Notes from Problems Class 2.

Let $X = (X_1, \dots, X_p)^T \sim N_p(\theta, I_p)$ where

$\theta = (\theta_1, \dots, \theta_p)^T$ and I_p is the $p \times p$ identity matrix

and $p \geqslant 3$.

Thus, the X_is are independent $N(\theta_i, 1)$ with

$$f_X(x_1, \dots, x_p \mid \theta) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \exp\left\{ - \frac{(x_i - \theta_i)^2}{2} \right\}$$

For a single observation $X = x$, the maximum likelihood estimate is x (i.e. $x_i = \theta_i$ for each i) which is unbiased.

Consider estimation of θ using quadratic loss

$$L(\theta, d) = (\theta - d)^T (\theta - d) = \sum_{i=1}^{p} (\theta_i - d_i)^2$$

for decision $d = (d_1, \dots, d_p)^T \in \mathbb{R}^p$.

The CLASSICAL RISK of $\delta^\circ(X) = X$ is

$$R(\theta, \delta^\circ) = \mathbb{E}[L(\theta, \delta^\circ(X)) \mid \theta]$$

This could be a function of θ.

X is treated as random for fixed θ.

$$= \sum_{i=1}^{p} \mathbb{E}[(\theta_i - X_i)^2 \mid \theta]$$

$$= \sum_{i=1}^{p} \mathrm{Var}(X_i \mid \theta)$$

$$= p.$$

Consider the set of JAMES-STEIN ESTIMATORS,

$$\delta^a(X) = \left(1 - \frac{a}{X^T X}\right) X$$

for $a \geq 0$. [$a=0$ gives $\delta^0(X) = X$]. Note that if $X^T X > 0$ then $\delta^a(X)$ shrinks X towards 0.

We now consider the classical risk of $\delta^a(X)$ under quadratic loss. In order to do so we will need to make use of STEIN'S LEMMA

For $X \sim N(\theta, I_p)$ and $g(X)$ suitably behaved real valued function,

$$\mathbb{E}\left(g(X)(X_i - \theta_i) \mid \theta\right) = \mathbb{E}\left[\frac{\partial g(X)}{\partial X_i} \,\Big|\, \theta\right]$$

In one-dimension but the extension is clear,

$$\mathbb{E}\left(g(X)(X_i - \theta_i) \mid \theta\right) = \int_{-\infty}^{\infty} g(X) \overbrace{\frac{1}{\sqrt{2\pi}} (x_i - \theta_i) \exp\left\{-\frac{(x_i-\theta_i)^2}{2}\right\}} \, dx_i$$

and so I can do this integral by parts

e.g. $u = g(x) \quad \partial u = \dfrac{\partial g(x)}{\partial x_i} \qquad \partial v = \dfrac{1}{\sqrt{2\pi}} (x_i - \theta_i) \exp\left\{-\dfrac{(x_i - \theta_i)^2}{2}\right\}$

$$v = -\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \theta_i)^2}{2}\right\}$$

Putting the parts together gives:

$$\left[-g(x) \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(x_i - \theta_i)^2}{2} \right\} \right]_{-\infty}^{\infty}$$

$$+ \int_{-\infty}^{\infty} \frac{\partial g(x)}{\partial x_i} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(x_i - \theta_i)^2}{2} \right\} dx_i$$

$$= \mathbb{E}\left[\frac{\partial g(X)}{\partial X_i} \,\middle|\, \theta \right].$$

$$R(\theta, \delta^a(X)) = \mathbb{E}\left[\left(\theta - \left(1 - \frac{a}{X^TX}\right)X\right)^T \left(\theta - \left(1 - \frac{a}{X^TX}\right)X\right) \,\middle|\, \theta \right]$$

$$= \mathbb{E}\left[\left((\theta - X) + \frac{a}{X^TX} X \right)^T \left((\theta - X) + \frac{a}{X^TX} X \right) \,\middle|\, \theta \right]$$

$$= \underbrace{\mathbb{E}[(\theta - X)^T(\theta - X) \mid \theta]}_{= p \text{ as } R(\theta, \delta^0(X))} + a^2 \mathbb{E}\left[\frac{1}{X^TX} \,\middle|\, \theta \right]$$

$$- 2a\, \mathbb{E}\left[\frac{X^T(X - \theta)}{X^TX} \,\middle|\, \theta \right]$$

use Stein's Lemma with $g(X) = \frac{X_i}{X^TX}$

Now, $\mathbb{E}\left[\frac{X^T(X - \theta)}{X^TX} \,\middle|\, \theta \right] = \sum_{i=1}^{p} \mathbb{E}\left[\frac{X_i}{X^TX}(X_i - \theta_i) \,\middle|\, \theta \right]$

$$= \sum_{i=1}^{p} \mathbb{E}\left[\frac{\partial}{\partial X_i} \frac{X_i}{X^TX} \,\middle|\, \theta \right]$$

$$= \sum_{i=1}^{p} \mathbb{E}\left[\frac{X^TX - 2X_i^2}{(X^TX)^2} \,\middle|\, \theta \right]$$

$$= \mathbb{E}\left[\frac{p\, X^T X - 2\sum_{i=1}^{p} X_i^2}{(X^T X)^2} \,\middle|\, \theta \right]$$

$$= (p-2)\, \mathbb{E}\left[\frac{1}{X^T X} \,\middle|\, \theta \right].$$

Hence,

remember that $p \geq 3$

$$R(\theta, \delta^a(X)) = \underbrace{p}_{R(\theta, \delta^0(X))} + (a^2 - 2a(p-2))\, \mathbb{E}\left[\frac{1}{X^T X} \,\middle|\, \theta \right]$$

Now, $X^T X \geq 0$ and so $\mathbb{E}\left[\frac{1}{X^T X} \,\middle|\, \theta \right] \geq 0$ then if

$$a^2 - 2a(p-2) < 0$$

i.e. $0 < a < 2(p-2)$ (such an a exists for $p \geq 3$)

we have $R(\theta, \delta^a(X)) < R(\theta, \delta^0(X))$ for all θ

Thus, $\delta^0(X)$ is INADMISSIBLE.

Note $\frac{d}{da} R(\theta, \delta^a(X)) = (2a - 2(p-2))\, \mathbb{E}\left[\frac{1}{X^T X} \,\middle|\, \theta \right]$

$\Rightarrow a = p-2$ is the minimum

Note if $\theta = 0$ then $X^T X \sim \chi^2_p$ so that $\mathbb{E}\left[\frac{1}{X^T X} \,\middle|\, \theta \right] = \frac{1}{p-2}$

so then $R(\theta, \delta^0(X)) = p$ and $R(\theta, \delta^{p-2}(X)) = 2$

and so $R(\theta, \delta^{p-2}(X)) << R(\theta, \delta^0(X))$ for p large.

Note: $\delta^a(X) = \left(1 - \frac{a}{X^T X}\right) X$ and so the ith term is

$$\left(1 - \frac{a}{X^T X}\right) X_i$$ and so depends on ALL $X_1, \dots, X_p$.

This phenomenon is known as STEIN'S PHENOMENON and it can be shown to occur in many situations when comparing three or more populations.

Occurs because the loss function is dealing with SIMULTANEOUS estimation of ALL parameters. It's an ON AVERAGE property. If you had a loss function that related to just an individual component then x_i would be fine.

Shrinkage: reduce variance at the expense of bias.

In Bayesian statistics, the prior and the likelihood are combined within the posterior: we "pull" the likelihood towards the prior.
Posterior estimate often sees a classical estimate pulled towards a corresponding prior.

[loss function is $L(\theta, d) = (\theta - d)^T (\theta - d)$
$$= \sum_{i=1}^{p} (\theta_i - d_i)^2$$ ← total sum of the individual losses

Aim to minimise loss: this loss functions does them on average across all θ_i]

$\alpha\mu_0 + (1-\alpha)x_i$

Thus the variance of the posterior is smaller that the classical variance.

Suppose I take $Q_i \sim N(\mu_0, \sigma_0^2)$ independent then

$$Q_i \mid x_i = N\left(\underbrace{\left(\frac{1}{\sigma_0^2}+\frac{1}{1}\right)^{-1}\left(\frac{\mu_0}{\sigma_0^2}+\frac{x_i}{1}\right)}_{\text{weighted average of } \mu_0 \text{ and } x},\ \left(\frac{1}{\sigma_0^2}+\frac{1}{1}\right)^{-1}\right)$$

Under quadratic loss, the Bayes rule for this prior is the posterior expectation $E[\theta \mid X]$ and if $\sigma_0^2 \to \infty$ $E[\theta \mid X] \to X$.

Inadmissible $\delta^*(X)$ means I can't find a proper prior for which $\delta^\circ(X)$ is the Bayes rule [of the posterior decision]. Here it is essentially, the Bayes rule of an improper uniform

Normal model: mle x, variance σ^2 $[\bar{x}, \sigma^2/n]$