Stat 982: Module 2, Homework 1

Due on October 11, 2022

Problem 1

Let $\mathcal{F}$ be a convex class of absolutely continuous distribution functions which contains all uniform densities. If $X = (X_1, ..., X_n)$ are iid $F \in \mathcal{F}$, then $T(X) = (X_{(1)}, ...X_{(n)})$ is a complete and sufficient statistic for $F \in \mathcal{F}$.

Problem 2

Suppose that $\Theta = \Theta_1 \times \Theta_2$ and that a decision rule $\phi \in \mathcal{D}^*$ is such that for each $\theta_2 \in \Theta_2$, ϕ is admissible when the parameter space is $\Theta_1 \times \{\theta_2\}$. Show that ϕ is then admissible when the parameter space is Θ. (Note: The result holds for any class of decision rules.)

Problem 3

Suppose that $w(\theta) > 0$ for any $\theta \in \Theta$. Show that $\phi \in \mathcal{D}^*$ is admissible with the loss function $L(\theta, a)$ if and only if ϕ is admissible with the loss function $w(\theta)L(\theta, a)$. (Note: The result holds for any class of decision rules.)

Problem 4

Consider estimation of $p \in [0, 1]$ with the squared error loss, based on $X \sim \text{Binomial}\,(n, p)$, and the two non-randomized decision rules $\delta_1(x) = \frac{x}{n}$ and $\delta_2(x) = \frac{1}{2}\left(\frac{x}{n} + \frac{1}{2}\right)$. Let ψ be a randomized decision rule that chooses δ_1 with probability $\frac{1}{2}$ and δ_2 with probability $\frac{1}{2}$.

(a) Write out expressions for the risk functions of δ_1, δ_2, and ψ.

(b) Find a behavioral rule ϕ that is risk equivalent to ψ (that is, ϕ and ψ have the same risk function).

(c) Identify a non-randomized estimator (decision rule) that is strictly better than ψ (or ϕ).

See the attached information on behavioral decision rules at the end of this assignment.

Problem 5

Give a counterexample to the following statement: If $\mathcal{C}_1$ and $\mathcal{C}_2$ are complete classes of a class of decision rules $\mathcal{H}$, then $\mathcal{C}_1 \bigcap \mathcal{C}_2$ is essentially complete for $\mathcal{H}$. (Note: The definitions and theorems in ISU Notes Section 3.3 still hold if you replace $\mathcal{D}^*$ there by any class of decision rules $\mathcal{H}$.)

Hint: Consider the following decision problem with $\mathcal{X} = \Theta = \{0, 1\}$, $\mathcal{A} = [0, 1]$, $L(\theta, a) = |\theta - a|$, and $P_0(X = 0) = P_1(X = 1) = 1$. Find complete classes $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathcal{H}$ that are disjoint, where $\mathcal{H} \subset \mathcal{D}$ is the subset of $\mathcal{D}$ (the class of non-randomized decision rules) that does not contain the (only) admissible decision rule $\delta(x) = x$ for $x = 0, 1$. Think about why we cannot use $\mathcal{D}$ or $\mathcal{D}^*$ in place of $\mathcal{H}$ here.

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*

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

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.