STATS 300A: Theory of Statistics I **Autumn 2016/17**

Lecture 11: November 3

Lecturer: Joseph Romano *Scribes: Kenneth Tay*

11.1 Minimax Estimators

For a prior Λ, let r_Λ be the Bayes risk of the Bayes estimator δ_Λ w.r.t. Λ_n.

Definition 11.1 *A sequence of priors $\{\Lambda_m\}$ on a parameter space Θ is said to be **least favorable** if $r_{\Lambda_m} \to r$ as $m \to \infty$, and for any prior Λ', $r_{\Lambda'} \leq r$.*

Theorem 11.2 *Suppose that we have an estimator δ and a sequence of priors $\{\Lambda_m\}$ such that $\sup_\theta R(\theta, \delta) = r$ and $r_{\Lambda_m} \to r$. Then*

1. *δ is minimax, and*
2. *$\{\Lambda_m\}$ is least favorable.*

Proof:

1. Let δ' be any other estimator. Then
$$\begin{aligned} \sup R(\theta, \delta') &\geq \int R(\theta, \delta') d\Lambda_m(\theta) \\ &\geq \int R(\theta, \delta_{\Lambda_m}) d\Lambda_m(\theta) \\ &= r_{\Lambda_m} \end{aligned}$$
for all m. Taking $m \to \infty$, we get $\sup R(\theta, \delta') \geq r$. Thus, δ is minimax.

2. If Λ' is any prior, then
$$\begin{aligned} r_{\Lambda'} &= \int R(\theta, \delta_{\Lambda'}) d\Lambda'(\theta) \\ &\leq \int R(\theta, \delta) d\Lambda'(\theta) \\ &\leq \sup R(\theta, \delta) \\ &= r. \end{aligned}$$
Thus $\{\Lambda_m\}$ is least favorable.

$\blacksquare$

The following lemma gives us a way to find minimax estimators for a larger family (by finding minimax estimators from a smaller family satisfying a certain condition):

Lemma 11.3 *Suppose X has an unknown distribution F from a family $\mathcal{F}$, and we wish to estimate $g(F)$. Suppose δ is minimax when estimating $g(F)$ for $F \in \mathcal{F}_0$, where $\mathcal{F}_0 \subseteq \mathcal{F}$ is some submodel of $\mathcal{F}$. If*

$$\sup_{F \in \mathcal{F}_0} R(\delta, F) = \sup_{F \in \mathcal{F}} R(\delta, F),$$

then δ is minimax for $\mathcal{F}$.

11.1.1 Example: Normal setting

Let $X_1, \ldots, X_n \sim \mathcal{N}(\theta, \sigma^2)$, X_i's independent. To estimate θ with squared error loss.

1. **Case 1:** σ^2 known.

   Take $\Lambda_m = \mathcal{N}(0, b_m^2)$ with $b_m \to \infty$ as $m \to \infty$. In previous lectures, we calculated the Bayes risk of the Bayes estimator to be

   $$r_{\Lambda_m} = \left(\frac{n}{\sigma^2} + \frac{1}{b_m} \right)^{-1}.$$

   As $m \to \infty$, $r_{\Lambda_m} \to \dfrac{\sigma^2}{n}$. On the other hand, we know that $R(\theta, \bar{X}) = \dfrac{\sigma^2}{n}$, so by Theorem 11.2, $\bar{X}$ is minimax.

2. **Case 2:** σ^2 unknown.

   In this case,

   $$\sup_{\theta, \sigma^2} R((\theta, \sigma^2), \bar{X}) = \sup_{\theta, \sigma^2} \frac{\sigma^2}{n} = \infty.$$

   In fact, every estimator has sup risk $= \infty$.

3. **Case 3:** σ^2 unknown, but $\sigma^2 \leq M$ for some known $M < \infty$.

   Considering the subfamily $\mathcal{F}_0 = \{\mathcal{N}(\theta, M)\}$, we can use Case 1 to conclude that $\bar{X}$ is minimax for $\mathcal{F}_0$, and $\sup_{F \in \mathcal{F}_0} R(\theta, \bar{X}) = \frac{M}{n}$. Since

   $$\sup_{\theta, \sigma^2} R((\theta, \sigma^2), \bar{X}) = \sup_{\theta, \sigma^2} \frac{\sigma^2}{n} = \frac{M}{n} = \sup_{F \in \mathcal{F}_0} R(\theta, \bar{X}),$$

   we can use Lemma 11.3 to conclude that $\bar{X}$ is minimax.

11.1.2 Example: Non-parametric setting

Let $X_1, \ldots, X_n \sim F$, X_i's independent. F unknown, Var$F = \sigma_F^2$ unknown. To estimate $g(F) = \int x dF(x) = \mathbb{E}_F X_i$.

1. **Case 1:** $\sigma_F^2 \leq M$, where M is known.

   In this case, we can set

   $$\mathcal{F} = \{\text{all CDFs with } \sigma_F^2 \leq M\},$$
   $$\mathcal{F}_0 = \{N(\theta, M), -\infty < \theta < \infty\}.$$

   Using Lemma 11.3, we can conclude that $\bar{X}$ is minimax.

2. **Case 2:** $a \le X_i \le b$, where a and b are known.

   Without loss of generality, assume that $a = 0$ and $b = 1$.

   Recall that in a previous lecture, we proved that if X_i's are supported on $\{0, 1\}$, then the minimax estimator is
$$\delta = \frac{\sqrt{n}}{1+\sqrt{n}}\bar{X} + \frac{1}{2(1+\sqrt{n})}.$$
   Let $\theta = \mathbb{E}_F X_i$, and define $\mathcal{F}_0 = \{F : F \text{ supported on } \{0,1\}\}$. We now claim that for all F,
$$\mathbb{E}_F[(\delta - \theta)^2] \le \frac{1}{4(1+\sqrt{n})^2},$$
   with equality iff $F \in \mathcal{F}_0$. Once we show this, we can use Lemma 11.3 to conclude that δ is minimax.

   Proof: Using the fact that $X_i^2 \le X_i$ for $0 \le X_i \le 1$,
$$\begin{aligned}
\mathbb{E}_F[(\delta-\theta)^2] &= \mathbb{E}\left[\left(\frac{\sqrt{n}}{1+\sqrt{n}}(\bar{X}-\theta) + \left(\frac{\sqrt{n}}{1+\sqrt{n}} - 1\right)\theta + \frac{1}{2(1+\sqrt{n})}\right)^2\right] \\
&= \left(\frac{1}{1+\sqrt{n}}\right)^2 \mathbb{E}_F\left[\left[\sqrt{n}(\bar{X}-\theta) + \left(\frac{1}{2} - t\right)\right]^2\right] \\
&= \left(\frac{1}{1+\sqrt{n}}\right)^2 \left[\mathrm{Var}_F X_i + \left(\frac{1}{2} - t\right)^2\right] \\
&= \left(\frac{1}{1+\sqrt{n}}\right)^2 \left[\mathbb{E}_F X_i^2 - \theta^2 + \left(\frac{1}{2} - t\right)^2\right] \\
&\le \left(\frac{1}{1+\sqrt{n}}\right)^2 \left[\theta - \theta^2 + \left(\frac{1}{2} - t\right)^2\right] \\
&= \frac{1}{4(1+\sqrt{n})^2},
\end{aligned}$$
   with equality when $\mathbb{E}_F X_i = \mathbb{E}_F X_i^2$, i.e. F supported on $\{0, 1\}$. ∎

11.1.3 Example: Minimax but inadmissible

Let $X_1, \dots, X_n \sim \mathcal{N}(\theta, 1)$, X_i's independent. Suppose further that $\Theta = \{\theta : \theta \ge 0\}$.

Note that $\delta = \bar{X}$ is inadmissible: the estimator $\delta' = \max(\bar{X}, 0)$ dominates it.

However, $\bar{X}$ is still minimax! (As $\theta \to \infty$, $R(\theta, \delta') \nearrow R(\theta, \delta)$.)

11.1.4 Example: Two densities, 0-1 loss

Assume that X has either density f_0 or f_1, and that we have 0-1 loss. Assume further that $\dfrac{f_1(X)}{f_0(X)}$ has a continuous distribution under f_0 and f_1.

In a previous lecture, we calculated the Bayes estimator for this set-up. If the prior has mass π for $\theta = 1$

and $1-\pi$ for $\theta = 0$, then the Bayes estimator is:

$$\delta(x) = \begin{cases} 1 & \text{if } \dfrac{\pi f_1(x)}{(1-\pi)f_0(x)} > 1, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the Bayes estimator has the form "choose $\theta = 1$ if $\frac{f_1(X)}{f_0(X)} > c$ for some constant c".

If we can find some π (or equivalently, some c) such that the Bayes estimator has constant risk, then it is minimax as well. Under 0-1 loss, we have

$$\text{risk if } f_0 \text{ is true} = P_{f_0}\left\{\frac{f_1(X)}{f_0(X)} > c\right\},$$
$$\text{risk if } f_1 \text{ is true} = P_{f_1}\left\{\frac{f_1(X)}{f_0(X)} < c.\right\}$$

Thus, the Bayes estimator is minimax if c is such that

$$P_{f_0}\left\{\frac{f_1(X)}{f_0(X)} > c\right\} = P_{f_1}\left\{\frac{f_1(X)}{f_0(X)} < c\right\}.$$

As c goes from $-\infty$ to ∞, the LHS decreases from 1 to 0 while the RHS increases from 0 to 1. By our continuity assumption, such a c always exists.

11.2 Hypothesis Testing

Assume that X has either density f_0 or f_1. We can set up null and alternative hypotheses:

$$H_0 : X \sim f_0,$$
$$H_1 : X \sim f_1.$$

Our goal is to find a decision rule which guesses which distribution X has well. However, in this set-up we treat errors differently: choosing f_1 when the distribution is actually f_0 is not viewed with the same weight as choosing f_0 when the distribution is actually f_1.

(This is different from the example in the previous section, where both errors are treated the same, and we are trying to minimize the probability of being wrong.)

The table below maps out the possible scenarios:

	H_0 is true	H_1 is true
Reject H_0	Type 1 error	Good decision
Don't reject H_0	Good decision	Type 2 error

In general, we are more concerned with Type 1 error, and so, we want to find a decision rule such that, for some pre-specified α,

$$P_{f_0}(\text{commit a Type 1 error}) \leq \alpha.$$

Subject to this constraint above, we want to find a decision rule such that $P_{f_1}(\text{reject } H_0)$ is as large as possible. This is called the **power** of the test/decision rule.

Definition 11.4 *A* ***test function*** $\varphi(X)$ *is a function taking values in* $[0,1]$ *with the interpretation that*

$$\begin{cases} \varphi(x) = 1 & \textit{reject } H_0, \\ \varphi(x) \in (0,1) & \textit{reject } H_0 \textit{ with probability } \varphi(X), \\ \varphi(x) = 0 & \textit{don't reject } H_0. \end{cases}$$

We can restate the problem of hypothesis testing:

Choose $\varphi = \varphi(X)$ to maximize $\mathbb{E}_{f_1}\varphi(X)$ subject to the constraint $\mathbb{E}_{f_0}\varphi(X) \leq \alpha$.

The Neyman-Pearson Lemma tells us how to do this:

Lemma 11.5 (Neyman-Pearson Lemma) *A most powerful test which solves the problem above has the form*

$$\varphi(X) = \begin{cases} 1 & \textit{if } \frac{f_1(X)}{f_0(X)} > c, \\ \gamma & \textit{if } \frac{f_1(X)}{f_0(X)} = c, \\ 0 & \textit{if } \frac{f_1(X)}{f_0(X)} < c, \end{cases}$$

where c *and* γ *satisfy* $\mathbb{E}_{f_0}\varphi(X) = \alpha$.