\theta_i, 1)$.

$$f(x|\Theta=\theta) = \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{p}(x_i-\theta_i)^2}$$

$$f(\theta) = \frac{1}{(2\pi\tau^2)^{\frac{p}{2}}} e^{-\frac{1}{2\tau^2}\sum_{i=1}^{p}\theta_i^2}$$

$$f(x,\theta) = \frac{1}{(2\pi\tau)^{p}} e^{-\frac{1}{2}\sum_{i=1}^{p}(x_i-\theta_i)^2 - \frac{1}{2\tau^2}\sum_{i=1}^{p}\theta_i^2}$$

$$f(\Theta|X) = \frac{1}{\left(2\pi\frac{\tau^2}{1+\tau^2}\right)^{\frac{p}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{p}\frac{\left(\theta_i - \frac{x_i}{1+1/\tau^2}\right)^2}{\tau^2/1+\tau^2}}.$$

$$\theta_i \mid X = x \sim N\left(\frac{x_i}{1+\frac{1}{\tau^2}}, \frac{\tau^2}{1+\tau^2}\right) \qquad i=1,\dots p.$$

$$E(\theta_i \mid X) = \frac{X_i}{1+\frac{1}{\tau^2}}. \qquad Var(\theta_i \mid X) = \frac{\tau^2}{1+\tau^2}.$$

The expected loss under the Bayesian model:

$$E L(\theta, \delta(X)) = E \sum_{i=1}^{p} \left(\delta_i(X) - \theta_i\right)^2$$

$$= E\left[E \sum_{i=1}^{p} (\delta_i(X) - \theta_i)^2 \,\middle|\, X\right]$$

$$= E \sum_{i=1}^{p} \left[Var(\theta_i \mid X) + \left(\frac{X_i}{1+\frac{1}{\tau^2}} - \delta_i(X)\right)^2\right]$$

The risk is minimized when

$$\delta_i(X) = \frac{X_i}{1+\frac{1}{\tau^2}} = \left(1 - \frac{1}{1+\tau^2}\right) X_i = \frac{\tau^2}{1+\tau^2} X_i.$$ <u>shrinkage</u>!

---

choosing τ is crucial.

under the Bayesian model: $X_1 \cdots X_p \overset{iid}{\sim} N(0, 1+\tau^2)$.

the UMVUE of $1+\tau^2$ is $\frac{\sum_{i=1}^{p} X_i^2}{p}$.

the James-Stein estimator of θ is based on estimating $\frac{1}{1+\tau^2}$ by $\frac{p-2}{\|X\|^2}$. $\Rightarrow \delta_{JS}(X) = \left(1 - \frac{p-2}{\|X\|^2}\right) X$.

τ: hyperparameter.

Bayesian: treat τ as a RV with its own prior.

$\longrightarrow$ hierarchical Bayes.

Empirical Bayes: use data to estimate it.

- Risk of the James-Stein Estimator.

Lemma (Stein): $X \sim N(\mu, \sigma^2)$. $h: \mathbb{R} \to \mathbb{R}$ differentiable. $E(|h'(X)|) < \infty$.

Then. $E\big((X-\mu)\, h(X)\big) = \sigma^2 E\, h'(X)$.

pf: $\begin{cases} \mu = 0 \\ \sigma^2 = 1 \end{cases}$. $h(0) = 0$.

$$\int_0^\infty x\, h(x)\, e^{-\frac{x^2}{2}}\, dx = \int_0^\infty x \left[\int_0^x h'(y)\, dy\right] e^{-\frac{x^2}{2}}\, dx$$

$$= \int_0^\infty \int_0^\infty I(y<x) \cdot x\, h'(y)\, e^{-\frac{x^2}{2}}\, dy\, dx$$

$$= \int_0^\infty h'(y) \left[\int_y^{+\infty} x\, e^{-\frac{x^2}{2}}\, dx\right] dy$$

$$= \int h'(y)\, e^{-\frac{y^2}{2}}\, dy.$$

Similarly: $\int_{-\infty}^{0} x h(x) e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{0} h'(y) e^{-\frac{y^2}{2}} dy.$

$\Rightarrow \quad E X h(X) = E h'(X). \quad X \sim N(0.1).$

In general: $Z = \frac{X-\mu}{\sigma} \sim N(0,1) \quad X = \mu + \sigma Z.$

$$E\left((X-\mu) h(X) \right) = \sigma E\left(Z h(\mu+\sigma Z)\right)$$
$$= \sigma^2 E h'(\mu+\sigma Z)$$
$$= \sigma^2 E h'(X)$$

Lemma: $X_1 \cdots X_p$ independent. $X_i \sim N(\theta_i, 1)$

$h: \mathbb{R}^p \to \mathbb{R}^p$. $\quad E\|Dh(X)\| < \infty$. Euclidean norm

$$\Rightarrow \quad E\left((X-\theta)' h(X)\right) = E\, tr\left(Dh(X)\right)$$

$X_i \sim N(\theta_i, 1) \quad i = 1, \cdots p.$ $\qquad$ define $h(X) = X - \delta(X)$

$\delta(X) = X - h(X).$ $\qquad$ ex: J-S estimator: $h(X) = \frac{p-2}{\|X\|^2} X$

Define: $\hat{R} = p + \|h(X)\|^2 - 2\, tr(Dh(X))$

Then $R(\theta, \delta) = E_\theta \|\delta(X) - \theta\|^2 = E_\theta \hat{R}$

$$R(\theta, \delta) = E_\theta \sum_{i=1}^{p} \left(X_i - \theta_i - h_i(X) \right)^2$$

$$= E_\theta \left[\sum_{i=1}^{p} (X_i - \theta_i)^2 + \sum_{i=1}^{p} h_i^2(X) - 2 \sum_{i=1}^{p} (X_i - \theta_i) h_i(X) \right]$$

$$= p + E_\theta \| h(X) \|^2 - 2 E_\theta (X - \theta)' h(X)$$

$$= p + E_\theta \| h(X) \|^2 - 2 E_\theta \, tr \left(D h(X) \right)$$

J-S: $\quad h(X) = \dfrac{p-2}{\|X\|^2} X. \qquad h_i(x) = \dfrac{(p-2) x_i}{x_1^2 + \cdots + x_p^2}.$

$$\frac{\partial h_i(x)}{\partial x_i} = \frac{p-2}{\|x\|^2} - \frac{2(p-2) x_i^2}{\|x\|^4}$$

$$tr\, D h(x) = \frac{(p-2)^2}{\|x\|^2}.$$

$$\| h(x) \|^2 = \frac{(p-2)^2}{\|x\|^2}.$$

$$\hat{R} = p - \frac{(p-2)^2}{\|X\|^2}$$

$$R(\theta, \delta_{JS}) = E_\theta \hat{R} = E_\theta \left[p - \frac{(p-2)^2}{\|X\|^2} \right] < p = R(\theta, X).$$

when $p > 2$. J-S estimator risk $<$ risk of X.

X is inadmissable!