Stat 982: Module 2, Homework 2

Due on November 1, 2022

Problem 1

Let $X_1, \ldots, X_n$ be iid binary random variables with $P(X_i = 1) = \theta \in (0, 1)$ and $P(X_i = 0) = 1 - \theta$, for $i = 1, \ldots, n$. Consider estimating θ with the squared error loss. Let $X = (X_1, \ldots, X_n)$. Find the risk functions of the following estimators:

(a) the non-randomized estimators $\overline{X}$ (the sample mean) and

$$T_0(X) = \begin{cases} 0 & \text{if more than half of the } X_i\text{'s are } 0, \\ 1 & \text{if more than half of the } X_i\text{'s are } 1, \\ 0.5 & \text{if exactly half of the } X_i\text{'s are } 0; \end{cases}$$

(b) the randomized estimators

$$T_1(X) = \begin{cases} \overline{X} & \text{with probability } 0.5, \\ T_0(X) & \text{with probability } 0.5, \end{cases}$$

and

$$T_2(X) = \begin{cases} \overline{X} & \text{with probability } \overline{X}, \\ 0.5 & \text{with probability } 1 - \overline{X}. \end{cases}$$

Problem 2

Find decision rules that are better than T_1 and T_2, respectively, in Problem 1, for $n \geq 3$. (Hint: Think about how a result similar to Lemma 101 can be established for a randomized decision rule. However, you do not need to establish any general result for solving this problem.)

Problem 3

Consider the two-state decision problem with $\Theta = \{1, 2\}$, P_1 the Bernoulli $(\frac{1}{4})$ distribution, P_2 the Bernoulli $(\frac{1}{2})$ distribution, $\mathcal{A} = \Theta$, and $L(\theta, a) = I[\theta \neq a]$.

(a) Find $\mathcal{S}^0$, the set of risk vectors (risk points) for the four non-randomized decision rules. Plot these risk points in the plane, and then sketch $\mathcal{S}$, the set of all randomized risk vectors.

(b) Identify $A(\mathcal{S})$, the set of all admissible risk vectors. Is there a minimal complete class for this decision problem? If there is one, what is it? (Note: It can be shown by direct, but tedious, calculation that any element of $\mathcal{D}^*$ has a corresponding element of $\mathcal{D}_*$ with the same risk vector, and vice versa. You may use this fact without proof.)

Problem 4

Consider again the two-state decision problem with $\Theta = \{1, 2\}$, P_1 the Bernoulli $(\frac{1}{4})$ distribution, P_2 the Bernoulli $(\frac{1}{2})$ distribution, $\mathcal{A} = \Theta$, and $L(\theta, a) = I[\theta \neq a]$.

(a) For each $p \in [0, 1]$, identify those risk vectors that are Bayes with respect to the prior $g = (p, 1 - p)$. For which priors are there more than one Bayes rule?

(b) Verify directly that the prescription "choose an action that minimizes the posterior expected loss" produces a Bayes rule with respect to the prior $g = (\frac{1}{2}, \frac{1}{2})$.

Problem 5

Consider a two-state decision problem where $\Theta = \mathcal{A} = \{0, 1\}$, P_0 and P_1 have densities f_0 and f_1, respectively, with respect to a dominating σ-finite measure μ, and the loss function is $L(\theta, a)$.

(a) For an arbitrary prior distribution G, find a formal Bayes rule with respect to G.

(b) Specialize your result from part (a) to the case where $L(\theta, a) = I[\theta \neq a]$. What connection does the form of these Bayes rules have to the theory of simple-versus-simple hypothesis testing?

Problem 6

Suppose that $X \sim$ Bernoulli (p) and that one wishes to estimate $p \in [0, 1]$ with the loss function $L(p, a) = |p - a|$. Consider the estimator δ with $\delta(0) = \frac{1}{4}$ and $\delta(1) = \frac{3}{4}$.

(a) Write out the risk function of δ and show that $R(p, \delta) \leq \frac{1}{4}$ for $p \in [0, 1]$.

(b) Show that there exists a prior distribution placing all its mass on $\{0, \frac{1}{2}, 1\}$ with respect to which δ is Bayes.

(c) Prove that δ is minimax in this problem and identify a least favorable prior.

We will use the notation $\mathcal{D} = \{\delta\}$ = the class of (non-randomized) decision rules It is technically useful to extend the notion of decision procedures to include the possibility of randomizing in various ways.

Definition 98. *If for each $x \in \mathcal{X}$, ϕ_x is a distribution on $(\mathcal{A}, \mathcal{E})$, then ϕ_x is called a **behavioral decision rule**.*

The notion of a behavioral decision rule is that one observes $X = x$ and then makes a random choice of an element of $\mathcal{A}$ using distribution ϕ_x. We'll let

$$\mathcal{D}^* = \{\phi_x\} = \text{the class of behavioral decision rules}$$

It's possible to think of $\mathcal{D}$ as a subset of $\mathcal{D}^*$ by associating with $\delta \in \mathcal{D}$ a behavioral decision rule ϕ_x^δ that is a point mass distribution on $\mathcal{A}$ concentrated at $\delta(x)$. The natural definition of the risk function of a behavioral decision rule is (abusing notation and using "R" here too)

$$R(\theta, \phi) = \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a)\, d\phi_x(a)\, dP_\theta(x)$$

A second (less intuitively appealing) notion of randomizing decisions is one that might somehow pick an element of $\mathcal{D}$ at random (and then plug in X). Let $\mathcal{F}$ be a σ-algebra on $\mathcal{D}$ that contains all singleton sets.

Definition 99. *A* ***randomized decision function (or rule)*** ψ *is a probability measure on* $(\mathcal{D}, \mathcal{F})$.

δ with distribution ψ becomes a random object. Let

$$\mathcal{D}_* = \{\psi\} = \text{the class of randomized decision rules}$$

It's possible to think of $\mathcal{D}$ as a subset of $\mathcal{D}_*$ by associating with $\delta \in \mathcal{D}$ a randomized decision rule ψ_δ placing mass 1 on δ. The natural definition of the risk function of a randomzied decision rule is (yet again abusing notation and using "R" here too)

$$R(\theta, \psi) = \int_{\mathcal{D}} R(\theta, \delta)\, d\psi(\delta) = \int_{\mathcal{D}} \int_{\mathcal{X}} L(\theta, \delta(x))\, dP_\theta(x)\, d\psi(\delta)$$

(assuming that $R(\theta, \delta)$ is properly measurable).

The behavioral decision rules are most natural, while the randomized decision rules are easiest to deal with in some proofs. So it is a reasonably important question when $\mathcal{D}^*$ and $\mathcal{D}_*$ are equivalent in the sense of generating the same set of risk functions. Some properly qualified version of the following is true.

Proposition 100. *If $\mathcal{A}$ is a complete separable metric space with $\mathcal{E}$ the Borel σ-algebra and ??? regarding the distributions P_θ and ???, then $\mathcal{D}^*$ and $\mathcal{D}_*$ are equivalent in terms of generating the same set of risk functions.*

$\mathcal{D}^*$ and $\mathcal{D}_*$ are clearly more complicated than $\mathcal{D}$. A sensible question is when they really provide anything $\mathcal{D}$ doesn't provide. One kind of negative answer can be given for the case of convex loss. The following is like 2.5a of Shao, page 151 of Schervish, page 40 of Berger, or page 78 of Ferguson.

Lemma 101. *Suppose that $\mathcal{A}$ is a convex subset of $\mathbb{R}^d$ and ϕ_x is a behavioral decision rule. Define a non-randomized decision rule by*

$$\delta(x) = \int_{\mathcal{A}} a\, d\phi_x(a)$$

(In the case that $d > 1$, interpret $\delta(x)$ as vector-valued, the integral as a vector of integrals over d coordinates of $a \in \mathcal{A}$.) Then

1. *if $L(\theta, \cdot) : \mathcal{A} \to [0, \infty)$ is convex, then*

$$R(\theta, \delta) \le R(\theta, \phi)$$

   and

2. *if $L(\theta, \cdot) : \mathcal{A} \to [0, \infty)$ is strictly convex, $R(\theta, \phi) < \infty$ and $P_\theta(\{x | \phi_x \text{ is non-degenerate}\}) > 0$, then*

$$R(\theta, \delta) < R(\theta, \phi)$$

Corollary 102. *Suppose that $\mathcal{A}$ is a convex subset of $\mathbb{R}^d$, ϕ_x is a behavioral decision rule and*

$$\delta(x) = \int_{\mathcal{A}} a\, d\phi_x(a)$$

Then

1. *if $L(\theta, a)$ is convex in a $\forall \theta$, δ is at least as good as ϕ,*

2. *if $L(\theta, a)$ is convex in a $\forall \theta$ and for some θ_0 the function $L(\theta_0, a)$ is strictly convex in a, $R(\theta_0, \phi) < \infty$ and $P_{\theta_0}(\{x | \phi_x \text{ is non-degenerate}\}) > 0$, then δ is better than ϕ.*

The corollary shows, e.g., that for squared error loss estimation, averaging out over non-trivial randomization in a behavioral decision rule will in fact improve that estimator.