STA732
Statistical Inference

Lecture 13: Minimax estimators

Yuansi Chen

Spring 2022

Duke University

https://www2.stat.duke.edu/courses/Spring22/sta732.01/

Recap from Lecture 12

1. Minimax risk as the minimum worst-case risk
2. Minimax estimator is
    - the one that minimizes the worst-case risk
    - the optimal strategy when Nature is adversarial
3. Least favorable priors can be used to construct minimax estimator from Bayes estimator

Goal of Lecture 13

1. Least favorable prior sequence
2. Minimaxity via submodel restriction
3. Minimaxity vs. admissibility

5.1, 5.2 of Lehmann and Casella

Least favorable prior sequence

Minimax risk

The minimum achievable sup-risk is called the minimax risk of the estimation problem

$$r^* = \inf_{\delta} \sup_{\theta \in \Omega} R(\theta, \delta)$$

Recall Least favorable prior

Recall the Bayes risk under prior Λ is

$$r_\Lambda = \inf_\delta \int R(\theta, \delta) d\Lambda(\theta).$$

Def. least favorable prior

A prior Λ is a least favorable prior if $r_\Lambda \geq r_{\Lambda'}$ for any other prior Λ'.

The Bayes risk for the least favorable prior should give the tightest lower bound for the minimax risk

Motivation

From last lecture, if we

- find a least favorable prior
- show that its Bayes risk equals to the worst-case risk

then we find the minimax estimator which is the corresponding Bayes estimator.

It turns out that minimax estimators may not be Bayes!

Intuition: sometimes the least favorable prior is not a proper prior

Motivating example: minimax for normal mean estimation

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ with σ^2 known. We use squared error loss.

- Compute the risk of $\bar{X}$
- Is $\bar{X}$ Bayes with some prior? $\bar{X}$ is not Bayes but a limit of Bayes estimators
- Is $\bar{X}$ minimax?

Least favorable prior sequence

Def. least favorable prior sequence

Let $\{\Lambda_m\}$ be a sequence of priors with minimal average risks $\left\{r_{\Lambda_m}\right\}$ where $r_{\Lambda_m} = \inf_\delta \int R(\theta, \delta) d\Lambda_m(\theta)$. 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'

Remark: less restrictive than the definition of a least-favorable prior. Useful when the space of achievable risk with proper prior is not compact

Analogue theorem of 5.1.4

Thm. 5.1.12 in Lehmann and Casella

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

1. δ is minimax
2. $\{\Lambda_m\}$ is a least-favorable prior sequence

Proof of Thm. 5.1.12:

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

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

Back to our example of normal mean estimation

Use Thm. 5.1.12 to show that $\bar{X}$ is minimax

Minimaxity via submodel restriction

Motivation

If an estimator is minimax in submodel and its risk doesn’t change when we go to a larger model then it is minimax in this larger class.

Example: minimax for i.i.d. normal random variables unknown mean and variance

$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$, with both θ and σ^2 unknown. We would like to estimate θ with squared error loss.

- Find a minimax estimator for $\Omega = \left\{(\theta, \sigma^2) : \theta \in \mathbb{R}, \sigma^2 \leq B_0\right\}$

Minimaxity via submodel restriction

Lem. Lehmann and Casella 5.1.15

Suppose δ is minimax for a submodel $\theta \in \Omega_0 \subset \Omega$ and satisfies

$$\sup_{\theta \in \Omega_0} R(\theta, \delta) = \sup_{\theta \in \Omega} R(\theta, \delta)$$

Then, δ is minimax for the full model $\theta \in \Omega$.

proof idea: same as in the normal mean example

Example: nonparametric minimax estimator, LC 5.1.16

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. with common CDF F, with mean $\mu(F) < \infty$ and variance $\sigma^2(F) < \infty$. Find a minimax estimator of $\mu(F)$ under squared error loss under each of the following csontraint

1. Assume $\sigma^2(F) \leq B$
2. Assume $F \in \mathcal{F}$ where $\mathcal{F}$ is the set of all CDFs with support contained in $[0, 1]$.

proof:

1. Try $\bar{X}$
2. Try the minimax estimator for binomial

Admissibility of minimax estimators

Admissible estimator with constant risk is minimax

Lemma

If δ is admissible with constant risk, then δ is also minimax

In general, minimaxity does not guarantee admissibility!

But unique minimax estimator is admissible

admissible vs. minimax in normal mean estimation

Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ where σ^2 is known, and θ is the estimand. Then we know that $\bar{X}$ is minimax under squared error loss.

- Is $\bar{X}$ admissible?
- More generally, for $a, b \in \mathbb{R}$, is $a\bar{X} + b$ admissible?

See Example 5.2.5 in Lehmann and Casella proof idea:

- To show admissibility: unique Bayes, or show no dominating estimators
- To show inadmissiblity: find an dominating estimator

1. $0 < a < 1$
2. $a = 0$
3. $a = 1, b \neq 0$
4. $a > 1$
5. $a < 0$
6. $a = 1, b = 0$

Summary

- Least favorable prior sequence as an extension to least favorable prior, can be used to find minimax estimators
- Submodel restriction is another strategy to find minimax estimators
- Minimaxity does not guarantee admissibility in general, while unique Bayes optimality does

Point estimation recap

Useful concepts

- Exponential family
- Data reduction: sufficient statistics, completeness

Optimality in point estimation

- Uniform optimality is rare
- Restrict to a smaller class of estimators
    - Unbiased
    - Equivariant
- Global measures
    - Bayes
    - Minimax
- Large sample theory (to be continued)

What is next?

Basics in large sample theory

- Convergence in probability/distribution
- continuous mapping theorem, Slutsky's theorem
- Delta method

Thank you for attending
See you on Wednesday in Old Chem 025