STAT5010 Advanced Statistical Inference **Fall 2023**

Lecture 6: Minimaxity

Lecturer: Tony Sit *Scribe: Jishuai MIAO and Yuwen WANG*

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.*

6.1 Minimax Estimators and Worst-Case Optimality

In minimax estimation, we collapse our risk function by looking at the worst-case risk. Given $X \sim P_\theta$, where $\theta \in \Omega$, and a loss function $L(\theta, d)$, we want to find an estimator δ that minimizes the maximum risk:

$$\sup_{\theta \in \Omega} R(\theta, \delta).$$

Any such δ is called a minimax estimator. In Bayes estimation, we essentially had a single, general-purpose way of deriving a Ba estimator: minimize the posterior risk. The derivation of minimax estimators is often prescriptive and more problem-specific. However, we will develop a few tools which, w they apply, will identify minimax estimators.

Perhaps surprisingly, one of the most effective ways of finding minimax estimator: to restrict our attention to Bayes estimators. To understand the connection, we will n to introduce some additional notation. First we recall the definition of the Bayes risk minimum average risk) under any prior distribution Λ,

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

Definition 6.1 *We say that a prior Λ is a least favorable prior if $r_\Lambda \geq r_{\Lambda'}$ for any other prior distribution Λ'. (Note that we always use the unmodified word "prior" to mean a proper prior.)*

Theorem 6.2 (TPE 5.1.4) *Suppose δ_Λ is Bayes for Λ with*

$$r_{\Lambda \in \Omega} = \sup_\theta R\left(\theta, \delta_\Lambda\right)$$

That is, the Bayes risk of δ_Λ is the maximum risk of δ_Λ. Then,

1. *δ_Λ is minimax*
2. *Λ is a least favorable prior*
3. *If δ_Λ is the unique Bayes estimator for Λ (a.s. for all P_θ), then it is the unique minimax estimator.*

Proof: If δ is any other estimator, then we have that

$$\sup_{\theta \in \Omega} R(\theta, \delta) \geq \int R(\theta, \delta) d\Lambda(\theta) \geq \int R\left(\theta, \delta_\Lambda\right) d\Lambda(\theta) = \sup_{\theta \in \Omega} R\left(\theta, \delta_\Lambda\right)$$

where the first step holds because the worst-case risk of δ is greater than (or equal to) the average risk of δ, the second step holds because δ_Λ is Bayes (and hence has an average risk no higher than that of δ), and the third step holds because of our assumption that the Bayes risk of δ_Λ is equal to the worst-case risk. This implies that δ_Λ is minimax.

If δ_Λ is the unique Bayes estimator, then the second inequality above is strict for $\delta \neq \delta_\Lambda$, which implies that δ_Λ is the unique minimax. Let Λ' be any other prior distribution. Then, we have that

$$r_{\Lambda'} = \inf_\delta \int R(\theta, \delta) d\Lambda'(\theta) \leq \int R\left(\theta, \delta_\Lambda\right) d\Lambda'(\theta) \leq \sup_\theta R\left(\theta, \delta_\Lambda\right) = r_\Lambda.$$

The first step first and second steps are by the definition of Bayes risk, and the third step holds because the worst-case risk of δ_Λ is no less than its average risk over the distribution Λ'. Since the worst-case risk of δ_Λ is its Bayes risk over Λ (by our assumption), we can infer that Λ is a least favorable prior distribution. ■

Thus, we can find a minimax estimator by finding a Bayes estimator with Bayes risk equal to its maximum risk. The following corollary highlights an important special case of this strategy.

Corollary 6.3 (TPE 5.1.5) *If a Bayes estimator δ_Λ has constant risk (that is, $R\left(\theta, \delta_\Lambda\right) = R\left(\theta', \delta_\Lambda\right)$ for all θ and θ'), then δ_Λ is minimax. Note that this is a sufficient but not necessary condition.*

It is often relatively easy to check whether an estimator has constant risk, and this is typically our first line of attack for determining whether an estimator is minimax. More generally we could find a prior support set ω such that $\Lambda(\omega) = 1$ and for which $R\left(\theta, \delta_\Lambda\right)$ is maximum for all $\theta \in \omega$.

Corollary 6.4 (TPE 5.1.6) *Define*

$$\omega_\Lambda = \left\{\theta : R\left(\theta, \delta_\Lambda\right) = \sup_{\theta'} R\left(\theta', \delta_\Lambda\right)\right\}$$

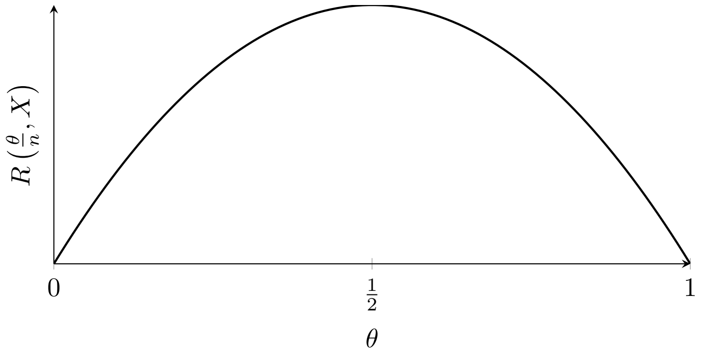
Then, a Bayes estimator δ_Λ is minimax if $\Lambda\left(\omega_\Lambda\right) = 1$. (TPE misstates this result as if and only if, but the only if component is false. In other words, this condition is sufficient, but not necessary.)

Example 6.5 *Suppose $X \sim \mathrm{Binom}(n, \theta)$ for some $\theta \in (0, 1)$ and that we use the squared error loss function. Is the sample proportion $\frac{X}{n}$ minimax? The risk of this estimator is*

$$R\left(\theta, \frac{X}{n}\right) = \frac{\theta(1-\theta)}{n}.$$

The graph of $R\left(\theta, \frac{X}{n}\right)$ versus θ looks like the following:

The graph of $R\left(\frac{\theta}{n}, X\right)$ versus θ



The risk has a unique maximum at $\theta = \frac{1}{2}$, so the worst-case risk is

$$\sup_{\theta \in \Omega} R\left(\theta, \frac{X}{n}\right) = R\left(\frac{1}{2}, \frac{X}{n}\right) = \frac{1}{4n}.$$

Unfortunately, we cannot apply TPE 5.1.6 directly because if $\Lambda\left(\left\{\frac{1}{2}\right\}\right) = 1$, then $\delta_\Lambda(X) = \frac{1}{2} \neq \frac{X}{n}$.

However, we can use the TPE 5.1.5 to find a minimax estimator and then compare the risk of the minimax estimator with that of $\frac{X}{n}$. To find a minimax estimator, we will search for a prior such that the Bayes estimator has constant risk.

Recall the following useful fact. Under the prior distribution Beta (a, b), the Bayes estimator under the squared error loss is

$$\delta_{a,b}(X) = \frac{X + a}{n + a + b}.$$

For any a and b,

$$\begin{aligned} R\left(\theta, \delta_{a,b}\right) &= E_\theta\left[\left(\frac{X+a}{n+a+b} - \theta\right)^2\right] \\ &= \frac{1}{(n+a+b)^2} E_\theta\left[(X + a - (n+a+b)\theta)^2\right] \\ &= \frac{1}{(n+a+b)^2} E_\theta\left[(X - n\theta - a(\theta - 1) - \theta b)^2\right] \\ &= \frac{1}{(n+a+b)^2}\left(n\theta(1-\theta) + (a(\theta-1) + \theta b)^2\right). \end{aligned}$$

This is a quadratic function of θ. To eliminate the θ dependence in $R\left(\theta, \delta_{a,b}\right)$, we need the coefficients of the linear and quadratic terms to equal zero. The coefficient of θ^2 is

$$-n + (a+b)^2$$

so we need $a + b = \sqrt{n}$ (since $a, b > 0$). The coefficient of θ is

$$n - 2a(a+b) = n - 2a\sqrt{n},$$

so we need $a = b = \frac{\sqrt{n}}{2}$. With these choices of a and b, the risk of $R\left(\theta, \delta_{a,b}\right)$ is constant, which implies that Beta $\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$ is a least favorable prior with constant risk. Then our Bayes estimator

$$\delta_{\frac{\sqrt{\pi}}{2}, \frac{\sqrt{n}}{2}}(X) = \frac{X + \frac{\sqrt{n}}{2}}{n + \sqrt{n}},$$

is minimax with constant risk of

$$\frac{1}{4(\sqrt{n}+1)^2}.$$

Since the worst-case risk of $\frac{X}{n}$ is $\frac{1}{4n} > \frac{1}{4(\sqrt{n}+1)^2}$, we can conclude that $\frac{X}{n}$ is not minimax.

6.2 Minimaxity and least favorable prior sequences

In this part, we will extend our tools for deriving minimax estimators. Last time, we discovered that minimax estimators can arise from Bayes estimators under least favorable priors. However, it turns out that minimax estimators may not be Bayes estimators. Consider the following example, where our old approach fails.

Example 6.6 (Minimax for i.i.d. Normal random variables with unknown mean θ) *Let $X_1, \ldots, X_\mathrm{n} \overset{iid}{\sim} \mathcal{N}\left(\theta, \sigma^2\right)$, with σ^2 known. Our goal is to estimate θ under squared-error loss. For our first guess, pick the natural estimator X. Note that it has constant risk $\frac{\sigma^2}{n}$, which suggests minimaxity because we know that Bayes estimators with constant risk are also minimax estimators. However, X is not Bayes for any prior, because under squared-error loss unbiased estimators are Bayes estimators only in the degenerate situations of zero risk (TPE Theorem 4.2.3), and X is unbiased. Thus, we cannot conclude by our previous results (e.g., TPE Corollary 5.1.5) that $\bar{X}$ is minimax.*

We might try to consider the wider class of estimators $\delta_{a,\mu_0}(X) = aX + (1-a)\mu_0$ for $a \in (0,1)$ and $\mu_0 \in R$, because many of the Bayes estimators we've encountered are convex combinations of a prior and a data mean. Note however that the worst case risk for these estimators is infinite:

$$\begin{aligned}
\sup_\theta E_\theta[\theta - \delta(X)]^2 &= \sup_\theta \left\{ a^2 \operatorname{Var}_\theta(\bar{X}) + (1-a)^2 \left(\theta - \mu_0\right)^2 \right\} \\
&= \frac{a^2\sigma^2}{n} + (1-a)^2 \sup_\theta \left(\theta - \mu_0\right)^2 \\
&= +\infty
\end{aligned}$$

Since these estimators have poorer worst case risk than $\bar{X}$, they certainly cannot be minimax. We could keep trying to find Bayes estimators with better worst-case performance than X, but we would fail: it turns out that $\bar{X}$ is in fact minimax. To establish this, we will extend our minimax results to the limits of Bayes estimators, rather than restricting attention to Bayes estimators only.

Definition 6.7 (Least Favorable Sequence of Priors) *Let $\{\Lambda_m\}$ be a sequence of priors with minimal average risk $r_{\Lambda_m} = \inf_\delta \int R(\theta, \delta) d\Lambda_m(\theta)$. Then, $\{\Lambda_m\}$ is a least favorable 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'.*

The reason for studying the limit of priors is that it may help us establish minimaxity. Since there need not exist a prior Λ such that the associated Bayes estimator has average risk r, this definition is less restrictive than that of a least-favorable prior. We can prove an analogue of TPE Theorem 5.1.4 in this new setting.

Theorem 6.8 (TPE 5.1.12) *Suppose there is 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,*

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

Proof: 1. Let δ' be any other estimator. Then, for any m,

$$\sup_\theta R\left(\theta, \delta'\right) \geq \int R\left(\theta, \delta'\right) d\Lambda_m(\theta) \geq r_{\Lambda_m},$$

so that sending $m \to \infty$ yields

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

which means that δ is minimax.
2. Let Λ' be any prior, then

$$r_{\Lambda'} = \int R\left(\theta, \delta_{\Lambda'}\right) d\Lambda'(\theta) \leq \int R(\theta, \delta) d\Lambda'(\theta) \leq \sup_\theta R(\theta, \delta) = r,$$

which means that $\{\Lambda_m\}$ is least favorable. ■

Unlike Theorem 5.1.4, this result does not guarantee uniqueness, even if the Bayes estimators δ_{Λ_m} are unique. This is because the limiting step in the proof of (1) changes any strict inequality to nonstrict inequality. However, this result allows to check much wider class of estimators, since to check that the estimator is indeed a minimax estimator we need to find only the sequence of Bayes risks convergent to maximum risk of our candidate.

Example 6.9 (Minimax for i.i.d. Normal random variables, continued) *We now have the tools to confirm our suspicion that X is minimax. By Theorem TPE 5.1.12 above, it suffices to find a sequence $\{\Lambda_m\}$ such that $r_{\Lambda_m} \to \frac{\sigma^2}{n} =: r$. Using the conjugate prior is a good starting point, so we let $\{\Lambda_m\}$ be the conjugate priors $\left\{N\left(0, m^2\right)\right\}$ with variance tending to ∞, so that Λ_m tends to the (improper with $\pi(\theta) = 1, \forall \theta \in R$) uniform prior on R. By TPE Example 4.2.2, the posterior for θ associated with each Λ_m is*

$$\theta \mid X_1, \ldots, X_n \sim N\left(\frac{\frac{nX}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{m^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{m^2}}\right).$$

In particular, the posterior variance does not depend on $X_1, \ldots, X_n$, so Lemma 1 below automatically yields the Bayes risk

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

It follows from Theorem TPE 5.1.12 that X is minimax and $\{\Lambda_m\}$ is least favorable.

Lemma 6.10 (TPE 5.1.13) *If the posterior variance* $\operatorname{Var}_{\Theta|X}(g(\Theta) \mid X = x$ *) is constant in x, then under squared error loss, $r_\Lambda = \operatorname{Var}_{\Theta|X}(g(\Theta) \mid X = x)$.*

We know that the posterior mean minimizes Bayes risk, so this result can be obtained by plugging in the posterior mean of $g(\theta)$ into the average risk.

6.3 Minimaxity via submodel restriction

The following example illustrates the technique of deriving a minimax estimator for a general family of models by restricting attention to a subset of that family. The idea comes from simple observation that if the estimator is minimax in submodel and its risk doesn't change when we go to a larger model then estimator is minimax in this larger class.

Example 6.11 (Minimax for i.i.d. Normal random variables, unknown mean and variance) *Reconsider Example 1 in the case that the variance is unknown. That is, let* $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\theta, \sigma^2\right)$*, with both* θ *and* σ^2 *unknown. Note that*

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

and in fact, the maximum risk of any estimator in this setting is infinite, so the question of minimaxity is uninteresting. Therefore, we restrict attention to the family parameterized by $\Omega = \left\{\left(\theta,\sigma^2\right) : \theta \in R, \sigma^2 \leq B\right\}$*, where* B *is a known constant. Assume* δ *is any other estimator. Calculating the risk of* $\bar{X}$ *within this family, we find*

$$\begin{aligned}
\sup_{\theta\in R,\sigma^2\leq B} R\left(\left(\theta,\sigma^2\right), \bar{X}\right) &= \frac{B}{n} \\
&= \sup_{\theta\in R,\sigma^2=B} R\left(\left(\theta,\sigma^2\right), \bar{X}\right) \\
&\leq \sup_{\theta\in R,\sigma^2=B} R\left(\left(\theta,\sigma^2\right), \delta\right) \ \textit{[submodel minimax]} \\
&\leq \sup_{\theta\in R,\sigma^2\leq B} R\left(\left(\theta,\sigma^2\right), \delta\right),
\end{aligned}$$

where the first inequality follows from the fact that $\bar{X}$ *is minimax for i.i.d. normals with known* σ^2*, and the second inequality follows from the fact that we are taking the supremum over a larger set. Hence, we are able to show that* $\bar{X}$ *is minimax over* Ω *by focusing on the case where* σ^2 *is known. Notice further that the form of the estimator does not depend on the upper bound* B*, though the bound is necessary for minimaxity to be worth investigating.*