STAT 5010 Tutorial 3*

Nov. 2022

Definitions

1. Let X be a sample from a population $P \in \mathcal{P}$. A *statistical decision* is an *action* that we take after we observe X, for example, a conclusion about P or a characteristic of P. Throughout this section, we use $\mathbb{A}$ to denote the set of allowable actions. Let $\mathcal{F}_{\mathbb{A}}$ be a σ-field on $\mathbb{A}$. Then the measurable space $(\mathbb{A}, \mathcal{F}_{\mathbb{A}})$ is called the *action space*. Let $\mathcal{X}$ be the range of X and $\mathcal{F}_{\mathcal{X}}$ be a σ-field on $\mathcal{X}$. A *decision rule* is a measurable function (a statistic) T from $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$ to $(\mathbb{A}, \mathcal{F}_{\mathrm{A}})$. If a decision rule T is chosen, then we take the action $T(X) \in \mathbb{A}$ whence X is observed.

2. The construction or selection of decision rules cannot be done without any criterion about the performance of decision rules. In statistical decision theory, we set a criterion using a *loss function* L, which is a function from $\mathcal{P} \times \mathbb{A}$ to $[0, \infty)$ and is Borel on $(\mathbb{A}, \mathcal{F}_{\mathbb{A}})$ for each fixed $P \in \mathcal{P}$. If $X = x$ is observed and our decision rule is T, then our "loss" (in making a decision) is $L(P, T(x))$. The average loss for the decision rule T, which is called the risk of T, is defined to be
$$R_T(P) = E[L(P, T(X))] = \int_X L(P, T(x)) dP_X(x).$$

3. The loss and risk functions are denoted by $L(\theta, a)$ and $R_T(\theta)$ if $\mathcal{P}$ is a parametric family indexed by θ. A decision rule with small loss is preferred. But it is difficult to compare $L\left(P, T_1(X)\right)$ and $L\left(P, T_2(X)\right)$ for two decision rules, T_1 and T_2, since both of them are random. A rule T_1 is *as good as* another rule T_2 if and only if
$$R_{T_1}(P) \leq R_{T_2}(P) \quad \text{for any } P \in \mathcal{P},$$
and is *better* than T_2 if and only if the inequality above holds and $R_{T_1}(P) < R_{T_2}(P)$ for at least one $P \in \mathcal{P}$. Two decision rules T_1 and T_2 are *equivalent* if and only if $R_{T_1}(P) = R_{T_2}(P)$ for all $P \in \mathcal{P}$. If there is a decision rule T_* that is as good as any other rule in $\Im$, a class of allowable decision rules, then T_* is said to be *$\Im$-optimal* (or optimal if $\Im$ contains all possible rules).

4. Let $\Im$ be a class of decision rules (randomized or nonrandomized). A decision rule $T \in \Im$ is called *$\Im$-admissible* (or *admissible* when $\Im$ contains all possible rules) if and only if there does not exist any $S \in \Im$ that is better than T (in terms of the risk).

5. Now we consider an average of $R_T(P)$ over $P \in \mathcal{P}$:
$$r_T(\Pi) = \int_{\mathcal{P}} R_T(P) d\Pi(P),$$
where Π is a known probability measure on $(\mathcal{P}, \mathcal{F}_{\mathcal{P}})$ with an appropriate σ-field $\mathcal{F}_{\mathcal{P}}$. $r_T(\Pi)$ is called the *Bayes risk* of T w.r.t. Π. If $T_* \in \Im$ and $r_{T_*}(\Pi) \leq r_T(\Pi)$ for any $T \in \Im$, then T_* is called a *$\Im$-Bayes rule* (or *Bayes rule* when $\Im$ contains all possible rules) w.r.t. Π. The second method is to consider the worst situation, i.e., $\sup_{P \in \mathcal{P}} R_T(P)$. If $T_* \in \Im$ and $\sup_{P \in \mathcal{P}} R_{T_*}(P) \leq \sup_{P \in \mathcal{P}} R_T(P)$ for any $T \in \Im$, then T_* is called a *$\Im$-minimax* rule (or *minimax* rule when $\Im$ contains all possible rules).

6. A prior is called a *conjugate* prior if the posterior is in the same parametric family of distributions as that of the prior.

Propositions and Theorems

1. With squared error loss, the Bayes estimator is the mean of the posterior distribution.

2. Suppose that $\mathbb{A}$ is a convex subset of $\mathcal{R}^k$ and that for any $P \in \mathcal{P}, L(P, a)$ is a convex function of a. Let T be a sufficient statistic for $P \in \mathcal{P}, T_0 \in \mathcal{R}^k$ be a nonrandomized rule satisfying $E\left\|T_0\right\| < \infty$, and $T_1 = E\left[T_0(X) \mid T\right]$. Then $R_{T_1}(P) \leq R_{T_0}(P)$ for any $P \in \mathcal{P}$. If L is strictly convex in a and T_0 is not a function of T, then T_0 is inadmissible.

3. Suppose that $\mathbb{A}$ is a subset of $\mathcal{R}^k$. Let $T(X)$ be a sufficient statistic for $P \in \mathcal{P}$ and let δ_0 be a decision rule. Then
$$\delta_1(t, A) = E\left[\delta_0(X, A) \mid T = t\right],$$
which is a randomized decision rule depending only on T, is equivalent to δ_0 if $R_{\delta_0}(P) < \infty$ for any $P \in \mathcal{P}$.

*Dept. of Stat., CUHK. TA: YX

Question 1

1. Let $\bar{X}$ be the sample mean of a random sample of size n from $N\left(\theta, \sigma^2\right)$ with a known $\sigma > 0$ and an unknown $\theta \in \mathcal{R}$. Let $\pi(\theta)$ be a prior density with respect to a σ-finite measure ν on $\mathcal{R}$.

   Show that the posterior mean of θ, given $\bar{X} = x$, is of the form

$$\delta(x) = x + \frac{\sigma^2}{n} \frac{d \log(p(x))}{dx},$$

   where $p(x)$ is the marginal density of $\bar{X}$, unconditional on θ.

Solution:

Note that $\bar{X}$ has distribution $N\left(\theta, \sigma^2/n\right)$. The product of the density of $\bar{X}$ and $\pi(\theta)$ is

$$\frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(x-\theta)^2/\left(2\sigma^2\right)} \pi(\theta).$$

Hence,

$$p(x) = \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(x-\theta)^2/\left(2\sigma^2\right)} \pi(\theta) d\nu$$

and

$$p'(x) = \frac{n}{\sigma^2} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} (\theta - x) e^{-n(x-\theta)^2/\left(2\sigma^2\right)} \pi(\theta) d\nu.$$

Then, the posterior mean is

$$\begin{aligned}
\delta(x) &= \frac{1}{p(x)} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \theta e^{-n(x-\theta)^2/\left(2\sigma^2\right)} \pi(\theta) d\nu \\
&= x + \frac{1}{p(x)} \int \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} (\theta - x) e^{-n(x-\theta)^2/\left(2\sigma^2\right)} \pi(\theta) d\nu \\
&= x + \frac{\sigma^2}{n} \frac{p'(x)}{p(x)} \\
&= x + \frac{\sigma^2}{n} \frac{d \log(p(x))}{dx}.
\end{aligned}$$

Question 2

1. Prove Proposition 3.

Solution:

Proof. Note that δ_1 is a decision rule since δ_1 does not depend on the unknown P by the sufficiency of T.

$$\begin{aligned}
R_{\delta_1}(P) &= E\left\{\int_{\mathrm{A}} L(P, a) d\delta_1(X, a)\right\} \\
&= E\left\{E\left[\int_{\mathbb{A}} L(P, a) d\delta_0(X, a) \mid T\right]\right\} \\
&= E\left\{\int_{\mathbb{A}} L(P, a) d\delta_0(X, a)\right\} \\
&= R_{\delta_0}(P)
\end{aligned}$$

■