ISyE 6412A **Theoretical Statistics** **Fall 2023**

HW #5 (due at Canvas midnight on Wednesday, September 27, ET)

(There are 6 questions. The hints are available on the second page of this pdf file.)

1. Let $Y_1, \ldots, Y_n$ be an iid random sample from a population with unknown mean θ and a finite variance $\sigma^2 > 0$. In the problem of estimating $\theta \in \Omega = \mathbb{R}^1$ (real numbers) with $D = \mathbb{R}^1$ and $L(\theta, d) = (\theta - d)^2$ (squared error loss), consider a general statistical procedure of the form $\delta_{a,b} = a\bar{Y} + b$, where a and b are general constants.

   (a) Calculate the risk function of $\delta_{a,b} = a\bar{Y} + b$.

   Answer: The risk function is

$$
\begin{aligned}
R_{\delta_{a,b}}(\theta) &= \mathbf{E}_\theta(\theta - \delta_{a,b})^2 \\
&= \mathbf{E}_\theta(\delta_{a,b} - \theta)^2 \qquad \text{(since } (-x)^2 = x^2\text{)} \\
&= \mathbf{E}_\theta(a\bar{Y} + b - \theta)^2 \\
&= (a\mathbf{E}_\theta\bar{Y} + b - \theta)^2 + a^2\mathrm{Var}_\theta(\bar{Y}) \\
&= \Big[(a-1)\theta + b\Big]^2 + \frac{a^2\sigma^2}{n}.
\end{aligned}
$$

□

   (b) Show that $a\bar{Y} + b$ is **inadmissible** whenever (i) $a > 1$; or (ii) $a < 0$; or (iii) $a = 1, b \neq 0$.

   Answer: (i) If $a > 1$, then by part (a),

$$
R_{\delta_{a,b}}(\theta) \geq \frac{a^2\sigma^2}{n} > \frac{\sigma^2}{n} = R_{\delta_{1,0}}(\theta) = R_{\bar{Y}}(\theta),
$$

   so that $\bar{Y}$ is better than $\delta_{a,b}$, and thus $\delta_{a,b}$ is inadmissible when $a > 1$.
   (ii) If $a < 0$, then $1 - a > 0$ and $(1-a)^2 > 1$. Hence

$$
\begin{aligned}
R_{\delta_{a,b}}(\theta) &\geq \Big[(a-1)\theta + b\Big]^2 = \Big[(1-a)\theta - b\Big]^2 \\
&= (1-a)^2\Big[\theta - \frac{b}{1-a}\Big]^2 \\
&> \Big[\theta - \frac{b}{1-a}\Big]^2 \\
&= R_{\delta_{0,b/(1-a)}}(\theta).
\end{aligned}
$$

   In other words, the constant estimator $\delta_{0,b/(1-a)} \equiv \frac{b}{1-a}$ will be better than $\delta_{a,b}$ when $a < 0$.
   (iii) In this case, $a = 1, b \neq 0$, we have

$$
R_{\delta_{a,b}}(\theta) = b^2 + \frac{\sigma^2}{n} > \frac{\sigma^2}{n} = R_{\delta_{1,0}}(\theta) = R_{\bar{Y}}(\theta),
$$

   and thus $\delta_{a,b}$ is also dominated by $\bar{Y}$.

□

2. In Problem 1, assume further that $Y_i \sim N(\theta, \sigma^2)$, where $\sigma^2 > 0$ is known. Show that $a\bar{Y} + b$ is **admissible** if $0 \le a < 1$.

   Remark: Combining problem 1 and problem 2 studies the admissibility of $a\bar{Y} + b$ for the normal distribution for all cases except $\bar{Y}$ itself which will be shown to be admissible in class. Therefore, under the normality assumption, the procedure $\delta_{a,b} = a\bar{Y} + b$ is **admissible** if and only if (i) $0 \le a < 1$ or (ii) $a = 1, b = 0$.]

   Answer: Here we need to consider two subcases separately: (i) $0 < a < 1$ and (ii) $a = 0$. For the case of $0 < a < 1$, recall that the unique Bayes procedure with respect to the prior distribution $\theta \sim N(\mu, \tau^2)$ is

$$\frac{n/\sigma^2}{n/\sigma^2 + 1/\tau^2}\bar{Y} + \frac{1/\tau^2}{n/\sigma^2 + 1/\tau^2}\mu.$$

   It is clear that for given σ^2, $0 < a < 1$ and any b, we can always choose μ and τ so that

$$\frac{n/\sigma^2}{n/\sigma^2 + 1/\tau^2} = a \qquad \text{and} \qquad \frac{1/\tau^2}{n/\sigma^2 + 1/\tau^2}\mu = b$$

   In fact, a simply calculation shows that

$$\tau^2 = \frac{a}{1-a}\frac{\sigma^2}{n} \qquad \text{and} \qquad \mu = \frac{b}{1-a}.$$

   In other words, as long as $0 < a < 1$, $\delta_{a,b} = a\bar{Y} + b$ is the unique Bayes procedure with respect to a well-defined prior distribution, as the variance $\tau^2 > 0$ is well-defined. Since any unique Bayes procedure is admissible, we can conclude that $a\bar{Y} + b$ is admissible if $0 < a < 1$.
   Meanwhile, for the case of $a = 0$, the above argument does not work, since $\tau^2 = 0$ cannot be the variance of a normal distribution. In this case, $\delta_{a,b} = a\bar{Y} + b = b$ is the constant estimator, which has **zero** risk at $\theta = b$. We did a similar problem in part (h) of HW#1, and see the solution there for the detailed arguments! $\square$

3. In Problem 1, assume further that $Y_i \sim Bernoulli(\theta)$, i.e., $\mathbf{P}_\theta(Y_i = 1) = 1 - \mathbf{P}_\theta(Y_i = 0) = \theta$. Assume that the statistician decides to restrict consideration to procedures of the form $\delta_{a,b} = a\bar{Y} + b$, and want to always yield only decisions in $D = [0, 1]$ regardless of the observed value $\mathbf{Y} = (Y_1, \cdots, Y_n)$. Show that the pair (a, b) has to be chosen to satisfy $0 \le b \le 1$ and $-b \le a \le 1 - b$.

   Answer: Observe that the term $ay + b$ is a linear function of y, and thus it is minimized/maximized at the end points. Since $\bar{y}$ takes the values between 01 and 1, it suffices that $ay + b \in [0, 1]$ for $y = 0$ and 1. This implies that

$$0 \le b \le 1 \qquad \text{and} \qquad 0 \le a + b \le 1,$$

   or equivalently,

$$0 \le b \le 1 \qquad \text{and} \qquad -b \le a \le 1 - b.$$

$\square$

4. In Problem 3 for Bernoulli distribution, when $0 < b < 1$ and $-b < a < 0$, is the procedure $\delta_{a,b} = a\bar{Y} + b$ admissible? Note that the variance $\sigma^2 = \theta(1-\theta)$ depends on θ here.

Answer: The answer is "No." To see this, denote the risk function of $\delta_{a,b}(X) = a\bar{Y} + b$ by

$$\rho(a,b) = R_{\delta_{a,b}}(\theta) = \mathbf{E}_\theta(a\bar{Y} + b - \theta)^2 = ((a-1)\theta + b)^2 + a^2\theta(1-\theta)/n.$$

If $a < 0$, then $(a-1)^2 > 1$ and hence

$$\begin{aligned} \rho(a,b) &\geq ((a-1)\theta + b)^2 \qquad \text{(since the second term } a^2\theta(1-\theta)/n \geq 0) \\ &= (a-1)^2\Big(\theta - \frac{b}{1-a}\Big)^2 \\ &> \Big(\theta - \frac{b}{1-a}\Big)^2 = \rho(\frac{b}{1-a}, 0). \end{aligned}$$

Thus, $a\bar{Y} + b$ is dominated by the constant estimator $\delta \equiv b/(1-a)$.
Remark: In general, let $Y_1, \cdots, Y_n$ be i.i.d. with mean θ and finite variance. Then in the problem of estimating θ under squared error loss, $a\bar{Y}+b$ is an inadmissible procedure whenever $a < 0$. □

5. In Problem 3 for Bernoulli distribution, show that when $0 < b < 1$ and $0 \leq a < 1 - b$, the procedure $\delta_{a,b} = a\bar{Y} + b$ is admissible.

Remark: For the purpose of completeness, for Bernoulli distribution, it can be shown that $a\bar{Y}+b$ is admissible in the closed triangle $\{(a,b) : a \geq 0, b \geq 0, a+b \leq 1\}$, and it is inadmissible for the remaining values of a and b.

Answer: Here we need to prove it in two separate cases: (i) $0 < a < 1$ and (ii) $ba = 0$. First, let us consider the case when $0 < a < 1 - b$ and $0 < b < 1$. Recall that the Bayes procedure for the prior distribution is $Beta(\alpha, \beta)$ (this prior is well-defined if and only if $\alpha > 0$ and $\beta > 0$) is

$$\delta^*_\pi = \frac{\alpha + \sum_i nY_i}{\alpha + \beta + n} = \frac{n}{\alpha+\beta+n}\bar{Y} + \frac{\alpha}{\alpha+\beta+n}.$$

Set

$$\frac{n}{\alpha+\beta+n} = a \qquad \text{and} \qquad \frac{\alpha}{\alpha+\beta+n} = b,$$

we have

$$\alpha = \frac{b}{a}n \qquad \text{and} \quad \beta = \frac{1-a-b}{a}n.$$

The key observation is that in the above equation, $\alpha > 0$ and $\beta > 0$ when $0 < a < 1 - a$ and $0 < b < 1$. That is, $\delta_{a,b}$ is the unique Bayes procedure relative to a well-defined prior $Beta(\alpha = \frac{b}{a}n, \beta = \frac{1-a-b}{a}n)$ when $0 < a < 1$ and $0 \leq b < 1 - a$. Thus it is admissible.
Meanwhile, in the case of $a = 0$, the above argument does not work, since $\alpha = \beta = \infty$. In this case, $\delta_{a,b} = a\bar{Y} + b = b$ is the constant estimator, which has **zero** risk at $\theta = b$ when $0 < b < 1$. The remaining proof will be similar to those in HW#1. Below is the detailed arguments.
Assume that the constant estimator $\delta_{a=0,b} = b$ with $0 < b < 1$ were inadmissible and there existed a procedure δ' which were better than $\delta_{0,b} = b$. Then

$$R_{\delta'}(\theta) \leq R_{\delta_{0,b}}(\theta) = (\theta - b)^2 \qquad \text{for all} \quad 0 \leq \theta \leq 1,$$

$$R_{\delta'}(\theta_0) \le R_{\delta_{0,b}}(\theta_0) = (\theta - b)^2 \qquad \text{for at least one} \quad 0 \le \theta_0 \le 1.$$

Since

$$\begin{aligned} R_{\delta'}(\theta) &= \mathbf{E}_\theta L(\theta, \delta') = \mathbf{E}_\theta(\delta' - \theta)^2 \\ &= \sum_{a_1=0}^{1} \cdots \sum_{a_n=0}^{1} (\delta'(a_1, \cdots, a_n) - \theta)^2 \mathbf{P}_\theta(Y_1 = a_1, \cdots, Y_n = a_n) \\ &= \sum_{a_1=0}^{1} \cdots \sum_{a_n=0}^{1} (\delta'(a_1, \cdots, a_n) - \theta)^2 \theta^{\sum_i^n a_i}(1-\theta)^{n-\sum_i^n a_i}, \end{aligned}$$

we have that for any $0 \le \theta \le 1$,

$$\sum_{a_1=0}^{1} \cdots \sum_{a_n=0}^{1} (\delta'(a_1, \cdots, a_n) - \theta)^2 \theta^{\sum_i^n a_i}(1-\theta)^{n-\sum_i^n a_i} \quad \le \quad (\theta - b)^2.$$

Letting $\theta = b$ both sides yields that

$$\sum_{a_1=0}^{1} \cdots \sum_{a_n=0}^{1} (\delta'(a_1, \cdots, a_n) - b)^2 b^{\sum_i^n a_i}(1-b)^{n-\sum_i^n a_i} \le 0.$$

Since all terms on the left-hand side are non-negative and $0 < b < 1$, we have $\delta'(a_1, \cdots, a_n) = b$ for all $a_1, \cdots, a_n = 0, 1$ and this implies that $\delta'(Y_1, \cdots, Y_n) = b$ and $R_{\delta'}(\theta) = (\theta - b)^2$ for all $0 \le \theta \le 1$. This is a contradiction with our assumption that the procedure δ' is better than $\delta_{0,b} = b$! So $\delta_{0,b} \equiv b$ is admissible when $0 < b < 1$. This completes the proof. □

6. Let $Y_1, \cdots, Y_n$ be i.i.d. according to a $N(0, \sigma^2)$ density, and let $S^2 = \sum_{i=1}^n Y_i^2$. We are interested in estimating $\theta = \sigma^2$ under the squared error loss $L(\theta, d) = (\theta - d)^2 = (\sigma^2 - d)^2$ using linear estimator $\delta_{a,b} = aS^2 + b$, where a and b are constants. **Show that**

   (a) The risk of $\delta_{a,b}$ is given by

   $$R_{\delta_{a,b}}(\sigma^2) = \mathbf{E}_\sigma\Big(\sigma^2 - (aS^2 + b)\Big)^2 = 2na^2\sigma^4 + [(an-1)\sigma^2 + b]^2.$$

   (b) The constant estimator $\delta_{a=0,b=0} = 0$ is inadmissible.

   Remark: this exercise illustrates the fact the constants are not necessarily admissible.

   Answer: (a) Let $Y_i = \sigma Z_i$, where $Z_i \sim N(0, 1)$. For the standard normal distributions, we have

   $$\mathbf{E}(Z_i) = 0, \mathbf{E}(Z_i^2) = 1, \mathbf{E}(Z_i^3) = 0, \mathbf{E}(Z_i^4) = 3.$$

   From this, let us now find the mean and variance of $W_i = Y_i^2$, which becomes $\sigma^2 Z_i^2$ under $\mathbf{P}_{\theta=\sigma^2}$.

   $$\begin{aligned} \mathbf{E}_\theta(W_i) &= \mathbf{E}(\sigma^2 Z_i^2) = \sigma^2 \mathbf{E}(Z_i^2) = \sigma^2; \\ Var_\theta(W_i) &= \mathbf{E}_\theta(W_i^2) - [\mathbf{E}_\theta(W_i)]^2 = \mathbf{E}(\sigma^4 Z_i^4) - [\sigma^2]^2 \\ &= \sigma^4 \mathbf{E}(Z_i^4) - \sigma^4 = 3\sigma^4 - \sigma^4 \\ &= 2\sigma^4. \end{aligned}$$

Thus under the squared error loss, the risk function of $\delta_{a,b} = a\sum_i^n W_i + b$ when estimating $\theta = \sigma^2$ is

$$\begin{aligned}
R_{\delta_{a,b}}(\theta) &= \mathbf{E}_\theta(\theta - \delta_{a,b})^2 \\
&= \mathbf{E}_\theta(\delta_{a,b} - \theta)^2 \\
&= \mathbf{E}_\theta(a\sum_i^n W_i + b - \theta)^2 \\
&= (a\sum_i^n \mathbf{E}_\theta W_i + b - \theta)^2 + a^2\sum_i^n \mathrm{Var}_\theta(W_i) \\
&= (an\sigma^2 - \sigma^2 + b)^2 + a^2 n(2\sigma^4) \\
&= [(an-1)\sigma^2 + b]^2 + 2na^2\sigma^4.
\end{aligned}$$

(b) By (a), when $b = 0$, let us now consider the estimator of the form $\delta_{a,b=0} = a\sum_i^n Y_i^2$, and its risk function is

$$\begin{aligned}
R_{\delta_{a,b=0}}(\theta) &= [(an-1)^2 + 2na^2]\sigma^4 = [(n^2+2n)a^2 - (2n)a + 1]\sigma^4 \\
&= [(n^2+2n)(a - \frac{1}{n+2})^2 + \frac{2}{n+2}]\sigma^4,
\end{aligned}$$

which reaches the minimum value of $\frac{2}{n+2}\sigma^4$ when $a = \frac{1}{n+2}$. In particular,

$$R_{\delta_{a=0,b=0}}(\theta) = \sigma^4 > \frac{2}{n+2}\sigma^4 = R_{\delta_{a=\frac{1}{n+2},b=0}}(\theta)$$

for all $\sigma^2 > 0$. Thus the procedure $\delta_{a=0,b=0}$ is inadmissible. □

Hints for problem 1 (b) Find a better procedure than $\delta_{a,b} = a\bar{Y} + b$: try $\delta_{1,0} = \bar{Y}$ for case (i) ($a > 1$) or case (iii) ($a = 1$ and $b \neq 0$); and try some constant estimators for case (ii) ($a < 0$). In particular, when $a < 0$, we have $1 - a > 1$, and thus

$$\Big[(a-1)\theta + b\Big]^2 = \Big[(1-a)\theta - b\Big]^2 = (1-a)^2\Big[\theta - \frac{b}{1-a}\Big]^2 \geq \Big[\theta - \frac{b}{1-a}\Big]^2.$$

From this, can you guess the desired constant estimator?

Hints for problem 2 Show that $\delta_{a,b}$ is a Bayes procedure if $0 < a < 1$, and can you find μ, τ^2 (in term of a, b) so that $\delta_{a,b}$ is Bayes with respect to the prior distribution $\theta \sim N(\mu, \tau^2)$? Meanwhile, when $a = 0$, note that $\delta_{a=0,b}$ is the only estimator with zero risk at $\theta = b$, and have we done similar questions in part (h) of HW#1?

Hints for problem 3 it suffices to make sure that $\delta_{a,b} \in [0, 1]$ when $(Y_1, \cdots, Y_n) = (0, \cdots, 0)$ or $(1, \cdots, 1)$, why? Hints: where does the linear function achieve the minimum or maximum values?

Hints for problem 4 Here the variance $\sigma^2 = \theta(1-\theta)$, and it is a special case of problem #1(b).

Hints for problem 5 If $a = 0$, what is risk at $\theta = b$? If $0 < a < 1 - b$, what is the Bayes solution relative to the prior distribution $\pi(\theta) = \text{Beta}(\alpha, \beta)$ with $\alpha > 0$ and $\beta > 0$?

Hints for problem 6 in part (a), let $Y_i = \sigma Z_i$, where $Z_i \sim N(0, 1)$. For the standard normal distributions, we have

$$\mathbf{E}(Z_i) = 0, \mathbf{E}(Z_i^2) = 1, \mathbf{E}(Z_i^3) = 0, \mathbf{E}(Z_i^4) = 3.$$

From this, can you find the mean and variance of $W_i = Y_i^2$? The question can be reduced to the linear estimator of $\delta_{a,b} = a\sum_{i=1}^n W_i + b$ when estimating $\theta = \mathbf{E}_\theta(W_i) = \sigma^2$ under the squared error loss function.
In part (b), let $b = 0$, find a that minimizes the risk function, and such a will yield a better procedure.