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

HW #2 (due at Canvas @12:59pm midnight on Friday, September 8 ET)

(There are 4 questions, and please look at both sides).

Problem 1 (Confidence Interval). In a general statistical decision problem, the risk function of a statistical procedure often depends on the unknown parameter θ. However, in some special cases, the risk functions of some families of the procedures may not depend on θ, and when this occurs, it will be straightforward to derive the optimal procedure within these specific families of the procedures, as shown in the following confidence interval estimation problem.

Assume that the random variables $Y_1, Y_2, \cdots, Y_n$ are independent and identically distributed (i.i.d.) with $N(\theta, \sigma^2)$, and σ known. In the interval estimation problem, the decision space for estimating θ is the forms of $d = [L, U]$ with $L = L(Y_1, \cdots, Y_n)$ and $U = U(Y_1, \cdots, Y_n)$, and a widely used loss function is $L(\theta, d) = r * \text{length}(d) - \mathbf{I}(\theta \in d)$ for some constant $r > 0$, where $\mathbf{I}(\theta \in d) = 1$ if $\theta \in d = [L, U]$ and 0 otherwise. That is, the loss function includes two quantities: one is the length of the interval, and the other is whether the interval correctly includes the true θ. Here, we focus on the specific family of the interval estimator of the form

$$\delta_c(\mathbf{Y}) = [\bar{Y}_n - c\sigma, \bar{Y}_n + c\sigma], \qquad \text{where } \bar{Y}_n = (Y_1 + \cdots + Y_n)/n.$$

(a) For each $c \geq 0$, show that the risk function of $\delta_c(\mathbf{Y})$ is given by

$$R_{\delta_c}(\theta) = r(2c\sigma) - \mathbf{P}(-c\sqrt{n} \leq Z \leq c\sqrt{n}) = 2cr\sigma - 2\Phi(c\sqrt{n}) + 1,$$

where $Z \sim N(0,1)$ and $\Phi(z) = \mathbf{P}(Z \leq z)$.

Answer: Here the length of $\delta_c(\mathbf{Y})$ is a constant $2c\sigma$ that does not depend on θ, and thus

$$\begin{aligned} R_{\delta_c}(\theta) &= r\mathbf{E}_\theta(\text{length}(\delta_c(\mathbf{Y}))) - \mathbf{P}_\theta(\bar{Y}_n - c\sigma \leq \theta \leq \bar{Y}_n + c\sigma) \\ &= r(2c\sigma) - \mathbf{P}_\theta(-\sqrt{n}c \leq \sqrt{n}(\bar{Y}_n - \theta)/\sigma \leq \sqrt{n}c) \\ &= 2cr\sigma - \mathbf{P}(-\sqrt{n}c \leq Z \leq \sqrt{n}c), \end{aligned}$$

since $Z = \sqrt{n}(\bar{Y}_n - \theta)/\sigma \sim N(0,1)$ when $Y_1, \cdots, Y_n$ are iid $N(\theta, 1)$ under $\mathbf{P}_\theta$. The second equation follows at once from the fact that $\Phi(-z) = 1 - \Phi(z)$ for all $-\infty < z < \infty$ and $\mathbf{P}(-z \leq Z \leq z) = \Phi(z) - \Phi(-z)$ for any $z \geq 0$. □

(b) Show that the derivative of the risk function in (a) with respect to c is

$$\frac{d}{dc} R_{\delta_c}(\theta) = 2r\sigma - \frac{2\sqrt{n}}{\sqrt{2\pi}} e^{-nc^2/2},$$

which is an increasing function of c for $c \geq 0$.

Answer: Note that $\Phi'(z) = \phi(z) = (1/\sqrt{2\pi}) \exp(-z^2/2)$, and by the chain rule,

$$\frac{d}{dc} R_{\delta_c}(\theta) = 2r\sigma - 2\Phi'(c\sqrt{n})(\sqrt{n}) = 2r\sigma - \frac{2\sqrt{n}}{\sqrt{2\pi}} e^{-nc^2/2}.$$

This is clearly an increasing function of c as only the second term depends on c. □

(c) Show that if $r\sigma > \sqrt{n}/\sqrt{2\pi}$, the derivative is positive for all $c \geq 0$ and hence $R_{\delta_c}(\theta)$ is minimized at $c = 0$. That is, the best interval estimator is the point estimator $\delta_0(\mathbf{Y}) = [\bar{Y}_n, \bar{Y}_n]$.

Answer: If $r\sigma > \sqrt{n}/\sqrt{2\pi}$, then by (b), we have $\frac{d}{dc} R_{\delta_c}(\theta) > 0$ for all $c \geq 0$, and thus the risk function $R_{\delta_c}(\theta)$ is an increasing function of $c \geq 0$. Hence it is minimized at $c = 0$. □

(d) When $r\sigma \le \sqrt{n}/\sqrt{2\pi}$, find the optimal c_{opt} that minimizes the risk function in (a).

Answer: By (b), setting the derivative to be 0 will yield the optimal c_{opt}. Solve

$$2r\sigma - \frac{2\sqrt{n}}{\sqrt{2\pi}} e^{-nc^2/2} = 0,$$

we have

$$c_{opt} = \sqrt{-\frac{2}{n}\log\left(\frac{r\sigma\sqrt{2\pi}}{\sqrt{n}}\right)}.$$

□

(e) Find the specific r^* value so that the usual $1-\alpha$ confidence interval, $[\bar{Y}_n - z_{\alpha/2}\sigma/\sqrt{n}, \bar{Y}_n + z_{\alpha/2}\sigma/\sqrt{n}]$, minimizes the risk function in (a) among all procedures of the form $\delta_c(\mathbf{Y})$.

Answer: This corresponds to $c_{opt} = z_{\alpha/2}/\sqrt{n}$. The corresponding r value is given by

$$r^* = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-nc_{opt}^2/2} = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z_{\alpha/2})^2}{2}\right).$$

Note that this r value satisfies the assumption in (d). □

Problem 2 (Hypothesis Testing). A coin which has probability 1/3 or probability 1/2 of coming up heads (no other values are possible) is flipped once. Y is the number of heads obtained on that flip, i.e., Y will take one of two possible values: 0 or 1. The decision space $D = \{d_0, d_1\}$, where d_i is the decision "my guess is that the coin has probability $1/(2+i)$ of coming up heads." The loss is 1 for an incorrect decision, 0 for a correct decision. [This is a 2-decision, 2-state setting, often referred as "hypothesis testing" in statistics.]

(a) Specify S, Ω, D, and L (i.e., the sample space, the set of all possible distribution functions, the decision space, and the loss function).

Answer: $S = \{0, 1\}$ (tails or heads), $\Omega = \{1/3, 1/2\}, D = \{d_0, d_1\}$ and the loss function is

$$L(\theta, d_0) = \begin{cases} 1, & \text{if } \theta = 1/3 \\ 0, & \text{if } \theta = 1/2 \end{cases}; \qquad L(\theta, d_1) = \begin{cases} 0, & \text{if } \theta = 1/3 \\ 1, & \text{if } \theta = 1/2 \end{cases}.$$

□

(b) There are four possible (non-randomized) procedures:

$$\begin{aligned} \delta_1(0) = \delta_1(1) = d_0; &\qquad \delta_2(0) = \delta_2(1) = d_1; \\ \delta_3(0) = d_1, \quad \delta_3(1) = d_0; &\qquad \delta_4(0) = d_0, \quad \delta_4(1) = d_1. \end{aligned}$$

Show that for a given procedure δ, the risk function $R_\delta(\theta) = \mathbf{P}_\theta(\delta \text{ reaches wrong decision})$ and use this to find the risk function of each procedure. (There are only two possible values for θ : 1/3 or 1/2, and so R_δ can be thought of as a 2-vector for each δ.)

Answer: Here we have the sample loss function, and we proved in class that

$$R_\delta(\theta) = \mathbf{P}_\theta(\delta \text{ reaches wrong decision}) = \begin{cases} \mathbf{P}_{\theta=1/3}(\delta = d_0), & \text{if } \theta = 1/3 \\ \mathbf{P}_{\theta=1/2}(\delta = d_1), & \text{if } \theta = 1/2 \end{cases}.$$

For the procedure δ_1, we have

$$R_{\delta_1}(\theta) = \mathbf{P}_\theta(\delta_1 \text{ reaches wrong decision}) = \begin{cases} \mathbf{P}_{\theta=1/3}(\delta_1 = d_0) = 1, & \text{if } \theta = 1/3 \\ \mathbf{P}_{\theta=1/2}(\delta_1 = d_1) = 0, & \text{if } \theta = 1/2 \end{cases},$$

since δ_1 always makes a decision d_0. Similarly,

$$R_{\delta_2}(\theta) = \mathbf{P}_\theta(\delta_2 \text{ reaches wrong decision}) = \begin{cases} \mathbf{P}_{\theta=1/3}(\delta_2 = d_0) = 0, & \text{if } \theta = 1/3 \\ \mathbf{P}_{\theta=1/2}(\delta_2 = d_1) = 1, & \text{if } \theta = 1/2 \end{cases},$$

For the procedure δ_3, we have

$$R_{\delta_3}(\theta) = \begin{cases} \mathbf{P}_{\theta=1/3}(\delta_3 = d_0) = \mathbf{P}_{\theta=1/3}(Y = 1) = 1/3, & \text{if } \theta = 1/3 \\ \mathbf{P}_{\theta=1/2}(\delta_3 = d_1) = \mathbf{P}_{\theta=1/2}(Y = 0) = 1 - 1/2 = 1/2, & \text{if } \theta = 1/2 \end{cases},$$

where we use the definition of δ_3 in which $\{\delta_3 = d_0\}$ if and only if $Y = 1$. Similarly,

$$R_{\delta_4}(\theta) = \begin{cases} \mathbf{P}_{\theta=1/3}(\delta_4 = d_0) = \mathbf{P}_{\theta=1/3}(Y = 0) = 1 - 1/3 = 2/3, & \text{if } \theta = 1/3 \\ \mathbf{P}_{\theta=1/2}(\delta_4 = d_1) = \mathbf{P}_{\theta=1/2}(Y = 1) = 1/2, & \text{if } \theta = 1/2 \end{cases}.$$

In summary, the risk function can be summarized as follows: □

procedure	Risk Function $R_\delta(\theta)$	
	$\theta = 1/3$	$\theta = 1/2$
δ_1	1	0
δ_2	0	1
δ_3	1/3	1/2
δ_4	2/3	1/2

(c) Use the results of (b) to determine which of the four procedures among the nonrandomized procedures is (or are)

(i) admissible;

(ii) Bayes with respect to a prior distribution $\mathbf{P}_\pi(1/3) = 0.10 = 1 - \mathbf{P}_\pi(1/2)$;

(iii) Bayes with respect to a prior distribution $\mathbf{P}_\pi(1/3) = 3/5 = 1 - \mathbf{P}_\pi(1/2)$.

Note that there are only 4 procedures and only 2 values of θ when considering admissible or Bayes procedures.

Answer: (i) δ_1, δ_2 and δ_3 are admissible, and δ_4 is inadmissible, since δ_3 is better than δ_4. Using the definition that

$$r_\delta(\pi) = \sum_{\theta=1/3 \text{ or } 1/2} R_\delta(\theta)\mathbf{P}_\pi(\theta),$$

we can calculate the Bayes risk as follows:
Hence, δ_1 is Bayes in (ii) and both δ_2 and δ_3 are Bayes in (iii). □

(d) Is the procedure which was Bayes in (c)(ii) also Bayes relative to any other prior distributions (or laws)? If so, which?

procedure	Risk Function $R_\delta(\theta)$		Bayes risk $r_\delta(\pi)$	
	$\theta = 1/3$	$\theta = 1/2$	$\pi(1/3) = 0.10$	$\pi(1/3) = 3/5$
δ_1	1	0	0.10	0.60
δ_2	0	1	0.90	0.40
δ_3	1/3	1/2	0.4833	0.40
δ_4	2/3	1/2	0.5167	0.60

Answer: Yes, the procedure δ_1 is not only Bayes in (c)(ii), but also Bayes relative to other prior distributions. To see this, suppose that $\mathbf{P}_\pi(\frac{1}{3}) = \gamma = 1 - \mathbf{P}_\pi(\frac{1}{2})$. Then we have

$$\begin{aligned} h^*_\pi(y, d_0) &= \sum_{\theta=1/3 \text{ or } 1/2} L(\theta, d_0)\mathbf{P}_\theta(Y = y)\pi(\theta) \\ &= L(\frac{1}{3}, d_0)\mathbf{P}_{\theta=1/3}(Y = y)\pi(\frac{1}{3}) \\ &= (\frac{1}{3})^y(\frac{2}{3})^{1-y}\gamma, \end{aligned}$$

and

$$\begin{aligned} h^*_\pi(y, d_1) &= \sum_{\theta=1/3 \text{ or } 1/2} L(\theta, d_1)\mathbf{P}_\theta(Y = y)\pi(\theta) \\ &= L(\frac{1}{2}, d_1)\mathbf{P}_{\theta=1/2}(Y = y)\pi(\frac{1}{2}) \\ &= (\frac{1}{2})^y(\frac{1}{2})^{1-y}(1-\gamma) = \frac{1}{2}(1-\gamma), \end{aligned}$$

and

$$\frac{h^*_\pi(y, d_0)}{h^*_\pi(y, d_1)} = \frac{4\gamma}{3(1-\gamma)}(\frac{1}{2})^y.$$

Hence, any Bayes procedure must make decision d_0 if this ratio is < 1, make decision d_1 if the ratio is > 1, and can decide either d_0 or d_1 if the ratio is $= 1$. In particular, $\delta_1 \equiv d_0$ is Bayes iff the above ratio is ≤ 1 for both $y = 0$ and 1, or equivalently, if and only if

$$\frac{4\gamma}{3(1-\gamma)}(\frac{1}{2})^0 \leq 1 \text{ and } \frac{4\gamma}{3(1-\gamma)}(\frac{1}{2})^1 \leq 1, \qquad \text{i.e.,} \quad \gamma \leq \frac{3}{7} \text{ and } \gamma \leq \frac{3}{5},$$

which implies that $\gamma \leq 3/7$. In other words, δ_1 is also Bayes relative to a prior distribution (or law) $\mathbf{P}_\pi(\frac{1}{3}) = \gamma = 1 - \mathbf{P}_\pi(\frac{1}{2})$, as long as $0 \leq \gamma \leq 3/7$. □

Problem 3 (Hypothesis Testing). *Testing between simple hypotheses:* Suppose the sample space S is discrete; Ω consists of two possible probability functions of $\mathbf{Y}$, say $f_0(\mathbf{y})$ and $f_1(\mathbf{y})$; the decision space D consists of two elements d_0, and d_1; and the loss function

$$L(f_i, d_j) = \begin{cases} w_i, & \text{if } i \neq j, \\ 0, & \text{if } i = j, \end{cases}$$

where w_0 and w_1 are given positive numbers. Show that, for any prior distribution π for which $0 \leq \pi(f_i) \leq 1$, all (nonranomized) Bayes procedures are of the form

$$\delta(\mathbf{y}) = \left\{ \begin{array}{l} d_1, \\ d_1 \text{ or } d_0, \\ d_0, \end{array} \right\} \qquad \text{according as} \qquad \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} C,$$

where C is a constant (perhaps 0 or ∞) depending on the w_i's, $\pi(f_0)$ and $\pi(f_1)$, but not on $\mathbf{y}$.

[*Remark:* In the statistical literature, the corresponding Bayes procedure is often called as the **likelihood ratio test**, which is optimal in the frequentist setup in Neyman-Pearson lemma of minimizing Type II error probability subject to the Type I error probability constraint.]

Answer: In this problem, $\Omega = \{f_0, f_1\}, D = \{d_0, d_1\}$. Here we have

$$\begin{aligned} h^*_\pi(\mathbf{y}, d_0) &= \sum_{\theta = f_0, f_1} L(\theta, d_0)\mathbf{P}_\theta(\mathbf{Y} = \mathbf{y})\pi(\theta) \\ &= L(f_0, d_0)f_0(\mathbf{y})\pi(f_0) + L(f_1, d_0)f_1(\mathbf{y})\pi(f_1) \\ &= w_1 f_1(\mathbf{y})\pi(f_1) \end{aligned}$$

and

$$\begin{aligned} h^*_\pi(\mathbf{y}, d_1) &= L(f_0, d_1)f_0(\mathbf{y})\pi(f_0) + L(f_1, d_1)f_1(\mathbf{y})\pi(f_1) \\ &= w_0 f_0(\mathbf{y})\pi(f_0) \end{aligned}$$

and

$$\frac{h^*_\pi(\mathbf{y}, d_0)}{h^*_\pi(\mathbf{y}, d_1)} = \frac{w_1 f_1(\mathbf{y})\pi(f_1)}{w_0 f_0(\mathbf{y})\pi(f_0)} = \frac{w_1\pi(f_1)}{w_0\pi(f_0)}\frac{f_1(\mathbf{y})}{f_0(\mathbf{y})}$$

Now let

$$C = \frac{w_0\pi(f_0)}{w_1\pi(f_1)}$$

then this ratio is $<, =, > 1$ if and only if $\frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} <, =, > C$. The conclusion follows at once from the fact that any Bayes procedure must make decision d_0 if this ratio is < 1, make decision d_1 if the ratio is > 1, and can decide either d_0 or d_1 if the ratio is $= 1$. $\square$

Problem 4 (Point Estimation). Suppose Ω can be defined by the density functions $f_\theta(\cdot)$ according to the values of a real parameter θ, where $a \le \theta \le b$. The decisions are $D = \{d : a \le d \le b\}$, representing guesses as to the true value of θ. The loss function is $L(\theta, d) = |\theta - d|^r$, where r is a given positive value. The prior density on Ω is $\pi(\theta)$. Assume all $\pi(\theta)$'s or f_θ's are positive throughout the sample space S.

(a) Show that if $\pi(\theta|\mathbf{y})$ is the posterior density function of θ given that the observed data $\mathbf{Y} = \mathbf{y} \in S$, then a Bayes procedure is obtained by choosing $\delta(\mathbf{y}) = d'$ to minimize $\int_a^b |\theta - d'|^r \pi(\theta|\mathbf{y})d\theta$.

[It is OK to simply quote our in-class discussions of how to compute Bayes procedures. In any event, do *not* try to find a formula for the minimizing d' in this part (a).]

Answer: This follows at once from our discussion in class that a procedure δ is Bayes relative to π if and only if, for every $\mathbf{y}$, it assigns a decision $\delta(\mathbf{y})$ which minimizes (over D)

$$h^*_\pi(\mathbf{y}, d) = \int_\Omega L(\theta, d)p_\theta(\mathbf{y})\pi(\theta)d\theta,$$

or equivalently, to minimize

$$h_\pi(\mathbf{y}, d) = \int_\Omega L(\theta, d)\frac{p_\theta(\mathbf{y})\pi(\theta)}{m(\mathbf{y})}d\theta = \int_\Omega L(\theta, d)\pi(\theta|\mathbf{y})d\theta.$$

$\square$

(b) In particular, for "squared error loss" ($r = 2$), show that from (a) that a Bayes procedure is $\delta(\mathbf{y}) =$ mean of posterior law $\pi(\theta|\mathbf{y})$ of θ.

Answer: When $r = 2$, we have

$$\begin{aligned} h_\pi(\mathbf{y}, d) &= \int_a^b (\theta - d)^2 \pi(\theta|\mathbf{y}) d\theta \\ &= [\int_a^b \theta^2 \pi(\theta|\mathbf{y}) d\theta] - 2[\int_a^b \theta \pi(\theta|\mathbf{y}) d\theta] d + [\int_a^b \pi(\theta|\mathbf{y}) d\theta] d^2 \\ &= \mathbf{E}(\theta^2|\mathbf{y}) - 2\mathbf{E}(\theta|\mathbf{y}) d + d^2, \end{aligned}$$

which is minimized at $d = \mathbf{E}(\theta|\mathbf{y}) =$ mean of posterior law $\pi(\theta|\mathbf{y})$ of θ. □

(c) For $r = 1$ ("absolute error loss"), show that a Bayes procedure is obtained as any *median* (not necessarily unique!) of the posterior law of θ. Since the crucial result from probability theory used in demonstrating this may be unfamiliar, part of this problem is to prove it:

If g is a univariate probability density function with finite first moment, $\int_{-\infty}^{\infty} |\theta - c| g(\theta) d\theta$ is minimized if and only if c is a median of g.

[*Remark:* If you want, the hints for part (c) can be found on the last page of this homework.]

Answer: Suppose m is a median and c is not a median with $c > m$, then

$$|\theta - c| - |\theta - m| = \begin{cases} (c - \theta) - (m - \theta) = c - m, & \text{if } \theta \le m \\ (c - \theta) - (\theta - m) = c - 2\theta + m \ge m - c, & \text{if } m \le \theta \le c \\ (\theta - c) - (\theta - m) = m - c, & \text{if } \theta \ge c \end{cases},$$

which implies that $|\theta - c| - |\theta - m| \ge (c - m)\text{sgn}(m - \theta)$. Multiplying both sides by $g(\theta) \ge 0$ and integral over $\theta \in (-\infty, \infty)$, we have

$$\int_{-\infty}^{\infty} |\theta - c| g(\theta) d\theta - \int_{-\infty}^{\infty} |\theta - m| g(\theta) d\theta \ge (c - m) \int_{-\infty}^{\infty} \text{sgn}(m - \theta) g(\theta) d\theta.$$

Now since m is a median, we know

$$\int_{-\infty}^{m} g(\theta) d\theta = \frac{1}{2} = \int_{m}^{\infty} g(\theta) d\theta,$$

and thus

$$\int_{-\infty}^{\infty} \text{sgn}(m - \theta) g(\theta) d\theta = \int_{-\infty}^{m} g(\theta) d\theta + \int_{m}^{\infty} (-1) g(\theta) d\theta = 1/2 - 1/2 = 0.$$

Combining the above results, for all $c > m$,

$$\int_{-\infty}^{\infty} |\theta - c| g(\theta) d\theta \ge \int_{-\infty}^{\infty} |\theta - m| g(\theta) d\theta.$$

Similar arguments also hold for all $c < m$ (check it! when $c < m$, we also have $|\theta - c| - |\theta - m| \ge (c - m)\text{sgn}(m - \theta) = (m - c)\text{sgn}(\theta - m)$). Thus $\int_{-\infty}^{\infty} |\theta - c| g(\theta) d\theta$ is minimized if and only if c is a median of g.

When $r = 1$, we have

$$h_\pi(\mathbf{y}, d) = \int_a^b |\theta - d| \pi(\theta|\mathbf{y}) d\theta,$$

which is minimized at $d =$ any median of the posterior distribution $\pi(\theta|\mathbf{y})$ of θ. □

[*Hints for Problem 4 (c):* It suffices to show that if m is a median and c is not a median, then

$$\int_{-\infty}^{\infty} |\theta - c| g(\theta) d\theta - \int_{-\infty}^{\infty} |\theta - m| g(\theta) d\theta = \int_{-\infty}^{\infty} \Big(|\theta - c| - |\theta - m| \Big) g(\theta) d\theta \geq 0.$$

To prove this, assume for a moment that $c > m$, show (draw it!) that

$$\Big(|\theta - c| - |\theta - m| \Big) - (c - m) \mathrm{sign}(m - \theta) \geq 0,$$

where $\mathrm{sign}(u) = 1$ if $u > 0$; $= 0$ if $u = 0$; and $= -1$ if $u < 0$. Moreover, the "≥ 0" is "> 0" if $m < \theta < c$. The result can be proved by combining this and the fact that

$$\int_{-\infty}^{\infty} \mathrm{sign}(m - \theta) g(\theta) d\theta = \int_{-\infty}^{m} g(\theta) d\theta - \int_{m}^{\infty} g(\theta) d\theta = \frac{1}{2} - \frac{1}{2} = 0,$$

since m is median for (an absolutely continuous probability distribution with) probability density function g and thus $\int_{-\infty}^{m} g(\theta) d\theta = 1/2$.
Can you use the similar ideas to prove the case of $c < m$?]