STATS3100A - Lecture 10

Dominik Rothenhaeusler
Scribed by Michael Howes

10/20/21

Contents

1 Announcements

- No new homework this week.
- Last years midterm+solutions are on Canvas.
- Midterm review session Monday 6pm on Zoom.
- Midterm in one week time. Midterm is open book and 90 minutes long.
- The relevant content is everything up to and including today.

2 Minimaxity and limits of Bayes estimators

Recall that if a Bayes estimator δ_Λ has constant risk, then δ_Λ is minimax.

Example 1. Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$ where σ^2 is known and $L(\theta, d) = (\theta - d)^2$. We have seen that $\bar{X}$ has constant risk but it cannot be the Bayes estimator for any proper prior since it is unbiased. One can show that $\bar{X}$ is a generalized Bayes estimator w.r.t. the improper prior $\pi(\theta) = 1$. We wish to show that $\bar{X}$ is minimax. To do this we will look at limits of Bayes estimators

Definition 1. Let $(\Lambda_m)_{m=1}^\infty$ be a sequence of priors with $r_{\Lambda_m} := \inf_\delta r(\Lambda_m, \delta)$. The sequence (Λ_m) is called *least favourable* if $r_{\Lambda_m} \to r$ and $r \geq r_{\Lambda'}$ for all priors Λ'.

Theorem 1. [TPE 5.1.12] *Suppose (Λ_m) is a sequence of priors with $r_{\Lambda_m} \to r < \infty$. Let δ be an estimator such that $\sup_\theta R(\theta, \delta) = r$, Then δ is minimax and (Λ_m) is least favourable.*

Proof. Let δ' be an estimator. We know that

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

Letting $m \to \infty$ we can conclude that

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

Thus δ is minimax. Now let Λ' be a prior. We know that

$$\begin{aligned} r_{\Lambda'} &\leq \int_\Omega R(\theta, \delta) d\Lambda' \\ &\leq \sup_\theta R(\theta, \delta) \\ &= r. \end{aligned}$$

Thus (Λ_m) is least favourable. $\square$

Example 2. Returning to our example with $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$ and $L(\theta, d) = (\theta - d)^2$. Consider the prior $\Lambda_m \sim N(0, m^2)$. We have seen that

$$\Theta | X \sim N\left(\frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{m^2}} \bar{X}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{m^2}} \right).$$

Thus $\delta_{\Lambda_m} = \frac{n}{\sigma^2} \frac{n}{\sigma^2} + \frac{1}{m^2} \bar{X}$ is the posterior mean and hence the Bayes estimator for Λ_m. Note that $\delta_{\Lambda_n}(X)$ is the mean of $\Theta | X$ and thus

$$\begin{aligned} r_{\Lambda_m} &= \mathbb{E}\left[L(\Theta, \delta_{\Lambda_m}(X)) \right] \\ &= \mathbb{E}\left[(\Theta - \delta_{\Lambda_n}(X))^2 \right] \\ &= \mathbb{E}\left[\mathbb{E}[(\Theta - \delta_{\Lambda_n}(x))^2 | X = x] \right] \\ &= \mathbb{E}\left[\mathrm{Var}(\Theta | X = x) \right] \\ &= \frac{1}{\frac{n}{\sigma^2} + \frac{1}{m^2}}. \end{aligned}$$

Thus as $m \to \infty$ we have $r_{\Lambda_m} \to \frac{\sigma^2}{n} = R(\theta, \bar{X})$. Thus $\bar{X}$ is minimax.

3 Randomized estimators

Recall that a randomized estimator is one of the form $\delta(X, U)$ where $U \sim \mathrm{Unif}([0,1])$ and $U \perp\!\!\!\perp X$. This is in constrast to estimators of the form $\delta(X)$ which are deterministic functions of the data. We saw that for convex loss functions we can ignore randomized estimators since by Jensen's inequality

$$\mathbb{E}_\theta L(\theta, \delta(X, U)) \geq \mathbb{E}_\theta L(\theta, \mathbb{E}[\delta(X, U) | X]) = \mathbb{E}_\theta L(\theta, \eta(X)).$$

For non-convex loss functions the minimax estimator may be randomized.

Example 3. Consider X a binomial random variable with parameters (n, θ) where n is fixed and $\theta \in [0, 1]$. Consider the 0-1 loss

$$L(\theta, d) = \begin{cases} 0 & \text{if } |d - \theta| < \alpha, \\ 1 & \text{else.} \end{cases}$$

If δ is a non-randomized estimator, δ can take at most $n + 1$ values. For $\alpha < \frac{1}{2(n+1)}$, there exists θ_0 such that $|\delta(x) - \theta_0| \geq \alpha$ for all x and thus the worst case risk of δ is 1. Consider $\delta'(X, U) = U$. For every θ,

$$R(\theta, \delta') = \mathbb{P}(|U - \theta| \geq \alpha) \leq 1 - \alpha.$$

Thus the worst case risk of δ' is $1 - \alpha < 1$.

4 "Boosting" via submodel restrictions

Consider $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$ where σ^2 and θ are both unknown but $\sigma^2 \le B_0$. Consider squared error loss $L(\theta, d) = (\theta - d)^2$. Is $\bar{X}$ minimax? To prove this we need to show that

$$\sup_{\theta, \sigma^2 \le B_0} R((\theta, \sigma^2), \delta) \ge \sup_{\theta, \sigma^2 \le B_0} R((\theta, \sigma^2), \bar{X}),$$

where δ is any estimator. We know that

$$\sup_{\theta, \sigma^2 \le B_0} R((\theta, \sigma^2), \bar{X}) = \sup_{\theta, \sigma^2 \le B_0} \frac{\sigma^2}{n} = \frac{B_0}{n}.$$

Thus (and this is crucial)

$$\sup_{\theta, \sigma^2 \le B_0} R((\theta, \sigma^2), \bar{X}) = \sup_{\theta, \sigma^2 = B_0} R((\theta, \sigma^2), \bar{X}).$$

We know that for fixed variance $\bar{X}$ is minimax. Thus for any estimator δ we have

$$\begin{aligned}
\sup_{\theta, \sigma^2 \le B_0} R((\theta, \sigma^2), \delta) &\ge \sup_{\theta, \sigma^2 = B_0} R((\theta, \sigma^2), \delta) \\
&\ge \sup_{\theta, \sigma^2 = B_0} R((\theta, \sigma^2), \bar{X}) \\
&= \sup_{\theta, \sigma^2 \le B_0} R((\theta, \sigma^2), \bar{X}).
\end{aligned}$$

Thus $\bar{X}$ is minimax. Formalizing this example, we have:

Lemma 1. [TPE 5.1.15] *If δ is minimax for a submodel $\Omega_0 \subseteq \Omega$ and $\sup_{\theta \in \Omega} R(\theta, \delta) = \sup_{\theta \in \Omega_0} R(\theta, \delta)$, then δ is minimax for the full model Ω.*

Proof. This is the same argument we saw in the example. For any estimator δ',

$$\begin{aligned}
\sup_{\theta \in \Omega} R(\theta, \delta') &\ge \sup_{\theta \in \Omega_0} R(\theta, \delta') \\
&\ge \sup_{\theta \in \Omega_0} R(\theta, \delta) \\
&= \sup_{\theta \in \Omega} R(\theta, \delta)
\end{aligned}$$

□

Example 4 (Non-parametric)**.** Suppse $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$, where $F \in \mathcal{F}$ has mean $\mu(F)$ and variance $\sigma^2(F) < B$. Our goal is to estimate $\mu(F)$ under $L(F, d) = (\mu(F) - d)^2$. We wish to show that the minimax estimator is $\delta(X) = \bar{X}$. We know that $\bar{X}$ is minimax on the submodel $\mathcal{F}_0 = \{N(\theta, \sigma^2) : \sigma^2 \le B\}$. On the full model $\mathcal{F}$, the estimator $\bar{X}$ has risk $R(F, \bar{X}) = \frac{\sigma^2(F)}{n} \le \frac{B}{n}$. Thus we hav

$$\sup_{F \in \mathcal{F}_0} R(F, \bar{X}) = \sup_{F \in \mathcal{F}} R(F, \bar{X}).$$

Thus by boosting, $\bar{X}$ is sufficient on the full model.

Example 5 (Another non-parametric example)**.** Now suppose that $\mathcal{F}$ is the set of all distributions with support on $[0, 1]$. Again we will do estimation with squared error loss $L(F, d) = (\mu(F) - d)^2$. Suppose $X_i \overset{\text{iid}}{\sim} \mathcal{F}$. Consider the submodel $\mathcal{F}_0$ of all Bernoulli distributions with parameter $\theta \in (0, 1)$. We have seen that in this setting the minimax estimator is

$$\delta_n(X) = \frac{\sum_{i=1}^n X_i + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}.$$

By studying the risk of δ_n on the full model we can show that δ_n is minimax on $\mathcal{F}$.