STAT-510		2017 Fall
	Lecture 7: Decision	
Instructor: Xiaohui Chen	Scribe: Yutong Dai	Last Modified: 2018-02-25

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

7.1 Basic Setup

Statistical Model	$(X, \mathcal{X}, P)$,where $P = \{p_\theta; \theta \in \Theta\}$.
Action Space	$\mathcal{A}$; A set specifies the possible "actions" we might take. The element is denoted as a. Usually let it be identical to Θ.
Decision Procedure	$\delta(x) : \mathcal{X}\mapsto \mathcal{A}$; A function specifies which action to take for each possible value of the data.
Loss Function	$L(a, \theta) : \mathcal{A} \times \Theta \mapsto [0, +\infty]$; A function measures the equality of a paticular action.
Risk Function	$R(\theta; \delta) = E_\theta[L(\delta(X), \theta)] = E[L(\delta(X), \theta)\|\Theta = \theta]$. Meaure the given decision procedure averaing on all possbile test points. The two expectations are the same, but the first is written for frequentists and the second for Bayesian.
Bayes Risk	$R(\pi; \theta) = E_\pi[R(\theta; \delta)]$; Average the risk over all possible θ, where the prior is given by $\theta \sim \pi(\theta)$.

Definition 7.1 (Bayes Procedures) For given risk function $R(\theta; \delta)$, set of decision procedures D and θ's prior distribution π defined on Θ, then a Bayes procedure w.r.t D and π is a procedure $\delta_\pi \in D$ such that

$$\delta_\pi = \arg\min_{\delta \in D} E_\pi[R(\theta; \delta)]$$

Theorem 7.2 The δ_π can be found by solving the

$$\delta_\pi = \arg\min_{\delta \in D} E_{\pi|x}[L(\delta(x), \theta)|X = x],$$

for each $x \in \mathcal{X}$, where

$$E_{\pi|x}[L(\delta(x), \theta)|X = x] = \int_\theta \pi(\theta|x)L(\delta(x), \theta)d\theta.$$

Remark: $E_{\pi|x}[L(\delta(x), \theta)|X = x]$ is caled posterior expected loss.

Example 7.3 Assume the $L(a, \theta) = (a - \theta)^2$ and $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1), \mu \in R$. We wish to estimate μ. For constants c, d, define $\delta(x) = c + d\bar{x}_n$, where $\bar{x}_n$ is the sample mean.

Since $\delta(x) \sim N(c + d\mu, 1/n)$, the risk of $\delta(x)$ is given by

$$\begin{aligned} R(\delta(x), \mu) &= E_\theta[\delta(x) - \mu]^2 = \mathrm{Var}[\delta(x)] + [E\delta(x) - \mu]^2 \\ &= d^2 \frac{1}{n} + [c + (d-1)\mu]^2. \end{aligned} \tag{7.1}$$

If the prior of μ is $N(\mu_0, \sigma_0^2)$, theb the Bayes risk for the given $\delta(x)$ is given by

$$\begin{aligned} E_\pi[R(\delta(x), \mu)] &= E_\pi[d^2 \frac{1}{n} + [c + (d-1)\mu]^2] \\ &= \frac{d^2}{n} + c^2 + 2c(d-1)E\mu + (d-1)^2 E\mu^2 \\ &= (d-1)^2(\sigma_0^2 + \mu_0^2) + 2(d-1)c\mu_0 + c^2 + d/n. \end{aligned} \tag{7.2}$$

To find the Bayes procedure, we have to options.

Option 1:

$$\begin{cases} \frac{\partial}{\partial c} E_\pi[R(\delta(x), \mu)] = 0 \\ \frac{\partial}{\partial d} E_\pi[R(\delta(x), \mu)] = 0 \end{cases},$$

then $c = \frac{\mu_0}{1+n\sigma_0^2}, d = \frac{n\sigma_0^2}{1+n\sigma_0^2}$ and the Bayes estimator is $\delta(x) = \frac{n\sigma_0^2 \bar{x}_n + \mu_0}{1+n\sigma_0^2}$.

Option 2: Since $\mu \sim N(\mu_0, \sigma_0^2)$, then $\mu | x \sim N(\frac{n\sigma_0^2 \bar{x}_n + \mu_0}{1+n\sigma_0^2}, \frac{\sigma_0^2}{1+n\sigma_0^2})$. The posterior expected loss is given by

$$\begin{aligned} E_{\pi|x}[R(\delta(x), \mu) | X = x] &= E_{\pi|x}[(\delta(x) - \mu) | X = x]^2 \\ &= \delta(x)^2 - 2E_{\pi|x}[\mu] + E_{\pi|x}\mu^2 \\ &= (c + d\bar{x}_n - \frac{n\sigma_0^2 \bar{x}_n + \mu_0}{1 + n\sigma_0^2})^2 + \frac{\sigma_0^2}{1 + n\sigma_0^2}, \end{aligned} \tag{7.3}$$

which implies $\delta(x) = \frac{n\sigma_0^2 \bar{x}_n + \mu_0}{1+n\sigma_0^2}$.

7.2 Admissibility

Definition 7.4 (Dominate) The suppose D is a set of procedures. For $\delta_1, \delta_2 \in D$, the δ_1 is said dominate δ_2 if

$$\forall \theta \in \Theta, R(\delta_1, \theta) \leq R(\delta_2, \theta)$$

and

$$\exists \theta_0 \in \Theta, R(\delta_1, \theta_0) < R(\delta_2, \theta_0).$$

Definition 7.5 (admisibility) The suppose D is a set of procedures. A $\delta_\in D$, the δ is said inadmissible if $\exists \delta' \in D$ such that δ' domintae δ. If there no such δ', then δ is admissible among procedures in D.

Example 7.6

- If D is a set of unbiased estimators and the loss is squared loss, then the UMVUE is admissible among D.

- If D is a set of shift-equivariant estimators and the loss is squared loss, then Pitman estimator is admissible among D.

Corollary 7.7 $\delta \in D$ is admissible if

$$R(\delta', \theta) \le R(\delta, \theta) \text{ for all } \theta \in \Theta \Rightarrow R(\delta', \theta) = R(\delta, \theta) \text{ for all } \theta \in \Theta.$$

Lemma 7.8 A Bayes procedure δ_π w.r.t D and π is admissible if any of the following hold:

1. It is admissible among D_π, the set of estimators that are Bayes Procedures w.r.t D and π;

2. It is the unique Bayes procedure. That is, if δ'_π is another Bayer Procedure, then $R(\delta_\pi, \theta) = R(\delta'_\pi, \theta)$ for all $\theta \in \Theta$.

Proof:

1. Suppose $\delta' \in D$ satisfies $\forall \theta \in \Theta, R(\delta', \theta) \le R(\delta_\pi, \theta)$, then taking expectation w.r.t π, one have

$$E_\pi[R(\delta', \theta)] \le E_\pi[R(\delta_\pi, \theta)].$$

By definition, then δ' is also a Bayes procedure w.r.t π and D. Therefore, by the admissibility, $\forall \theta \in \Theta, R(\delta', \theta) = R(\delta_\pi, \theta)$, which proves δ_π is admissible (Corollary7.7).

$\forall \delta \in D$ and $\delta \ne \delta_\pi$, by the uniqueness and the definition of Bayes procedure, there exits $\theta^* \in \Theta$ such that

$$R(\delta_\pi, \theta^*) < R(\delta, \theta^*),$$

which implies that δ cannot domintae δ_π. ■

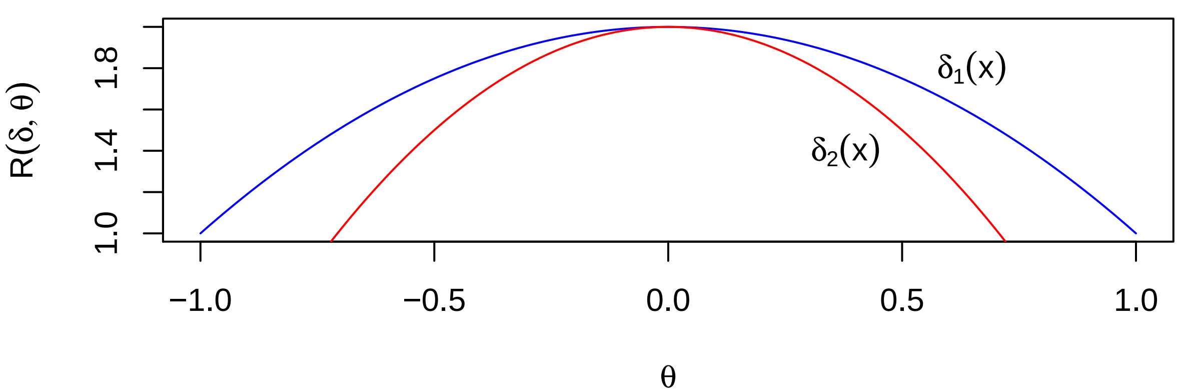
7.3 Minimax Procedure

Using a Bayes procedure involves choosing a prior π. One attempt to objectifying the choice of a procedure is for each procedure, see what its worst risk is. Then you choose the procedure that has the best worst, i.e., the minimax procedure. Next is the formal definition.

Definition 7.9 Let D be a set of decision procedures. A $\delta \in D$ is minimax among D if for any other δ',

$$\sup_{\theta \in \Theta} R(\delta, \theta) \le \sup_{\theta \in \Theta} R(\delta', \theta).$$

Remark: By the definition, minimax procedures are not necessarily unique. For example, if δ_1 is a minimax, and we have δ_2 dominates δ_1, then δ_2 is also a minimax (See below).



Lemma 7.10 If there exists a unique minmax procedure, then the esatimator is admissible.

Proof: Let δ^* be the only minimax procedure, then by the definition we have for all $\delta \in D$,

$$\sup_{\theta\in\Theta} R(\delta^*, \theta) < \sup_{\theta\in\Theta} R(\delta, \theta),$$

which implies δ^* is admissible. ■

Remark: If a minimax estimator is inadmissible, then it is not unique.

Lemma 7.11 Suppose δ_0 has a finite constant risk,

$$R(\delta_0, \theta) = c < +\infty \quad \forall\theta \in \Theta,$$

then δ_0 is minimax if it satisfies any of the following

1. It is a Bayes Procedure w.r.t a prior π.

2. It is admissible.

Proof:

1. If δ_0 is not the minimax, then there exists a δ' such that

$$\sup_{\theta\in\Theta} R(\delta', \theta) < \sup_{\theta\in\Theta} R(\delta_0, \theta) = c,$$

whcih implies

$$E_\pi[R(\delta', \theta)] < E_\pi[R(\delta_0, \theta)] = c.$$

This means δ_0 is not the Bayes procedure, hence contradition.

2. If δ_0 is not the minimax, then there exists a δ' such that $\sup_{\theta\in\Theta} R(\delta', \theta) < \sup_{\theta\in\Theta} R(\delta_0, \theta) = c$, and since $R(\delta_0, \theta) = c$, we know that δ' dominate δ_0,hence contradiction.

■