Coursework - Fundamentals of Statistical Inference

CID: 02091191

November 9, 2021

Abstract

This document summarises the work conducted for the first Fundamentals of Statistical Inference coursework of the year.

(1) Modified James-Stein estimator

Let X have a p-dimensional normal distribution ($p \geq 4$), with mean vector μ and covariance matrix I (the $p \times p$ identity matrix) so that $X_i \sim N(\mu_i, 1)$, independently, $i = 1, \ldots, p$.
Recall that the James-Stein estimator of μ is defined as

$$\delta_{JS}(X) = \left(1 - \frac{p-2}{\|X\|^2}\right) X$$

where $\|X\|^2 = \sum_{i=1}^{p} X_i^2$.
Consider the modified James-Stein estimator

$$d^a(X) = \bar{X} e_p + \left(1 - \frac{a}{V}\right)(X - \bar{X} e_p)$$

where $a \in \mathbb{R}$, $\bar{X} = \frac{1}{p}\sum_{i=1}^{p} X_i$, $V = \sum_{i=1}^{p}(X_i - \bar{X})^2$ and $e_p = (1, \ldots, 1)^T$ the p-dimensional vector of ones. Assume the loss function $L(\mu, d) = \|\mu - d\|^2$, known as the squared error loss. The risk function of $d^a(X)$ is defined as

$$R(\mu, d^a) = \mathbb{E}[L(\mu, d^a)] = \mathbb{E}\|\mu - d^a\|^2$$

For the sake of readability, we do not write $\mathbb{E}_\mu$ but we should recall that $\mathbb{E}$ denotes here the expectation with respect to the distribution of X for the given μ. Let us first focus on the case $a = 0 \Rightarrow d^0(X) = X$. Hence, the risk function of this obvious estimator of μ assuming the mean vector is equal to the observation vector X is

$$R(\mu, d^0(X)) = \mathbb{E}\|\mu - X\|^2 = \sum_{i=1}^{p} \mathbb{E}\left[(\mu_i - X_i)^2\right] = \sum_{i=1}^{p} \mathbb{V}\left[X_i\right] = p$$

as $\mathbb{V}[X_i] = 1$. Note that this risk is irrespective of μ and thus that d^0 is an equaliser decision rule.
Now consider the general case $a \in \mathbb{R}$.

$$\begin{aligned}
R\left(\mu, d^a(X)\right) &= \mathbb{E}\left\|\mu - d^a(X)\right\|^2 \\
&= \mathbb{E}\left\|\mu - \left(\bar{X} e_p + \left(1 - \frac{a}{V}\right)\left(X - \bar{X} e_p\right)\right)\right\|^2 \\
&= \mathbb{E}\left\|(\mu - X) + \frac{a}{V}\left(X - \bar{X} e_p\right)\right\|^2 \\
&= \underbrace{\mathbb{E}\left\|\mu - X\right\|^2}_{p} + \mathbb{E}\left[\frac{a^2}{V^2} \underbrace{\left(X - \bar{X} e_p\right)^T \left(X - \bar{X} e_p\right)}_{V}\right] - 2\mathbb{E}\left[\frac{a}{V}\left(X - \bar{X} e_p\right)^T (X - \mu)\right] \\
&= p + a^2 \mathbb{E}\left[\frac{1}{V}\right] - 2a\mathbb{E}\left[\frac{\left(X - \bar{X} e_p\right)^T (X - \mu)}{V}\right]
\end{aligned}$$

Focus on the last term

$$\mathbb{E}\left[\frac{\left(X-\bar{X}e_p\right)^T(X-\mu)}{V}\right] = \mathbb{E}\left[\sum_{i=1}^{p}\frac{\left(X_i-\bar{X}\right)(X_i-\mu_i)}{V}\right]$$
$$= \sum_{i=1}^{p}\mathbb{E}\left[(X_i-\mu_i)\frac{\left(X_i-\bar{X}\right)}{V}\right]$$

Stein's Lemma enables to write, provided $\sigma^2=\mathbb{V}[X_i]=1$

$$\mathbb{E}\left[(X_i-\mu_i)\frac{\left(X_i-\bar{X}\right)}{V}\right] = \mathbb{E}\left[\frac{\partial}{\partial X_i}\left\{\frac{X_i-\bar{X}}{V}\right\}\right]$$

On the one hand,

$$\frac{\partial}{\partial X_i}\left\{X_i-\bar{X}\right\} = \frac{\partial}{\partial X_i}\left\{X_i-\frac{1}{p}\sum_{k=1}^{p}X_k\right\} = 1-\frac{1}{p}$$

And on the other hand,

$$\frac{\partial V}{\partial X_i} = \frac{\partial}{\partial X_i}\left\{\sum_{j=1}^{p}\left(X_j-\frac{1}{p}\sum_{k=1}^{p}X_k\right)^2\right\} = \frac{\partial}{\partial X_i}\left\{\sum_{j=1}^{p}X_j^2-p\left(\frac{1}{p}\sum_{j=1}^{p}X_j\right)^2\right\}$$
$$= 2X_i-\frac{2}{p}\sum_{j=1}^{p}X_j = 2\left(X_i-\bar{X}\right)$$

Hence, re-writing the full derivative gives us

$$\mathbb{E}\left[\frac{\partial}{\partial X_i}\left\{\frac{X_i-\bar{X}}{V}\right\}\right] = \mathbb{E}\left[\frac{\left(1-\frac{1}{p}\right)V-2\left(X_i-\bar{X}\right)^2}{V^2}\right]$$

And finally

$$\mathbb{E}\left[\frac{\left(X-\bar{X}e_p\right)^T(X-\mu)}{V}\right] = \sum_{i=1}^{p}\mathbb{E}\left[\frac{\left(1-\frac{1}{p}\right)V-2\left(X_i-\bar{X}\right)^2}{V^2}\right]$$
$$= \mathbb{E}\left[\sum_{i=1}^{p}\frac{\left(1-\frac{1}{p}\right)}{V}-\frac{2}{V^2}\sum_{i=1}^{p}\left(X_i-\bar{X}\right)^2\right]$$
$$= \mathbb{E}\left[\frac{(p-1)}{V}-\frac{2}{V}\right]$$
$$= (p-3)\,\mathbb{E}\left[\frac{1}{V}\right]$$

Thus

$$R\left(\mu,d^a(X)\right) = p+a^2\mathbb{E}\left[\frac{1}{V}\right]-2a\left(p-3\right)\mathbb{E}\left[\frac{1}{V}\right]$$
$$= p-\left[2a(p-3)-a^2\right]\mathbb{E}\left[\frac{1}{V}\right]$$

□

We may now focus on how this modified James-Stein estimator improves on $d(X)=X$. We simply note that $R\left(\mu,d^a(X)\right)<R\left(\mu,d^0(X)\right)=p$ which is the case provided $2a(p-3)-a^2>0$ i.e. $0<a<2(p-3)$. For such values of a, $d^a(X)$ strictly dominates $d^0(X)$, thus for such values of a the obvious estimator of X is inadmissible. To compare, for the classical James-Stein estimator $\delta_{JS}(X)$, it was proved in the lectures notes [1] that for $0 < a < 2(p-2)$ the obvious estimation of X is inadmissible.

An optimal value of a would be a value of a which minimises the risk. By calculating the first derivative with respect to a of the risk and equalling to zero, $2(p-3) - 2a = 0$ i.e. the risk is minimised for $a = p - 3$ (the risk function is a convex second degree polynomial with respect to a, providing this is a minimum and not a maximum). To compare, for the classical James-Stein estimator $\delta_{JS}(X)$, the risk is minimised for $a = p - 2$.

Now let us compare the risks of the modified James-Stein estimator and the classical James-Stein estimator when $\mu = 0$.

- **James-Stein estimator:** Recall that the risk function we found in the lecture notes [1] when $a = p - 2$ (optimal value) can be written as

$$R(0, \delta_{JS}(X)) = p - (p-2)^2 \mathbb{E}\left[\frac{1}{\|X\|^2}\right]$$

  As $\|X\|^2 = X^T X \sim \chi^2_p$ (sum of squares of p independent standard normal distributions). Hence, $\mathbb{E}\left[\frac{1}{\|X\|^2}\right]$ follows an inverse χ^2 distribution with p degrees of freedom, so $\mathbb{E}\left[\frac{1}{\|X\|^2}\right] = \frac{1}{p-2}$. Thus,

$$R(0, \delta_{JS}(X)) = 2$$

- **Modified James-Stein estimator:** Recall that the risk function of the modified James-Stein estimator when a is optimal, i.e. $a = p - 3$ is

$$R\left(0, d^{p-3}(X)\right) = p - (p-3)^2 \mathbb{E}\left[\frac{1}{V}\right]$$

  We note that the unbiased estimation of the variance derived from the observations $(X_i)_{i=1}^p$ is defined as $S^2 = \frac{1}{p-1}\sum_{i=1}^p (X_i - \bar{X})^2 = \frac{V}{p-1}$ and we know that $\frac{(p-1)S^2}{\sigma^2} \sim \chi^2_{p-1}$. Here, $\sigma = 1$ then $V \sim \chi^2_{p-1}$. Therefore, $\mathbb{E}\left[\frac{1}{V}\right] = \frac{1}{(p-1)-2} = \frac{1}{p-3}$. Hence,

$$R\left(0, d^{p-3}(X)\right) = 3$$

Thus, when $\mu = 0$, the modified James-Stein estimator has a higher risk than the classical James-Stein estimator. $\square$

(2) Another estimator of the mean vector

Now suppose that $X_1, \ldots, X_p$ ($p \geq 4$) are independent random variables such that $X_i \sim N(\theta_i, \sigma^2)$ for $i = 1 \ldots, p$ with *unknown* variance $\sigma^2 > 0$.
Let $\tilde{\theta}_r$ be defined as

$$\tilde{\theta}_r = \left(1 - \frac{r(p-2)\sigma^2}{\|X\|^2}\right) X$$

As defined, $\tilde{\theta}_r$ is an estimator of the mean vector $\theta = (\theta_1 \ldots, \theta_p)^T$. Let us compute its risk function under the squared error loss.

$$\begin{aligned}
R(\theta, \tilde{\theta}_r) &= \mathbb{E}\|\theta - \tilde{\theta}_r\|^2 \\
&= \mathbb{E}\left\|(\theta - X) + \frac{r(p-2)\sigma^2}{\|X\|^2} X\right\|^2 \\
&= \mathbb{E}\|\theta - X\|^2 + \mathbb{E}\left[\frac{r^2(p-2)^2\sigma^4}{\|X\|^4}\|X\|^2\right] - 2\mathbb{E}\left[\frac{r(p-2)\sigma^2}{\|X\|^2} X^T(X - \theta)\right]
\end{aligned}$$

where $\mathbb{E}\|\theta - X\|^2 = \sum_{i=1}^p \mathbb{E}\left[(\theta_i - X_i)^2\right] = \sum_{i=1}^p \mathbb{V}[X_i] = p\sigma^2$. Then

$$R(\theta, \tilde{\theta}_r) = p\sigma^2 + r^2(p-2)^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\right] - 2r(p-2)\sigma^2\sum_{i=1}^p \mathbb{E}\left[(X_i - \theta_i)\frac{X_i}{\|X\|^2}\right]$$

Using *Stein's Lemma* with $\sigma^2 > 0$,

$$\begin{aligned}
R(\theta, \tilde{\theta}_r) &= p\sigma^2 + r^2(p-2)^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\right] - 2r(p-2)\sigma^2\sum_{i=1}^p \sigma^2\mathbb{E}\left[\frac{\partial}{\partial X_i}\left\{\frac{X_i}{\|X\|^2}\right\}\right] \\
&= p\sigma^2 + r^2(p-2)^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\right] - 2r(p-2)\sigma^4\sum_{i=1}^p \mathbb{E}\left[\frac{\|X\|^2 - 2X_i^2}{\|X\|^4}\right] \\
&= p\sigma^2 + r^2(p-2)^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\right] - 2r(p-2)^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \\
&= \sigma^2\left\{p - (2r - r^2)(p-2)^2\sigma^2\mathbb{E}\left[\frac{1}{\|X\|^2}\right]\right\}
\end{aligned}$$

$\square$

We may notice that since σ^2 is unknown, we can not use $\tilde{\theta}_r$ as an estimator of θ.

Suppose that there exists a statistic S^2 independent of $X = (X_1, \ldots, X_p)^T$ such that $\frac{S^2}{\sigma^2} \sim \chi_m^2$. Let $\hat{\sigma}^2 = tS^2$ for a constant $t > 0$. Let $\tilde{\theta}$ be a new estimator of θ defined as

$$\tilde{\theta} = \left(1 - \frac{(p-2)\hat{\sigma}^2}{\|X\|^2}\right)X$$

Let us first compute the risk of this estimator under the squared error loss.

$$\begin{aligned}
R(\theta, \tilde{\theta}) &= \mathbb{E}\left\|(\theta - X) + \frac{(p-2)t\sigma^2}{\|X\|^2}\frac{S^2}{\sigma^2}X\right\|^2 \\
&= \underbrace{\mathbb{E}\|\theta - X\|^2}_{p\sigma^2} + (p-2)^2t^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\left(\frac{S^2}{\sigma^2}\right)^2\right] - 2(p-2)t\sigma^2\mathbb{E}\left[\frac{X^T(X-\theta)}{\|X\|^2}\frac{S^2}{\sigma^2}\right] \\
&= p\sigma^2 + (p-2)^2t^2\sigma^4\mathbb{E}\left[\frac{1}{\|X\|^2}\right]\mathbb{E}\left[\left(\frac{S^2}{\sigma^2}\right)^2\right] - 2(p-2)t\sigma^2\mathbb{E}\left[\frac{X^T(X-\theta)}{\|X\|^2}\right]\mathbb{E}\left[\frac{S^2}{\sigma^2}\right]
\end{aligned}$$

since $\frac{S^2}{\sigma^2}$ and X are independent, by *Independence Lemma*: if X and Y are two independent random variables, then for any functions g and h, $g(X)$ and $h(Y)$ are also independent. Using *Stein's Lemma* exactly as previously, we have

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

Also, since $\frac{S^2}{\sigma^2} \sim \chi_m^2$,

$$\mathbb{E}\left[\frac{S^2}{\sigma^2}\right] = m$$

and

$$\mathbb{E}\left[\left(\frac{S^2}{\sigma^2}\right)\right] = \mathbb{V}\left[\frac{S^2}{\sigma^2}\right] + \left(\mathbb{E}\left[\frac{S^2}{\sigma^2}\right]\right)^2 = 2m + m^2 = m(m+2)$$

Thus, we obtain

$$\begin{aligned}
R(\theta, \tilde{\theta}) &= p\sigma^2 + (p-2)^2t^2\sigma^4m(m+2)\mathbb{E}\left[\frac{1}{\|X\|^2}\right] - 2(p-2)^2\sigma^4m\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \\
&= \sigma^2\left\{p - \left(2t - t^2(m+2)\right)m(p-2)^2\sigma^2\mathbb{E}\left[\frac{1}{\|X\|^2}\right]\right\}
\end{aligned}$$

$\square$

We may first notice that the risk function of $\tilde{\theta}$ is not exactly the same as the one of $\tilde{\theta}_r$ for some value of r: that is due to the fact that S^2 has to be seen as a random variable and not a simple constant quantity, and thus we can not just plug in a "good" value of r in the first risk function (we may be tempted here to plug $r = \frac{S^2}{\sigma^2}$ into the risk function of $\tilde{\theta}_r$ to obtain the risk function of $\tilde{\theta}$, but we would not get the correct risk function for $\tilde{\theta}$ as S^2 is a random variable and not a constant).

$\tilde{\theta}$ dominates X as an estimator of θ provided $2t - t^2(m+2) > 0$ i.e. $0 < t < \frac{2}{m+2}$.

The optimal value of t is such that it minimises the risk. By calculating the first derivative of the risk function with respect to t and equalling it to zero, we obtain $2 - 2t(m+2) = 0$ i.e. $t = \frac{1}{m+2}$. $\square$

References

[1] Pr G.A. Young (October 2021) *MATH70078 - Fundamentals of Statistical Inference Lecture Notes*, Imperial College London MSc Statistics resources