STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: Qinghua (Devon) Ding and Dongrong Li

Lecture 9

1 Minimax Estimators and Worst-Case Optimality

Given $X \sim P_\theta$, where $\theta \in \Omega$, and a loss function $L(\theta, d)$, we want to minimize the maximum risk: $\sup_{\theta\in\Omega} R(\theta,\delta)$, this minimizer is known as a minimax estimator.

Recall the definition of Bayes risk under an arbitrary prior distribution Λ:

$$r_\Lambda = \inf_\delta r(\Lambda,\delta) = \inf_\delta \int_{\theta\in\Omega} R(\theta,\delta)d\Lambda(\theta)$$

Definition 1. *A prior distribution is said to be a least favorable prior if $r_\Lambda \geq r_{\Lambda'}$, for any other prior distribtion Λ'.*

Following the definition is the theorem:

Theorem 2 (TPE 5.1.4)**.** *Suppose δ_Λ is Bayes for Λ with*

$$r_\Lambda = \sup_\theta R(\theta, \delta_\Lambda)$$

i.e. the Bayes risk of δ_Λ is the maximum risk of δ_Λ, then:

(i) δ_Λ is minimax,

(ii) Λ is a least favorable prior,

(iii) If δ_Λ is the unique Bayes estimator for Λ almost surely, for all P_θ, then it is a unique minimax estimator.

Proof. (i) Let δ be any other estimator, then we have that:

$$\sup_{\theta\in\Omega} R(\theta,\delta) \geq \int R(\theta,\delta)d\Lambda(\theta) \overset{(*)}{\geq} \int R(\theta,\delta)d\Lambda(\theta)$$

This implies that δ_Λ is minimax.

(ii) If $\delta\Lambda$ is the unique Bayes estimator, then the ineqality above $(*)$ is strict for $\delta \neq \delta_\Lambda$, which implies that δ_Λ is the unique minimax.

(iii) Let Λ' be any other prior distribution, then

$$\begin{aligned} r_{\Lambda'} &\leq \inf_\delta \int R(\theta,\delta)d\Lambda'(\theta) \leq \int R(\theta,\delta_\Lambda)d\Lambda'(\theta) \\ &\leq \sup_\theta R(\theta,\delta_\Lambda) = r_\Lambda \end{aligned}$$

Since the worst case risk of δ_Λ is its Bayes risk over Λ, we know that Λ is a least favorable prior distribution.

□

An implication is that we can find a minimax estimator by finding a Bayes estimator with Bayes risk equals its maximum risk, which gives the following corollary:

Corollary 3 (TPE 5.1.5)**.** *If a Bayes estimator of δ_Λ has constant risk, i.e. $R(\theta, \delta_\Lambda) = R(\theta', \delta_\Lambda)$ for any $\theta, \theta' \in \Omega$, then δ_Λ is minimax.*

An implication of this corollary is that, if a Bayes estimator has constant risk, it is minimax too. We may find a prior support set ω such that $\Lambda(\omega) = 1$ and for which $R(\theta, \delta_\Lambda)$ is maximum for any $\theta \in \Omega$.

Corollary 4 (TPE 5.1.6)**.** *Define $\omega_\Lambda = \{\theta : R(\theta, \delta_\Lambda) = \sup_{\theta'} R(\theta', \delta_\Lambda)\}$. A Bayesian estimator δ_Λ is minimax if $\Lambda(\omega_\Lambda) = 1$.*

Example 1. *Suppose $X \sim Binomial(n, \theta)$ for some $\theta \in (0, 1)$ and we adopt the squared loss function, is $\frac{x}{n}$ minimax?*

Notice that the corresponding risk is $R(\theta, \frac{x}{n}) = \frac{\theta(1-\theta)}{n}$. Observe that the risk has a unique maximum at $\theta = \frac{1}{2}$. The worst risk is:

$$\sup_{\theta \in \Omega} R(\theta, \frac{x}{n}) = R(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4n}$$

In this case, [TPE 5.1.6] is not helpful because if $\Lambda(\{\frac{1}{2}\}) = 1$, then $\delta_\Lambda(X) = \frac{1}{2} \neq \frac{x}{n}$.

However, [TPE 5.1.5] can be helpful instead. To find a minimax estimator, we will need to search for a prior such that the Bayes estimator has constant risk.

Recall that if the prior is $Beta(\alpha, \beta)$, the Bayes estimator under the squared loss is:

$$\delta_{\alpha,\beta}(X) = \frac{x + \alpha}{n + \alpha + \beta}$$

for any α, β.

$$\begin{aligned} R(\theta, \delta_{\alpha,\beta}) &= \mathbb{E}_\theta \left(\left\{ \frac{x + \alpha}{n + \alpha + \beta} - \theta \right\}^2 \right) \\ &= \frac{1}{(n + \alpha + \beta)^2} \mathbb{E}_\theta \left(\{x - n\theta - \alpha(\theta - 1) - \theta\beta\}^2 \right) \\ &= \frac{1}{(n + \alpha + \beta)^2} \left[n\theta(1 - \theta + \{\alpha(\theta - 1) + \theta\beta\}^2) \right] \end{aligned}$$

To eliminate the θ dependence in $R(\theta, \delta_{\alpha,\beta})$, we need to set the coefficients of θ^2 and θ be zero, that is:

$$\begin{aligned} -n + (\alpha + \beta)^2 &= 0 \\ n - 2\alpha(\alpha + \beta) &= 0, \end{aligned}$$

which solves $\alpha = \beta = \frac{\sqrt{n}}{2}$. The Bayes estimator $\delta_{\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}}(X) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}$ is minimax (TPE 5.1.4) with constant risk of $\frac{1}{4(\sqrt{n}+1)^2}$, we can conclude that $\frac{X}{n}$ is not minimax.

2 Generalization of Minimax-Bayes Theorems

We remark that minimax estimators may not be Bayes estimators. This is illustrated in the following example.

Example 2 (minimax for normal with unknown mean θ)**.** *Let $X_1, X_2, ..., X_n \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ with σ^2 unknown. Our goal is to estimate θ under the squared loss function. Our candidate is $\overline{X}$, which has constant risk $R(\theta, \overline{X}) = \mathbb{E}_\theta[(\overline{X} - \theta)^2] = \frac{\sigma^2}{n}$. This suggests that $\overline{X}$ can be a minimax estimator (TPE 5.1.4 and 5.1.5). However, $\overline{X}$ is not Bayes for any prior (Example 5, Lecture 8 and TPE 4.2.3).*

We also recall *TPE 4.2.3* here.

Theorem 5 (TPE 4.2.3)**.** *Unbiased estimators are Bayes only in the degenerate case of zero risk, i.e.,*

$$\mathbb{E}_{\Theta,X}\left[\{\delta(X) - g(\Theta)\}^2\right] = 0.$$

Thus we cannot yet conclude that $\overline{X}$ is minimax. We now consider the family of estimators with the form $\delta_{\omega,\mu_0}(X) = \omega\overline{X} + (1-\omega)\mu_0$, where $\omega \in (0,1)$ and $\mu_0 \in \mathbb{R}$. However, the worst case risk for this family of estimators is infinite.

$$\begin{aligned}
\sup_\theta \ \mathbb{E}_\theta\left[(\theta - \delta_{\omega,\mu_0}(X))^2\right] &= \sup_\theta \ \mathbb{E}_\theta\left[(\theta - \omega\overline{X} - (1-\omega)\mu_0)^2\right] \\
&= \sup_\theta \ \mathbb{E}_\theta\left[(\omega(\overline{X} - \theta) + (1-\omega)(\mu_0 - \theta))^2\right] \\
&= \sup_\theta \ \omega^2 \mathrm{Var}\left(\overline{X}\right) + (1-\omega)^2(\mu_0 - \theta)^2 \\
&= \sup_\theta \ \frac{\omega^2\sigma^2}{n} + (1-\omega)^2(\mu_0 - \theta)^2 \\
&= +\infty
\end{aligned}$$

These estimators have much poorer worst-case risk than $\overline{X}$, hence they are certainly not minimax. To prove that $\overline{X}$ is indeed a minimax estimator, we need to generalize the previous definitions and theorems in the following way.

Definition 6 (Least Favourable Sequence of Priors)**.** *Let $\{\Lambda_m\}$ be a sequence of priors with minimal average risk*

$$r_{\Lambda_m} = \inf_\delta \int_\Omega R(\theta, \delta)\, \mathrm{d}\Lambda_m(\theta).$$

Then $\{\Lambda_m\}$ is a least favourable sequence of priors if there is a real number r such that $r_{\Lambda_m} \to r < \infty$ and $r \geq r_{\Lambda'}$ for any prior Λ'.

Theorem 7 (TPE 5.1.12)**.** *Suppose there is a real number r such that $\{\Lambda_m\}$ is a sequence of priors with $r_{\Lambda_m} \to r < \infty$. Let δ be any estimator such that $\sup_\theta R(\theta, \delta) = r$. Then we have*

(i) δ is minimax;

(ii) $\{\Lambda_m\}$ is least-favourable.

Proof. (i) Let δ' be any other estimator. Then for any m, we have

$$\sup_\theta R(\theta, \delta') \geq \int_\Omega R(\theta, \delta')\, \mathrm{d}\Lambda_m(\theta) \geq r_{\Lambda_m}.$$

Then sending $m \to \infty$ yields

$$\sup_\theta R(\theta, \delta') \geq r = \sup_\theta R(\theta, \delta),$$

which implies that δ is minimax.

(ii) Let Λ' be any prior, then

$$r_{\Lambda'} = \int_\Omega R(\theta, \delta_{\Lambda'}) \, \mathrm{d}\Lambda'(\theta) \le \int_\Omega R(\theta, \delta) \, \mathrm{d}\Lambda'(\theta) \le \sup_\theta R(\theta, \delta) = r,$$

which means that $\{\Lambda_m\}$ is least favourable.

$\square$

Remark.

1. *Unlike Theorem 5.1.4 (TPE), this theorem does not guarantee the uniqueness of the minimax estimator even if the Bayes estimators δ_{Λ_m}'s are unique. The problem arises from the step that we send m to the limit.*

2. *This theorem allows us to consider a much wider class of estimators, instead of limiting our attentions to Bayes estimators only. Specifically, we may also consider the estimators that comes from a sequence of priors.*

Example 3 (cont'd). *If we manage to find a sequence of priors $\{\Lambda_m\}$ such that $r_{\Lambda_m} \to \frac{\sigma^2}{n} = r$, then we can obtain a minimax estimator for θ. Let consider the sequence of priors $\Lambda_m \sim \mathcal{N}(0, m^2)$ (Λ_m will tend to the uniform prior over $\mathbb{R}$ which is improper with $\pi(\theta) = 1$ for any $\theta \in \mathbb{R}$). This will yield the following posterior distribution.*

$$\begin{aligned}
f(\theta | x_1, ..., x_n) &\propto \pi(\theta) \cdot f(x_1, ..., x_n | \theta) \\
&\propto \exp\left(-\frac{\theta^2}{2m^2} - \frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2} \right) \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{1}{m^2} + \frac{n}{\sigma^2} \right) \theta^2 + \frac{n\overline{x}}{\sigma^2} \cdot \theta \right) \\
&\sim \mathcal{N}\left(\frac{\frac{n\overline{x}}{\sigma^2}}{\frac{1}{m^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{m^2} + \frac{n}{\sigma^2}} \right)
\end{aligned}$$

Note that the posterior variance does not depend on $(X_1, ..., X_m)$, hence

$$r_{\Lambda_m} = \frac{1}{\frac{1}{m^2} + \frac{n}{\sigma^2}} \to \frac{\sigma^2}{n} = \sup_\theta R(\theta, \overline{X}).$$

It now follows from Theorem 5.1.12 (TPE) that $\overline{X}$ is minimal and $\{\Lambda_m\}$ is least favourable.

We remind that the choice of loss function will also influence the corresponding minimax estimators. Specially, we consider the following example.

Example 4 (weighted squared loss). *Let $X \sim \text{Binomial}(n, \theta)$ with the loss function $L(\theta, d) = \frac{(d-\theta)^2}{\theta(1-\theta)}$. We may view this loss function as the weighted squared loss function with weights $w(\theta) = \frac{1}{\theta(1-\theta)}$.*

Note that for any θ, $R(\theta, X/n) = \frac{1}{n}$, which is constant in θ. This suggests that X/n can be minimax. But be reminded that we cannot directly apply TPE 4.2.3 because L is not the vanilla squared loss function.

Consider the prior $\Theta \sim \Lambda_{\alpha,\beta} = \text{Beta}(\alpha, \beta)$, for some $\alpha, \beta > 0$. By results in Lecture 8, we have $\Theta | X \sim \text{Beta}(X + \alpha, n - X + \beta)$ and we can find the Bayes estimator as

$$\delta_\Lambda(X) = \frac{\mathbb{E}_{\Theta|X}\left(\frac{1}{1-\Theta} \middle| X \right)}{\mathbb{E}_{\Theta|X}\left(\frac{1}{\Theta(1-\Theta)} \middle| X \right)}$$

Suppose we have observed $X = x$ with $\alpha + x > 1$ and $n + \beta + x > 1$, then the resulting Bayes estimator is

$$\delta_{\alpha,\beta}(x) = \frac{\alpha + x - 1}{\alpha + \beta + n - 2}.$$

In particular, when $\alpha = \beta = 1$, we have $\delta_{1,1}(x) = x/n$ minimizes posterior risk under prior $\Lambda_{1,1}$ after observing $0 < x < n$.

When $x \in \{0, n\}$, then the posterior risk under the prior $\Lambda_{1,1}$ after observing $X = x$ and deciding $\delta(x) = d$ is

$$\int_0^1 \frac{(d-\theta)^2}{\theta(1-\theta)} \cdot \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \cdot \theta^x (1-\theta)^{n-x} \, \mathrm{d}\theta,$$

which for $x = 0$ reduces to $\int_0^1 \frac{(n+1)(1-\theta)^{n-1}(d-\theta)^2}{\theta} \, \mathrm{d}\theta$. Note this converges only when $\delta(0) = 0$. Similarly, one can deduce that $\delta(n) = 1$.

Now we may conlcude that X/n minimizes the posterior risk under prior distribution $\Lambda_{1,1}$ for any outcome X. Hence X/n is indeed minimax under such weighted squared loss function.

3 Next Lecture

1. Admissibility of minimax estimators;
2. Hypothesis testing (NP lemma/UMP).