

Stat 710: Mathematical Statistics

Lecture 6

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 6: Minimax estimators

Consider estimators of a real-valued $\vartheta = g(\theta)$ based on a sample X from P_θ , $\theta \in \Theta$, under loss L and risk $R_T(\theta) = E[L(T(X), \theta)]$.

Minimax estimator

A *minimax estimator* minimizes $\sup_{\theta \in \Theta} R_T(\theta)$ over all estimators T

Discussion

- A minimax estimator can be very conservative and unsatisfactory. It tries to do as well as possible in the worst case.
- A unique minimax estimator is admissible, since any estimator better than a minimax estimator is also minimax.
- We should find an admissible minimax estimator.
- Different for UMVUE: if a UMVUE is inadmissible, it is dominated by a biased estimator)
- If a minimax estimator has some other good properties (e.g., it is a Bayes estimator), then it is often a reasonable estimator.

Lecture 6: Minimax estimators

Consider estimators of a real-valued $\vartheta = g(\theta)$ based on a sample X from P_θ , $\theta \in \Theta$, under loss L and risk $R_T(\theta) = E[L(T(X), \theta)]$.

Minimax estimator

A *minimax estimator* minimizes $\sup_{\theta \in \Theta} R_T(\theta)$ over all estimators T

Discussion

- A minimax estimator can be very conservative and unsatisfactory. It tries to do as well as possible in the worst case.
- A unique minimax estimator is admissible, since any estimator better than a minimax estimator is also minimax.
- We should find an admissible minimax estimator.
- Different for UMVUE: if a UMVUE is inadmissible, it is dominated by a biased estimator)
- If a minimax estimator has some other good properties (e.g., it is a Bayes estimator) then it is often a reasonable estimator.

Lecture 6: Minimax estimators

Consider estimators of a real-valued $\vartheta = g(\theta)$ based on a sample X from P_θ , $\theta \in \Theta$, under loss L and risk $R_T(\theta) = E[L(T(X), \theta)]$.

Minimax estimator

A *minimax estimator* minimizes $\sup_{\theta \in \Theta} R_T(\theta)$ over all estimators T

Discussion

- A minimax estimator can be very conservative and unsatisfactory. It tries to do as well as possible in the worst case.
- A unique minimax estimator is admissible, since any estimator better than a minimax estimator is also minimax.
- We should find an admissible minimax estimator.
- Different for UMVUE: if a UMVUE is inadmissible, it is dominated by a biased estimator)
- If a minimax estimator has some other good properties (e.g., it is a Bayes estimator), then it is often a reasonable estimator.

How to find a minimax estimator?

Candidates for minimax: estimators having constant risks

Theorem 4.11 (minimaxity of a Bayes estimator)

Let Π be a proper prior on Θ and δ be a Bayes estimator of ϑ w.r.t. Π . Suppose δ has constant risk on Θ_Π .

If $\Pi(\Theta_\Pi) = 1$, then δ is minimax.

If, in addition, δ is the unique Bayes estimator w.r.t. Π , then it is the unique minimax estimator.

Proof

Let T be any other estimator of ϑ . Then

$$\sup_{\theta \in \Theta} R_T(\theta) \geq \int_{\Theta_\Pi} R_T(\theta) d\Pi \geq \int_{\Theta_\Pi} R_\delta(\theta) d\Pi = \sup_{\theta \in \Theta} R_\delta(\theta).$$

If δ is the unique Bayes estimator, then the second inequality in the previous expression should be replaced by $>$ and, therefore, δ is the unique minimax estimator.

How to find a minimax estimator?

Candidates for minimax: estimators having constant risks

Theorem 4.11 (minimaxity of a Bayes estimator)

Let Π be a proper prior on Θ and δ be a Bayes estimator of ϑ w.r.t. Π . Suppose δ has constant risk on Θ_Π .

If $\Pi(\Theta_\Pi) = 1$, then δ is minimax.

If, in addition, δ is the unique Bayes estimator w.r.t. Π , then it is the unique minimax estimator.

Proof

Let T be any other estimator of ϑ . Then

$$\sup_{\theta \in \Theta} R_T(\theta) \geq \int_{\Theta_\Pi} R_T(\theta) d\Pi \geq \int_{\Theta_\Pi} R_\delta(\theta) d\Pi = \sup_{\theta \in \Theta} R_\delta(\theta).$$

If δ is the unique Bayes estimator, then the second inequality in the previous expression should be replaced by $>$ and, therefore, δ is the unique minimax estimator.

How to find a minimax estimator?

Candidates for minimax: estimators having constant risks

Theorem 4.11 (minimaxity of a Bayes estimator)

Let Π be a proper prior on Θ and δ be a Bayes estimator of ϑ w.r.t. Π . Suppose δ has constant risk on Θ_Π .

If $\Pi(\Theta_\Pi) = 1$, then δ is minimax.

If, in addition, δ is the unique Bayes estimator w.r.t. Π , then it is the unique minimax estimator.

Proof

Let T be any other estimator of ϑ . Then

$$\sup_{\theta \in \Theta} R_T(\theta) \geq \int_{\Theta_\Pi} R_T(\theta) d\Pi \geq \int_{\Theta_\Pi} R_\delta(\theta) d\Pi = \sup_{\theta \in \Theta} R_\delta(\theta).$$

If δ is the unique Bayes estimator, then the second inequality in the previous expression should be replaced by $>$ and, therefore, δ is the unique minimax estimator.

Example 4.18

Let X_1, \dots, X_n be i.i.d. binary random variables with $P(X_1 = 1) = p$.

Consider the estimation of p under the squared error loss.

The UMVUE \bar{X} has risk $p(1-p)/n$ which is not constant.

In fact, \bar{X} is not minimax (Exercise 67).

To find a minimax estimator by applying Theorem 4.11, we consider the Bayes estimator w.r.t. the beta distribution $B(\alpha, \beta)$ with known α and β (Exercise 1):

$$\delta(X) = (\alpha + n\bar{X})/(\alpha + \beta + n).$$

$$R_\delta(p) = [np(1-p) + (\alpha - \alpha p - \beta p)^2]/(\alpha + \beta + n)^2.$$

To apply Theorem 4.11, we need to find values of $\alpha > 0$ and $\beta > 0$ such that $R_\delta(p)$ is constant.

It can be shown that $R_\delta(p)$ is constant if and only if $\alpha = \beta = \sqrt{n}/2$, which leads to the unique minimax estimator

$$T(X) = (n\bar{X} + \sqrt{n}/2)/(n + \sqrt{n}).$$

The risk of T is $R_T = 1/[4(1 + \sqrt{n})^2]$.

Example 4.18 (continued)

Note that T is a Bayes estimator and has some good properties. Comparing the risk of T with that of \bar{X} , we find that T has smaller risk if and only if

$$p \in \left(\frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{n}{(1+\sqrt{n})^2}}, \frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{n}{(1+\sqrt{n})^2}} \right).$$

Thus, for a small n , T is better (and can be much better) than \bar{X} for most of the range of p (Figure 4.1).

When $n \rightarrow \infty$, the above interval shrinks toward $\frac{1}{2}$.

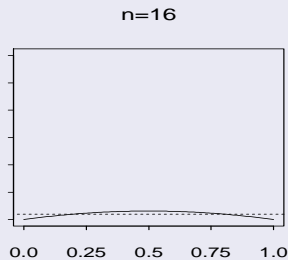
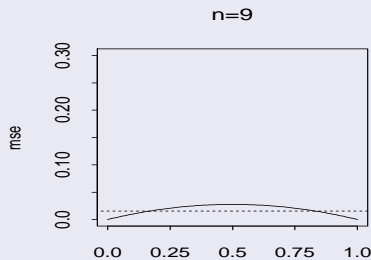
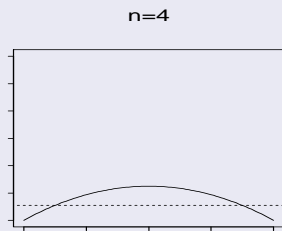
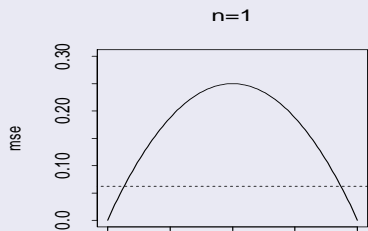
Hence, for a large (and even moderate) n , \bar{X} is better than T for most of the range of p (Figure 4.1).

The limit of the asymptotic relative efficiency of T w.r.t. \bar{X} is $4p(1-p)$, which is always smaller than 1 when $p \neq \frac{1}{2}$ and equals 1 when $p = \frac{1}{2}$. Minimality depends strongly on the loss function.

Under the loss function $L(p, a) = (a-p)^2/[p(1-p)]$, \bar{X} has constant risk and is the unique Bayes estimator w.r.t. the uniform prior on $(0, 1)$. By Theorem 4.11, \bar{X} is the unique minimax estimator.

The risk, however, of T is $1/[4(1+\sqrt{n})^2 p(1-p)]$, which is unbounded.

Figure 4.1. mse's of \bar{X} (curve) and $T(X)$ (straight line) in Example 4.18



A limit of Bayes estimators

In many cases a constant risk estimator is not a Bayes estimator (e.g., an unbiased estimator under the squared error loss), but a limit of Bayes estimators w.r.t. a sequence of priors.

The next result may be used to find a minimax estimator.

Theorem 4.12

Let $\Pi_j, j = 1, 2, \dots$, be a sequence of priors and r_j be the Bayes risk of a Bayes estimator of ϑ w.r.t. Π_j .

Let T be a constant risk estimator of ϑ .

If $\liminf_j r_j \geq R_T$, then T is minimax.

Proof

Similar to the proof of Theorem 4.11.

Although Theorem 4.12 is more general than Theorem 4.11 in finding minimax estimators, it does not provide uniqueness of the minimax estimator even when there is a unique Bayes estimator w.r.t. each Π_j .

A limit of Bayes estimators

In many cases a constant risk estimator is not a Bayes estimator (e.g., an unbiased estimator under the squared error loss), but a limit of Bayes estimators w.r.t. a sequence of priors.

The next result may be used to find a minimax estimator.

Theorem 4.12

Let $\Pi_j, j = 1, 2, \dots$, be a sequence of priors and r_j be the Bayes risk of a Bayes estimator of ϑ w.r.t. Π_j .

Let T be a constant risk estimator of ϑ .

If $\liminf_j r_j \geq R_T$, then T is minimax.

Proof

Similar to the proof of Theorem 4.11.

Although Theorem 4.12 is more general than Theorem 4.11 in finding minimax estimators, it does not provide uniqueness of the minimax estimator even when there is a unique Bayes estimator w.r.t. each Π_j .

A limit of Bayes estimators

In many cases a constant risk estimator is not a Bayes estimator (e.g., an unbiased estimator under the squared error loss), but a limit of Bayes estimators w.r.t. a sequence of priors.

The next result may be used to find a minimax estimator.

Theorem 4.12

Let $\Pi_j, j = 1, 2, \dots$, be a sequence of priors and r_j be the Bayes risk of a Bayes estimator of ϑ w.r.t. Π_j .

Let T be a constant risk estimator of ϑ .

If $\liminf_j r_j \geq R_T$, then T is minimax.

Proof

Similar to the proof of Theorem 4.11.

Although Theorem 4.12 is more general than Theorem 4.11 in finding minimax estimators, it does not provide uniqueness of the minimax estimator even when there is a unique Bayes estimator w.r.t. each Π_j .

A limit of Bayes estimators

In many cases a constant risk estimator is not a Bayes estimator (e.g., an unbiased estimator under the squared error loss), but a limit of Bayes estimators w.r.t. a sequence of priors.

The next result may be used to find a minimax estimator.

Theorem 4.12

Let $\Pi_j, j = 1, 2, \dots$, be a sequence of priors and r_j be the Bayes risk of a Bayes estimator of ϑ w.r.t. Π_j .

Let T be a constant risk estimator of ϑ .

If $\liminf_j r_j \geq R_T$, then T is minimax.

Proof

Similar to the proof of Theorem 4.11.

Although Theorem 4.12 is more general than Theorem 4.11 in finding minimax estimators, it does not provide uniqueness of the minimax estimator even when there is a unique Bayes estimator w.r.t. each Π_j .

In Example 2.25, we actually applied the result in Theorem 4.12 to show the minimaxity of \bar{X} as an estimator of $\mu = EX_1$ when X_1, \dots, X_n are i.i.d. from a normal distribution with a known $\sigma^2 = \text{Var}(X_1)$, under the squared error loss.

To discuss the minimaxity of \bar{X} in the case where σ^2 is unknown, we need the following lemma.

Lemma 4.3

Let Θ_0 be a subset of Θ and T be a minimax estimator of ϑ when Θ_0 is the parameter space. Then T is a minimax estimator if

$$\sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta).$$

Proof

If there is an estimator T_0 with $\sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta)$, then

$$\sup_{\theta \in \Theta_0} R_{T_0}(\theta) \leq \sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta),$$

which contradicts the minimaxity of T when Θ_0 is the parameter space. Hence, T is minimax when Θ is the parameter space.

In Example 2.25, we actually applied the result in Theorem 4.12 to show the minimaxity of \bar{X} as an estimator of $\mu = EX_1$ when X_1, \dots, X_n are i.i.d. from a normal distribution with a known $\sigma^2 = \text{Var}(X_1)$, under the squared error loss.

To discuss the minimaxity of \bar{X} in the case where σ^2 is unknown, we need the following lemma.

Lemma 4.3

Let Θ_0 be a subset of Θ and T be a minimax estimator of ϑ when Θ_0 is the parameter space. Then T is a minimax estimator if

$$\sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta).$$

Proof

If there is an estimator T_0 with $\sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta)$, then

$$\sup_{\theta \in \Theta_0} R_{T_0}(\theta) \leq \sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta),$$

which contradicts the minimaxity of T when Θ_0 is the parameter space. Hence, T is minimax when Θ is the parameter space.

In Example 2.25, we actually applied the result in Theorem 4.12 to show the minimaxity of \bar{X} as an estimator of $\mu = EX_1$ when X_1, \dots, X_n are i.i.d. from a normal distribution with a known $\sigma^2 = \text{Var}(X_1)$, under the squared error loss.

To discuss the minimaxity of \bar{X} in the case where σ^2 is unknown, we need the following lemma.

Lemma 4.3

Let Θ_0 be a subset of Θ and T be a minimax estimator of ϑ when Θ_0 is the parameter space. Then T is a minimax estimator if

$$\sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta).$$

Proof

If there is an estimator T_0 with $\sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta)$, then

$$\sup_{\theta \in \Theta_0} R_{T_0}(\theta) \leq \sup_{\theta \in \Theta} R_{T_0}(\theta) < \sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta),$$

which contradicts the minimaxity of T when Θ_0 is the parameter space. Hence, T is minimax when Θ is the parameter space.

Example 4.19

Let X_1, \dots, X_n be i.i.d. from $N(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2)$.

Consider the estimation of μ under the squared error loss.

Suppose first that $\Theta = \mathcal{R} \times (0, c]$ with a constant $c > 0$.

Let $\Theta_0 = \mathcal{R} \times \{c\}$.

From Example 2.25, \bar{X} is a minimax estimator of μ when the parameter space is Θ_0 .

By Lemma 4.3, \bar{X} is also minimax when the parameter space is Θ .

Although σ^2 is assumed to be bounded by c , the minimax estimator \bar{X} does not depend on c .

Consider next the case where $\Theta = \mathcal{R} \times (0, \infty)$, i.e., σ^2 is unbounded.

Let T be any estimator of μ . For any fixed σ^2 ,

$$\frac{\sigma^2}{n} \leq \sup_{\mu \in \mathcal{R}} R_T(\theta),$$

since σ^2/n is the risk of \bar{X} that is minimax when σ^2 is known.

Letting $\sigma^2 \rightarrow \infty$, we obtain that $\sup_{\theta} R_T(\theta) = \infty$ for any estimator T .

Thus, minimaxity is meaningless (any estimator is minimax).

Theorem 4.13

Suppose that T as an estimator of ϑ has constant risk and is admissible.

Then T is minimax.

If the loss function is strictly convex, then T is the unique minimax estimator.

Proof

The risk of T is R_T (not depending on θ).

By the admissibility of T , if there is another estimator T_0 with $\sup_{\theta} R_{T_0}(\theta) \leq R_T$, then $R_{T_0}(\theta) = R_T$ for all θ .

This proves that T is minimax.

If the loss function is strictly convex and T_0 is another minimax estimator, then

$$R_{(T+T_0)/2}(\theta) < (R_{T_0} + R_T)/2 = R_T$$

for all θ and, therefore, T is inadmissible.

This shows that T is unique if the loss is strictly convex.

Theorem 4.13

Suppose that T as an estimator of ϑ has constant risk and is admissible.

Then T is minimax.

If the loss function is strictly convex, then T is the unique minimax estimator.

Proof

The risk of T is R_T (not depending on θ).

By the admissibility of T , if there is another estimator T_0 with $\sup_{\theta} R_{T_0}(\theta) \leq R_T$, then $R_{T_0}(\theta) = R_T$ for all θ .

This proves that T is minimax.

If the loss function is strictly convex and T_0 is another minimax estimator, then

$$R_{(T+T_0)/2}(\theta) < (R_{T_0} + R_T)/2 = R_T$$

for all θ and, therefore, T is inadmissible.

This shows that T is unique if the loss is strictly convex.