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

HW #4 (due at Canvas midnight on Wednesday, September 20, ET)

(There are 6 questions, and please look at both sides.)

The hints to Problems 1, 2, 4, 5(b), and 6(c) can be found on page #3 of this pdf file.)

1. (**Finding minimax procedures, case I**). Suppose that Y is Bernoulli random variable with $\mathbf{P}_\theta(Y=1)=\theta$ and $\mathbf{P}_\theta(Y=0)=1-\theta$, and it is desired to guess the value of θ on the basis of X under the squared error loss function $L(\theta,d)=(\theta-d)^2$. Assume that the domain of θ is $\Omega=\{\frac{4}{9}\le\theta\le\frac{5}{9}\}$ (since we still want to estimate θ, the decision space D can still be $[0,1]$). Find the minimax procedure under the squared error loss function when $\Omega=\{\frac{4}{9}\le\theta\le\frac{5}{9}\}$.

Answer: The method is the almost identical as the example of we illustrated in class when $\Omega=\{\frac{1}{3}\le\theta\le\frac{2}{3}\}$. Suppose a prior π puts mass $\frac{1}{2}$ at $\theta=\frac{4}{9}$ and mass $\frac{1}{2}$ at $\theta=\frac{5}{9}$, i.e.,

$$\pi(\theta=\frac{4}{9})=\frac{1}{2}=\pi(\theta=\frac{5}{9}).$$

Also recall that $\mathbf{P}_\theta(Y=y)=\theta^y(1-\theta)^{1-y}$ for $y=0$ or 1. With these preparations, it is useful to split the following steps to find the Bayes procedure.

Step 1. Let us construct a 2×2 table with each cell being the joint probability $f(\theta,y)=\pi(\theta)\mathbf{P}_\theta(Y=y)$ for $\theta=\frac{4}{9}$ or $\frac{5}{9}$ and $y=0$ or 1.

	$\theta=\frac{4}{9}$	$\theta=\frac{5}{9}$
$y=0$	$\pi(\theta=\frac{4}{9})\mathbf{P}_{\theta=\frac{4}{9}}(Y=0)=\frac{1}{2}\frac{5}{9}=\frac{5}{18}$	$\pi(\theta=\frac{5}{9})\mathbf{P}_{\theta=\frac{5}{9}}(Y=0)=\frac{1}{2}\frac{4}{9}=\frac{2}{9}$
$y=1$	$\pi(\theta=\frac{4}{9})\mathbf{P}_{\theta=\frac{4}{9}}(Y=1)=\frac{1}{2}\frac{4}{9}=\frac{2}{9}$	$\pi(\theta=\frac{5}{9})\mathbf{P}_{\theta=\frac{5}{9}}(Y=1)=\frac{1}{2}\frac{5}{9}=\frac{5}{18}$

Step 2. Find the marginal distribution of Y, which is to compute the row sum of the above table:

	$\theta=\frac{4}{9}$	$\theta=\frac{5}{9}$	$m(y)$
$y=0$	$\frac{5}{18}$	$\frac{2}{9}$	$\frac{5}{18}+\frac{2}{9}=\frac{1}{2}$
$y=1$	$\frac{2}{9}$	$\frac{5}{18}$	$\frac{2}{9}+\frac{5}{18}=\frac{1}{2}$

In other words, the marginal distribution of Y is given by $m(Y=0)=\frac{1}{2}$ and $m(Y=1)=\frac{1}{2}$.

Step 3. Find the **posterior distribution** of $\theta|y$ for $y=0$ or $y=1$, which can be derived by dividing each row by the row sum:

	$\pi(\theta=\frac{4}{9}\|y)$	$\pi(\theta=\frac{5}{9}\|y)$
$y=0$	$\frac{\frac{5}{18}}{m(y=0)}=\frac{5}{9}$	$\frac{\frac{2}{9}}{m(y=0)}=\frac{4}{9}$
$y=1$	$\frac{\frac{2}{9}}{m(y=1)}=\frac{4}{9}$	$\frac{\frac{5}{18}}{m(y=1)}=\frac{5}{9}$

Step 4. Now given $Y=y=0$ or 1, we need to find the mean of the corresponding poster distributions. To do so, we need to rewrite the above table into two cases, one for each row. The first one is when $Y=0$, and

θ	$\frac{4}{9}$	$\frac{5}{9}$
$\pi(\theta\|Y=0)$	$\frac{5}{9}$	$\frac{4}{9}$

Thus the correspond mean is

$$\mathbf{E}(\theta|Y=0)=\sum_{\theta=\frac{4}{9} or \frac{5}{9}}(\theta\pi(\theta|Y=0))=\frac{4}{9}\frac{5}{9}+\frac{5}{9}\frac{4}{9}=\frac{40}{81}.$$

Likewise, the second case is when $Y=1$, and

θ	$\frac{4}{9}$	$\frac{5}{9}$
$\pi(\theta\|Y=1)$	$\frac{4}{9}$	$\frac{5}{9}$

Thus the correspond mean is

$$\mathbf{E}(\theta|Y=1) = \sum_{\theta=\frac{4}{9} or \frac{5}{9}} (\theta\pi(\theta|Y=1)) = \frac{4}{9}\frac{4}{9} + \frac{5}{9}\frac{5}{9} = \frac{41}{81}.$$

Hence, under the squared error loss, the corresponding Bayes procedure is

$$\begin{aligned} \delta_\pi(y) &= \text{mean of poster distribution} \pi(\theta|y) \\ &= \mathbf{E}(\theta|y) = \begin{cases} \frac{40}{81}, & \text{if } y=0; \\ \frac{41}{81}, & \text{if } y=1. \end{cases} \end{aligned}$$

Step 5. The risk function of the above-derived Bayes procedure δ_π is

$$\begin{aligned} R_{\delta_\pi}(\theta) &= \mathbf{E}(\theta - \delta_\pi(Y))^2 = (\theta - \delta_\pi(0))^2 \mathbf{P}_\theta(Y=0) + (\theta - \delta_\pi(1))^2 \mathbf{P}_\theta(Y=1) \\ &= (\theta - \frac{40}{81})^2(1-\theta) + (\theta - \frac{41}{81})^2\theta \\ &= (\theta^2 - \frac{80}{81}\theta + \frac{1600}{6561})(1-\theta) + (\theta^2 - \frac{82}{81}\theta + \frac{1681}{6561})\theta \\ &= \frac{79}{81}\theta^2 - \frac{79}{81}\theta + \frac{1600}{6561} \\ &= \frac{79}{81}(\theta - \frac{1}{2})^2 + \frac{1}{26244}. \end{aligned}$$

Since the leading coefficient of this quadratic polynomial is $\frac{79}{81}$, which is positive, on the domain $\Omega = [\frac{4}{9}, \frac{5}{9}]$, the risk function $R_{\delta_\pi}(\theta)$ is minimized at $\theta = \frac{1}{2}$, and is maximized at two endpoints: $\theta = \frac{4}{9}, \frac{5}{9}$ and

$$\begin{aligned} \sup_{\frac{4}{9} \le \theta \le \frac{5}{9}} R_{\delta_\pi}(\theta) &= R_{\delta_\pi}(\theta = \frac{4}{9}) = R_{\delta_\pi}(\theta = \frac{5}{9}) \\ &= \frac{79}{81}(\frac{4}{9} - \frac{1}{2})^2 + \frac{1}{26244} = \frac{80}{26244} = \frac{20}{6561}. \end{aligned}$$

Step 6. The Bayes risk of the Bayes procedure is

$$\begin{aligned} r^*_\pi &= r_{\delta_\pi}(\pi) = \sum_{\theta=\frac{4}{9},\frac{5}{9}} R_{\delta_\pi}(\theta)\mathbf{P}(\theta) \\ &= R_{\delta_\pi}(\theta = \frac{4}{9})\frac{1}{2} + R_{\delta_\pi}(\theta = \frac{5}{9})\frac{1}{2} = \frac{20}{6561} \\ &= \sup_{\frac{4}{9} \le \theta \le \frac{5}{9}} R_{\delta_\pi}(\theta). \end{aligned}$$

Step 7. By Theorem 2 of the mimimax procedure, the Bayes procedure $\delta_\pi(0) = \frac{40}{81}$ and $\delta_\pi(1) = \frac{41}{81}$ is minimax on $\Omega = \{\theta : \frac{4}{9} \le \theta \le \frac{5}{9}\}$. □

2. (**Finding minimax procedures, case II**). When finding the minimax procedures, sometimes the desired prior distribution might not necessarily put the point mass at the boundary of Ω. To illustrate this, let us consider the setting of the previous problem with a single Bernoulli random variable Y, but now we assume that the domain of θ is $\Omega = [\frac{1}{9}, \frac{8}{9}]$. Find the minimax estimator of θ under the squared error loss function $L(\theta, d) = (\theta - d)^2$ when $\Omega = [\frac{1}{9}, \frac{8}{9}]$. For that purpose, let us consider two kinds of Bayesian procedures.

(a) Assume that θ has a prior distribution on two endpoints of Ω with probability mass function $\pi_a(\theta = \frac{1}{9}) = \frac{1}{2}$ and $\pi_a(\theta = \frac{8}{9}) = \frac{1}{2}$. For the corresponding Bayes procedure, denoted by δ_a, show that its Bayes risk $r_{\delta_a}(\pi_a) < \sup_{\theta \in \Omega} R_{\delta_a}(\theta)$, and thus this direction does not work.

(b) Consider another prior distribution $\pi_b(\theta = \frac{2-\sqrt{2}}{4}) = \frac{1}{2}$ and $\pi_b(\theta = \frac{2+\sqrt{2}}{4}) = \frac{1}{2}$, which is a well-defined prior over $\Omega = [\frac{1}{9}, \frac{8}{9}]$, since $\frac{1}{9} < \frac{2-\sqrt{2}}{4} < \frac{2+\sqrt{2}}{4} < \frac{8}{9}$. Show that the corresponding Bayes procedure, denoted by δ_b, satisfies $r_{\delta_b}(\pi_b) = \sup_{\theta \in \Omega} R_{\delta_b}(\theta)$, and thus we can conclude that δ_b is minimax on $\Omega = [\frac{1}{9}, \frac{8}{9}]$.

Answer: To solve (a) and (b), let us first consider a general case of finding the Bayes procedure when the prior distribution is given by $\pi(\theta = r) = \frac{1}{2}$ and $\pi(\theta = 1 - r) = \frac{1}{2}$ for some $0 \leq r \leq \frac{1}{2}$. As in the previous question, under the squared error loss, the corresponding Bayes procedure is (**Please do it by yourself!!!**)

$$\delta_\pi(y) \quad = \quad \begin{cases} 2r(1-r) = u, & \text{if } y = 0 \\ r^2 + (1-r)^2 = 1 - u, & \text{if } y = 1 \end{cases},$$

where we denote $u = 2r - 2r^2$ to simplify the computations below. Next, the risk function of this Bayes procedure is given by

$$A\theta^2 - A\theta + B,$$

where

$$A = 4u - 1 = -8r^2 + 8r - 1 \qquad \text{and} \qquad B = u^2 = 4r^2(1-r)^2.$$

When $r = \frac{4}{9}$, the leading coefficient of the risk function $A > 0$, and the maximum risk is attained at the endpoints, which allows us to prove the minimax properties in Problem 2.
However, when $r = \frac{1}{9}$, we have $u = 16/81$ and the leading coefficient $A = -17/81 < 0$, and thus the maximum risk is attained at $\theta = 1/2$, not at the endpoints, which is the case in Problem 3(a).
Meanwhile, in case 3(b), we have $r = (2 - \sqrt{2})/4$ in case 3(b), and $u = 1/4$ and $A = 0$. Thus we actually have the constant risk function for all $0 \leq \theta \leq 1$, and of course, also constant over $\Omega = [\frac{1}{9}, \frac{8}{9}]$. Thus the corresponding procedure $\delta_b(0) = \frac{1}{4}$ and $\delta_b(1) = \frac{3}{4}$ is minimax over $\Omega = [\frac{1}{9}, \frac{8}{9}]$.
<u>Remarks:</u> recall that we have shown in class that $\delta_b(0) = \frac{1}{4}$ and $\delta_b(1) = \frac{3}{4}$ is minimax over $\Omega = [0, 1]$! Our results actually show that it is minimax over $\Omega = [r, 1 - r]$ for any $0 \leq r \leq \frac{2-\sqrt{2}}{2} \approx 0.1464$. □

3. (**Finding minimax procedures, case III**). When finding the minimax procedures, sometimes the desired prior distribution might not necessarily be symmetric! To see this, under the setting of Problem 1 with a single Bernoulli random variable Y, but we now assume that the domain of θ is $\Omega = [0, \frac{1}{2}]$. Find the minimax estimator of θ under the squared error loss function $L(\theta, d) = (\theta - d)^2$ when $\Omega = [0, \frac{1}{2}]$. To help you find the minimax estimator/procedure when $\Omega = [0, \frac{1}{2}]$, we can split into the following steps.

   (a) Assume that θ has a prior distribution on two endpoints of $\Omega = [0, \frac{1}{2}]$ with probability mass function $\pi(\theta = \frac{1}{2}) = \gamma$ and $\pi(\theta = 0) = 1 - \gamma$. Find the posterior distribution of θ given $Y = y$ for $y = 0$ or 1. [Note that your answer for $\pi(\theta|Y = 0)$ should be different from $\pi(\theta|Y = 1)$, and θ only has two values: 0 and $\frac{1}{2}$].

Answer: In this case, the prior distribution is $\pi(\theta = 0) = 1 - \gamma$ and $\pi(\theta = \frac{1}{2}) = \gamma$, whereas $\mathbf{P}_\theta(Y = y) = \theta^y(1-\theta)^{1-y}$ for $y = 0$ or 1.

<u>*Step 1.*</u> Let us construct a 2×2 table with each cell being the joint probability $f(\theta, y) = \pi(\theta)\mathbf{P}_\theta(Y = y)$ for $\theta = 0$ or $\frac{1}{2}$ and $y = 0$ or 1.

	$\theta = 0$	$\theta = \frac{1}{2}$
$y = 0$	$\pi(\theta = 0)\mathbf{P}_{\theta=0}(Y = 0) = (1-\gamma) * 1 = 1 - \gamma$	$\pi(\theta = \frac{1}{2})\mathbf{P}_{\theta=\frac{1}{2}}(Y = 0) = \gamma\frac{1}{2} = \frac{\gamma}{2}$
$y = 1$	$\pi(\theta = 0)\mathbf{P}_{\theta=0}(Y = 1) = (1-\gamma)0 = 0$	$\pi(\theta = \frac{1}{2})\mathbf{P}_{\theta=\frac{1}{2}}(Y = 1) = \gamma\frac{1}{2}\frac{5}{9} = \frac{\gamma}{2}$

<u>*Step 2.*</u> Find the marginal distribution of Y, which is to compute the row sum of the above table:

	$\theta = 0$	$\theta = \frac{1}{2}$	$m(y)$
$y = 0$	$1 - \gamma$	$\frac{\gamma}{2}$	$1 - \gamma + \frac{\gamma}{2} = 1 - \frac{\gamma}{2}$
$y = 1$	0	$\frac{\gamma}{2}$	$0 + \frac{\gamma}{2} = \frac{\gamma}{2}$

In other words, the marginal distribution of Y is given by $m(Y = 0) = 1 - \frac{\gamma}{2}$ and $m(Y = 1) = \frac{\gamma}{2}$.

<u>*Step 3.*</u> Find the **posterior distribution** of $\theta|y$ for $y = 0$ or $y = 1$, which can be derived by dividing each row by the row sum:

	$\pi(\theta = 0\|y)$	$\pi(\theta = \frac{1}{2}\|y)$
$y = 0$	$\frac{1-\frac{\gamma}{2}}{m(y=0)} = \frac{1-\frac{\gamma}{2}}{1-\frac{\gamma}{2}} = \frac{2-2\gamma}{2-\gamma}$	$\frac{\frac{\gamma}{2}}{m(y=0)} = \frac{\frac{\gamma}{2}}{1-\frac{\gamma}{2}} = \frac{\gamma}{2-\gamma}$
$y = 1$	$\frac{0}{m(y=1)} = \frac{0}{\frac{\gamma}{2}} = 0$	$\frac{\frac{\gamma}{2}}{m(y=1)} = \frac{\frac{\gamma}{2}}{\frac{\gamma}{2}} = 1$

Therefore, when $y = 0$, the posterior distribution of θ is

$$\pi(\theta|Y = 0) = \begin{cases} \frac{2-2\gamma}{2-\gamma}, & \text{if } \theta = 0; \\ \frac{\gamma}{2-\gamma}, & \text{if } \theta = 1/2. \end{cases};$$

and when $y = 1$, the posterior distribution of θ is

$$\pi(\theta|Y = 1) = \begin{cases} 0, & \text{if } \theta = 0; \\ 1, & \text{if } \theta = 1/2. \end{cases}$$

□

(b) Show that the Bayes procedure under the assumption of part (a) is $\delta_B(0) = \frac{\gamma}{4-2\gamma}$ and $\delta_B(1) = \frac{1}{2}$.

Answer: Note that Bayes procedure is given by

$$\begin{aligned} \delta_B(y) &= \text{mean of } \pi(\theta|y) = \sum_{\theta=0 \text{ or } 1/2} \theta\pi(\theta|y) = 0\pi(\theta = 0|y) + \frac{1}{2}\pi(\theta = \frac{1}{2}|y) \\ &= \frac{1}{2}\pi(\theta = \frac{1}{2}|y) = \begin{cases} \frac{1}{2}\frac{\gamma}{2-\gamma} = \frac{\gamma}{4-2\gamma}, & \text{if } y = 0; \\ \frac{1}{2}1 = \frac{1}{2}, & \text{if } y = 1. \end{cases}; \end{aligned}$$

□

(c) Computing the risk function of δ_B and solving the equation $R_{\delta_B}(\theta = 0) = R_{\delta_B}(\theta = \frac{1}{2})$. Show that $\gamma = 2 - \sqrt{2}$. Thus the Bayes procedure in part (b) becomes

$$\delta^*(Y) = \begin{cases} \frac{\sqrt{2}-1}{2}, & \text{if } Y = 0; \\ \frac{1}{2}, & \text{if } Y = 1. \end{cases}$$

Answer: Since $\delta_B(0) = \frac{\gamma}{4-2\gamma}$ and $\delta_B(1) = \frac{1}{2}$, the risk function of δ_B is

$$\begin{aligned} R_{\delta_B}(\theta) &= \mathbf{E}_\theta(\delta_B(Y) - \theta)^2 = (\delta_B(0) - \theta)^2 P_\theta(Y = 0) + (\delta_B(1) - \theta)^2 P_\theta(Y = 1) \\ &= (\frac{\gamma}{4 - 2\gamma} - \theta)^2 (1 - \theta) + (\frac{1}{2} - \theta)^2 \theta \end{aligned}$$

for all $\theta \in \Omega$. Setting $R_{\delta_B}(\theta = 0) = R_{\delta_B}(\theta = \frac{1}{2})$ yields

$$(\frac{\gamma}{4 - 2\gamma})^2 = (\frac{\gamma}{4 - 2\gamma} - \frac{1}{2})^2 \frac{1}{2},$$

which is equivalent to $\gamma^2 - 4\gamma + 2 = 0$ after simplifying. Thus $\gamma = 2 \pm \sqrt{2}$. Since $0 \le \gamma \le 1$ (as a prior probability $\mathbf{P}(\theta = \frac{1}{2})$), we have $\gamma = 2 - \sqrt{2}$. Plugging $\gamma = 2 - \sqrt{2}$ into part (b) yields the procedure $\delta^*(X)$. □

(d) Show that the procedure δ^* in part (c) is minimax on $\Omega = [0, \frac{1}{2}]$.

Answer: For the procedure δ^*, the risk function is

$$\begin{aligned} R_{\delta^*}(\theta) &= (\frac{\sqrt{2} - 1}{2} - \theta)^2 (1 - \theta) + (\frac{1}{2} - \theta)^2 \theta = (\sqrt{2} - 1)\theta(\theta - \frac{1}{2}) + \frac{(\sqrt{2} - 1)^2}{4} \\ &= (\sqrt{2} - 1)[(\theta - \frac{1}{4})^2 - \frac{1}{16}] + \frac{(\sqrt{2} - 1)^2}{4} \end{aligned}$$

Clearly, on the interval $\Omega = [0, \frac{1}{2}]$, it reaches the maximum value of $\frac{(\sqrt{2}-1)^2}{4}$ at the two endpoints: $\theta = 0$ and $1/2$. Also the prior distribution has masses only at the endpoints, and thus δ^* is a Bayes procedure with the corresponding Bayes risk $r_{\delta^*}(\pi) = \sup_{\theta \in \Omega} R_{\delta^*}(\theta)$. Hence, δ^* is minimax on $\Omega = [0, \frac{1}{2}]$.

Remark: in terms of elementary mathematics, the question is to find a and b that minimizes

$$\sup_{\theta \in \Omega} \Big[(a - \theta)^2 (1 - \theta) + (b - \theta)^2 \theta \Big] = \sup_{\theta \in \Omega} \Big[(2a - 2b + 1)\theta^2 - (a^2 + 2a - b^2)\theta + a^2 \Big]$$

for some subinterval Ω of $[0, 1]$. Our results show that when $\Omega = \{0 \le \theta \le 1/2\}$, then the solution is $a = \frac{\sqrt{2}-1}{2}$ and $b = \frac{1}{2}$. Also the previous exercise shows that when $\Omega = \{\frac{4}{9} \le \theta \le \frac{5}{9}\}$, the corresponding solution is $a = \frac{40}{81}$ and $b = \frac{41}{81}$. □

4. (**Finding minimax procedures, case IV**). Sometimes the desired prior distribution for the minimax procedure might not exist, but we can use a sequence of priors to find the minimax procedures.

(a) Let $\pi_k, k = 1, 2, \ldots,$ be a sequence of prior distributions on Ω. Let δ_k denote a Bayes procedure with respect to π_k, and define the Bayes risks of the Bayes procedures

$$r_k = \int_\Omega R_{\delta_k}(\theta) \pi_k(\theta) d\theta.$$

Show that if the sequence r_k converges to a real-valued number r and if δ_* is a statistical procedure with its risk function $R_{\delta_*}(\theta) \le r$ for all $\theta \in \Omega$, then δ_* is minimax.

Answer: Suppose that δ' is another procedure. Then for all k,

$$\begin{aligned}
\sup_{\theta\in\Omega} R_{\delta'}(\theta) &\geq \int_{\Omega} R_{\delta'}(\theta)\pi_k(\theta)d\theta \qquad (\text{since } \textstyle\int_\Omega \pi_k(\theta)d\theta = 1) \\
&\geq \int_{\Omega} R_{\delta_k}(\theta)\pi_k(\theta)d\theta \qquad (\text{since } \delta_k \text{ is Bayes w.r.t. } \pi_k) \\
&= r_k.
\end{aligned}$$

Since this holds for every k, we let k go to ∞, and thus

$$\begin{aligned}
\sup_{\theta\in\Omega} R_{\delta'}(\theta) &\geq r \qquad (\text{let } k \to \infty) \\
&\geq \sup_{\theta\in\Omega} R_{\delta_*}(\theta) \qquad (\text{by property of } \delta_*)
\end{aligned}$$

Thus δ_* is minimax. □

(b) Let $Y_1, \ldots, Y_n$ be iid $N(\theta, \sigma^2)$ with σ known. Consider estimating θ using squared error loss. Show that $\bar{Y} = (Y_1 + \cdots + Y_n)/n$ is a minimax procedure.

Answer: Using the hints, the proof is straightforward. As $\tau \to \infty$, we have $r = \lim_{\tau\to\infty} r_b = \sigma^2/n$. Meanwhile, for every θ, $R_{\bar{Y}}(\theta) = \mathbf{E}_\theta(\theta - \bar{Y})^2 = \sigma^2/n = r$. Thus, by part (a), $\bar{Y}$ is minimax (under the squared error loss).

<u>**Remark:**</u> Suppose, now that σ^2 is unknown. It follows from this problem that the maximum risk of every procedure will be *infinite* unless σ^2 is bounded. When it is assumed that $\sigma^2 \leq M$, then the maximum risk of $\bar{Y}$ is

$$\sup_{(\theta,\sigma^2)} R_{\bar{Y}}(\theta) = M/n.$$

By Problem 5(a) below, $\bar{Y}$ is also minimax subject to the constraint that $\sigma^2 \leq M$. □

5. (**Finding minimax procedures, case V**). Sometimes the minimax properties can be extended from a smaller domain to a larger domain.

(a) Let $Y_1, \ldots, Y_n$ be iid with distribution F and finite unknown expectation θ, where F belongs to a set $\mathcal{F}_1$ of distributions. Suppose we want to estimate θ under a given loss function $L(\theta, d)$. Show that if δ_* is a minimax procedure when F is restricted to some subset $\mathcal{F}_0$ of $\mathcal{F}_1$, and if $\sup_{F\in\mathcal{F}_0} R_{\delta_*}(F) = \sup_{F\in\mathcal{F}_1} R_{\delta_*}(F)$ (i.e., sup risk of δ_* over $\mathcal{F}_1$ is the same as sup risk over $\mathcal{F}_0$), then δ_* is also minimax when F is permitted to vary over $\mathcal{F}_1$.

Answer: If a procedure δ' existed with smaller sup risk over $\mathcal{F}_1$ than δ_*, then

$$\begin{aligned}
\sup_{F\in\mathcal{F}_0} R_{\delta'}(F) &\leq \sup_{F\in\mathcal{F}_1} R_{\delta'}(F) \qquad (\text{since } \mathcal{F}_0 \text{ is a subset of } \mathcal{F}_1) \\
&< \sup_{F\in\mathcal{F}_1} R_{\delta_*}(F) \qquad (\text{by assumption that } \delta' \text{ has smaller sup risk over } \mathcal{F}_1 \text{ than } \delta_*) \\
&= \sup_{F\in\mathcal{F}_0} R_{\delta_*}(F) \qquad (\text{by assumption of } \delta_*)
\end{aligned}$$

Thus δ' will also have smaller sup risk over $\mathcal{F}_0$ than δ_*. This contradict the minimax property of δ_* over $\mathcal{F}_0$. Hence, δ_* is also minimax when F is permitted to vary over $\mathcal{F}_1$.
□

(b) Suppose that $Y_1, \ldots, Y_n$ are iid with unknown mean θ. We further assume that the Y_i's can take any values in the interval $[0, 1]$, i.e., $0 \leq Y_i \leq 1$, and that $\Omega = \{\theta : 0 \leq \theta \leq 1\}$ and $D = \{d : 0 \leq d \leq 1\}$. Show that

$$\delta_* = \frac{\sqrt{n}}{1+\sqrt{n}}\bar{Y} + \frac{1}{1+\sqrt{n}}\frac{1}{2}$$

is minimax for estimating θ under the squared error loss function.

Answer: Following the hints, in order to prove that δ_* is also minimax when F varies over $\mathcal{F}_1$, it suffices to show that the the risk function of δ_* takes on its maximum over $\mathcal{F}_0$, as we have shown that δ_* is minimax as F varies over $\mathcal{F}_0 =$ the set of Bernoulli distributions. Note that

$$\begin{aligned} R_{\delta_*}(F) &= \mathbf{E}_F(\theta - \delta_*)^2 \qquad (\text{recall that } \theta = \mathbf{E}_F(Y_i) = \mathbf{E}_F(\bar{Y})) \\ &= (\mathbf{E}_F(\delta_*) - \theta)^2 + \mathrm{Var}_F(\delta_*) \qquad (\text{as } \mathbf{E}(Y-c)^2 = (\mathbf{E}(Y) - c)^2 + Var(Y)) \\ &= \Big(\frac{\sqrt{n}}{1+\sqrt{n}}\theta + \frac{1}{1+\sqrt{n}}\frac{1}{2} - \theta\Big)^2 + (\frac{\sqrt{n}}{1+\sqrt{n}})^2 \mathrm{Var}_F(\bar{Y}) \\ &= (\frac{1}{1+\sqrt{n}})^2(\frac{1}{2} - \theta)^2 + (\frac{\sqrt{n}}{1+\sqrt{n}})^2 \frac{1}{n}\mathrm{Var}_F(Y_i) \\ &= (\frac{1}{1+\sqrt{n}})^2\Big[(\frac{1}{2} - \theta)^2 + \mathrm{Var}_F(Y_i)\Big]. \end{aligned}$$

Now since $0 \leq Y_i \leq 1$, we have $Y_i^2 \leq Y_i$. Thus, $\mathbf{E}_F(Y_i^2) \leq \mathbf{E}_F(Y_i) = \theta$, and

$$\mathrm{Var}_F(Y_i) = \mathbf{E}_F(Y_i^2) - (\mathbf{E}_F(Y_i))^2 \leq \theta - \theta^2.$$

Combining this inequality with the above relation of $R_{\delta_*}(F)$ yields

$$\begin{aligned} R_{\delta_*}(F) &\leq (\frac{1}{1+\sqrt{n}})^2\Big[(\frac{1}{2} - \theta)^2 + \theta - \theta^2\Big] \\ &= \frac{1}{4(1+\sqrt{n})^2}, \end{aligned}$$

which is the (constant) risk of δ_* over $\mathcal{F}_0 =$ the set of Bernoulli distributions. By part (a), δ_* is also minimax when F varies over $\mathcal{F}_1$. □

6. (**Impact of Loss function on minimax properties**). In HW#1, we assume that $Y_1, \ldots, Y_n$ are iid normal $N(\theta, 1)$ with $\theta \in \Omega = (-\infty, \infty)$ and one of the proposed procedures of estimating θ is

$$\delta_{3,n}(Y_1, \ldots, Y_n) = \frac{\sqrt{n}\,\bar{Y}_n}{1+\sqrt{n}}.$$

(a) Under the loss function $L(\theta, d) = (\theta - d)^2/(1+\theta^2)$, **show** that k can be chosen in the prior density $\pi_a(\theta) = \text{const.} \times (1+\theta^2)k^{-1}\phi(\theta/k)$, where ϕ is the standard normal density, in such a way that $\delta_{3,n}$ is Bayes relative to π_a.

(b) **Show** that $\delta_{3,n}$ is actually minimax under the loss function $L(\theta, d) = (\theta - d)^2/(1+\theta^2)$.

(c) Let us still consider the problem of estimating the normal mean θ, but now under the squared error loss function $L(\theta, d) = (\theta - d)^2$. **Show** that $\delta_{3,n}$ is still Bayes but no longer minimax under the squared error loss function $L(\theta, d) = (\theta - d)^2$.

Answer: (a) When $\pi_a(\theta)$ is a prior density, for each $\mathbf{y} = (y_1, \ldots, y_n)$, a Bayes procedure assigns a decision $d = \delta(\mathbf{y})$ that minimizes

$$\begin{aligned} h^*_\pi(\mathbf{y}, d) &= \int_{-\infty}^{\infty} L(\theta, d) f_\theta(y_1, \cdots, y_n)\pi_a(\theta) d\theta \\ &= \int_{-\infty}^{\infty} \frac{(\theta - d)^2}{1+\theta^2} f_\theta(y_1, \cdots, y_n) C_2 (1+\theta^2) k^{-1}\phi(\theta/k) d\theta \\ &= C_2 \int_{-\infty}^{\infty} (\theta - d)^2 f_\theta(y_1, \cdots, y_n) \frac{1}{k}\phi(\frac{\theta}{k}) d\theta, \end{aligned}$$

where the integral is just the Bayes risk when the loss function function is squared error loss, θ has a prior distribution $N(\mu = 0, b^2 = k^2)$, and the Y_i's are iid $N(\theta, \sigma^2 = 1)$. Thus the integral is minimized at

$$\widehat{d} = \frac{\frac{\sum_i Y_i}{\sigma^2} + \frac{\mu}{b^2}}{\frac{n}{\sigma^2} + \frac{1}{b^2}} = \frac{\frac{n\bar{Y}}{1^2} + \frac{0}{k^2}}{\frac{n}{1^2} + \frac{1}{k^2}} = \frac{n\bar{Y}}{n + 1/k^2}.$$

Let $1/k^2 = \sqrt{n}$ or $k = n^{-1/4}$, then it becomes $\delta_{3,n}$. This implies that $\delta_{3,n}$ minimizes $h^*_\pi(\mathbf{y}, d)$ over D for all $\mathbf{y}$, and thus $\delta_{3,n}$ is Bayes relative to the prior density $\pi_b(\theta)$ with $k = n^{-1/4}$.

(b) In HW#1, we have shown that $\delta_{3,n}$ has a constant risk ($= \frac{1}{(1+\sqrt{n})^2}$). Using the theorem that a (unique) Bayes procedure that has constant risk is minimax, we can conclude that $\delta_{3,n}$ is minimax under the loss function $(\theta - d)^2/(1+\theta^2)$.

(c) The $\delta_{3,n}$ is still Bayes under the squared error loss function, since the proof in (a) actually shows that $\delta_{3,n}$ is Bayes relative to the prior density $\pi_b(\theta) = k^{-1}\phi(\theta/k)$ with $k = n^{-1/4}$ under the squared error loss function.
However, the $\delta_{3,n}$ is no longer minimax under the squared error loss! To see this, note that its risk function under the squared error loss is

$$R_{\delta_{3,n}}(\theta) = \mathbf{E}_\theta\Big(\theta - \frac{\sqrt{n}\ \bar{Y}_n}{1+\sqrt{n}}\Big)^2 = (\theta - \frac{\sqrt{n}\ \theta}{1+\sqrt{n}})^2 + \frac{n}{(1+\sqrt{n})^2}\frac{1}{n} = \frac{\theta^2+1}{(1+\sqrt{n})^2}$$

and thus $\sup_{\theta\in\Omega} R_{\delta_{3,n}}(\theta) = \infty$. On the other hand, under the squared error loss function, $R_{\bar{Y}}(\theta) = \frac{1}{n}$ for all $-\infty < \theta < \infty$ and thus $\sup_{\theta\in\Omega} R_{\bar{Y}}(\theta) = \frac{1}{n} < \infty = \sup_{\theta\in\Omega} R_{\delta_{3,n}}(\theta)$. Indeed, as we have shown (or will show) in class, the minimiax procedure under the squared error loss is $\bar{Y}$. $\square$

Hints for problems 1 and 2: you can consider a general case of finding the Bayes procedure when the prior distribution is given by $\pi(\theta = r) = \frac{1}{2}$ and $\pi(\theta = 1 - r) = \frac{1}{2}$ for some $0 \le r \le \frac{1}{2}$. The corresponding Bayes procedure is of the form $\delta(0) = u$ and $\delta(1) = 1 - u$, where u depends on r. Next, the risk function of this Bayes procedure is of the form

$$A\theta^2 - A\theta + B,$$

where $A = 4u - 1$ and $B = u^2$. When $r = \frac{4}{9}$, the coefficient $A > 0$, which allows us to prove the minimax properties in Problem 2. However, when $r = \frac{1}{9}$, the coefficient $A < 0$, which is the case in Problem 2(a). Meanwhile, problem 2(b) corresponds to the case of $u = 1/4$ and $A = 0$.

Hints for problem 4: In (a), it is possible that $r_k < r$ but $\lim_{k\to\infty} r_k = r$. In (b) consider the priors $N(\mu, \tau^2)$ for θ. We have shown in class that the Bayes risk of the corresponding Bayes procedure is $\frac{1}{n/\sigma^2 + 1/\tau^2}$, denoted by r_τ. What is $r = \lim_{\tau\to\infty} r_\tau$? What is the risk function of $\delta_* = \bar{Y}$? Does the condition in part (a) hold?

Hints for problem 5 (b): Let the subset $\mathcal{F}_0 =$ the set of Bernoulli distributions, and let $\mathcal{F}_1 =$ the class of all distribution functions F with $F(0) = 0$ and $F(1) = 1$, i.e., $0 \le Y_i \le 1$. We have shown in class that δ_* is minimax as F varies over $\mathcal{F}_0$ (i.e., Bayes with constant risk). In order to use part (a), we need to consider the risk function $R_{\delta_*}(F)$ when F varies over $\mathcal{F}_1$. Note that if $0 \le Y_i \le 1$, then $Y_i^2 \le Y_i$ and $\mathbf{E}_F(Y_i^2) \le \mathbf{E}_F(Y_i) = \theta$, with equality if $F \in \mathcal{F}_0$. Use this to prove that when F varies over $\mathcal{F}_1$, $\mathrm{Var}_F(Y_i) \le \theta - \theta^2$, and thus $R_{\delta_*}(F) \le \frac{1}{4(1+\sqrt{n})^2}$, where the right-hand side is the (constant) risk of δ_* over $\mathcal{F}_0$.

Hints for problem 6(c): will it be Bayes relative to the prior density $\pi_b(\theta) = C_1\pi_a(\theta)/(1+\theta^2) = k^{-1}\phi(\theta/k)$ under the squared error loss? What are $\sup_{\theta\in\Omega} R_{\delta_{3,n}}(\theta)$ and $\sup_{\theta\in\Omega} R_{\bar{Y}}(\theta)$ under the squared error loss?]