STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: Chuchu Wang and Gan Wu

Lecture #10

1 Admissibility of Minimax Estimator

Recall: δ^M is minimax if its maximum risk is minimal:

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

Admissibility implies maximaxity: if δ is admissible with constant risk, then δ is also minimax.

Proof 1 (Argument) *Let the constant risk of δ be r, then r is also the worst case risk of δ, as the risk is constant. Now if we assume that δ is not minimax, there exists a different estimator, say δ', such that it is minimax (with the corresponding risk as $r' < r$). But since this is the worst case risk of δ', it implies that the risk of δ' is lower than r throughout, and thus δ' dominates δ. This is a contradiction, as δ is admissible, which implies that δ is minimax.*

Note, however, that minimaxity does not guarantee admissibility (need to check case-by-case).

Example 1 *Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$, where σ^2 is known, and the parameter θ is the estimand. The minimax estimator is $\bar{X}$ under the squared error loss function.*

Question: is $\bar{X}$ admissible?
[A more general question: when is $a\bar{X} + b, a, b \in \mathbb{R}$ (any affine function of $\bar{X}$) admissible?]

Case 1 $0 < a < 1$*:*
Observe that $a\bar{X} + b$ is a convex combination of $\bar{X}$ and b. It is a Bayes estimator with respect to some Gaussian prior on θ. Since we are considering the squared error loss function, which is strictly convex, the Bayes estimator is unique. By Theorem 5.2.4 (TPE), $a\bar{X} + b$ is therefore admissible.

Case 2 $a = 0$*:*
In this case, b is also a unique Bayes estimator with respect to a degenerate prior distribution with unit mass at $\theta = b$. ($\Lambda(\theta) = N(b, 0^2)$). So by Theorem 5.2.4, b is admissible.

Case 3 $a = 1, b \neq 0$*:*
In this case, $\bar{X} + b$ is not admissible, because it is dominated by $\bar{X}$ because $\bar{X}$ has the same variance as $\bar{X} + b$, but is has a strictly smaller bias.

Case 4 $a > 1$*:*

$$\text{Risk of } a\bar{X} + b = \mathbb{E}[(a\bar{X} + b - \theta)^2] = \mathbb{E}[(a(\bar{X} - \theta) + b + \theta(a-1))^2] = \frac{a^2\sigma^2}{n} + (b + \theta(a-1))^2$$

So, when $a > 1$,

$$\mathbb{E}[(a\bar{X} + b - \theta)^2] \geq \frac{a^2\sigma^2}{n} > \frac{\sigma^2}{n} = R(\theta, \bar{X})$$

Hence, $\bar{X}$ dominates $a\bar{X} + b$ when $a > 1$, and so in this case $a\bar{X} + b$ is inadmissible.

Case 5 *$a < 0$:*

$$\mathbb{E}[(a\bar{X} + b - \theta)^2] > (b + \theta(a-1))^2 = (a-1)^2(\theta + \frac{b}{a-1})^2 > (\theta + \frac{b}{a-1})^2$$

and this is the risk of predicting the constant $-\frac{b}{a-1}$. So $-\frac{b}{a-1}$ dominates $a\bar{X} + b$. Hence, $a\bar{X} + b$ is inadmissible.

Case 6 *$a = 1, b = 0$:*
We use a limiting Bayes argument. Suppose $\bar{X}$ is inadmissible. WLOG, we assume that $\sigma^2 = 1$, and have

$$R(\theta, \bar{X}) = \frac{1}{n}$$

By our hypothesis, there must exist an estimator δ' such that $R(\theta, \delta') \leq \frac{1}{n}$ for all θ, and $R(\theta', \delta') < \frac{1}{n}$ for some $\theta' \in \Omega$ [at least one]. Because $R(\theta, \delta)$ is continuous in θ, there must exist $\epsilon > 0$ and an interval (θ_0, θ_1) containing θ' such that

$$R(\theta, \delta') < \frac{1}{n} - \epsilon, \forall \theta \in (\theta_0, \theta_1) \tag{*}$$

Let r'_τ be the average risk of δ' with respect to the prior distribution $N(0, \tau^2)$ on θ. Let r_τ be the average risk of a Bayes estimator δ_τ under the same prior.
Note that $\delta_\tau \neq \delta'$ because $R(\theta, \delta_\tau) \to \infty$ as $\theta \to \infty$, which is not consistent with $R(\theta, \delta') \leq \frac{1}{n}$ for all $\theta \in \Omega = \mathbb{R}$. So, $r_\tau < r'_\tau$ because the Bayes estimator is unique almost surely w.r.t. the marginal distribution of θ.

$$\frac{\frac{1}{n} - r'_\tau}{\frac{1}{n} - r_\tau} = \frac{\frac{1}{\sqrt{2\pi\tau^2}} \int_{-\infty}^{\infty} \{\frac{1}{n} - R(\theta, \delta')\} \exp(-\frac{\theta^2}{2\tau^2}) d\theta}{\frac{1}{n} - \frac{1}{n + \frac{1}{\tau^2}}} \tag{\#}$$

By (*), we can simplify (#) as follows

$$\frac{\frac{1}{n} - r'_\tau}{\frac{1}{n} - r_\tau} \geq \frac{\frac{1}{\sqrt{2\pi\tau^2}} \int_{\theta_0}^{\theta_1} \epsilon e^{-\frac{\theta^2}{2\tau^2}} d\theta}{\frac{1}{n(1+n\tau^2)}} = \frac{n(1+n\tau^2)}{\tau\sqrt{2\pi}} \epsilon \int_{\theta_0}^{\theta_1} e^{-\frac{\theta^2}{2\tau^2}} d\theta$$

As $\tau \to \infty$, the first expression $\frac{n(1+n\tau^2)\epsilon}{\tau\sqrt{2\pi}} \to \infty$, and since the integrand converges monotonically to 1, Lebesgue's monotone convergence theorem ensures that the integrand approaches to the quantity $\theta_1 - \theta_0$. So, for sufficiently large τ, we must have

$$\frac{\frac{1}{n} - r'_\tau}{\frac{1}{n} - r_\tau} > 1$$

This means that $r'_\tau < r_\tau$. But this is a contradiction because r_τ is the optimal average risk. So our assumption that there was a dominating estimator is incorrect, in which case $a\bar{X} + b = \bar{X}$ is admissible.

James-Stein (JS) Estimator (empirical Bayes)

Reference: TPE 5.4 - 5.5, simultaneous estimation
Let $X_1, \ldots, X_n$ with $X_i \sim N(\theta_i, \sigma^2)$ for $1 \leq i \leq n$. Our goal is to estimate $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ under the loss function

$$L(\boldsymbol{\theta}, \boldsymbol{d}) = \sum_{i=1}^{n} (d_i - \theta_i)^2$$

Then,

$$\delta_i(\boldsymbol{X}) = \max\left(1 - \frac{p-2}{\|\boldsymbol{X}\|_2^2}, 0\right) X_i$$

"OPTIMAL" INFERENCE (no uniform optimality) $\Longleftrightarrow$ Decision Theory

$$\begin{cases} \text{I. Constraint} \begin{cases} \text{UMVU (unbiasedness)} \\ \text{Equivariance (Ch.3, TPE)} \end{cases} & \textit{Related to data compression, sufficient statistics, testing, \ldots} \\ \text{II. Collapse} \begin{cases} \text{Average risk (Bayes)} \\ \text{Worst case (Minimax)} \end{cases} & \textit{Related properties: admissibility, uniqueness, \ldots} \end{cases}$$

Considered in finite sample case.

2 Testing of Statistical Hypotheses

2.1 Another decision problem: hypothesis testing

We assumme that the data is sampled according to $X \sim P_\theta$, where $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$.

Two disjoint subclasses of θ (hypotheses):

$$H_0 : \theta \in \Omega_0 \subset \Omega \ \text{ (null hypothesis)}$$
$$H_1 (\text{or } H_a) : \theta \in \Omega_1 = \Omega / \Omega_0 \ \text{ (alternative hypothesis)}$$

Decision space $\mathcal{D} = \{\text{Reject } H_0, \text{ not to reject } H_0 \text{ (or accept } H_0)\}$.
e.g. H_0 : no change in infection rate (default).
H_1 : improvements in lowering infection rate.

Truth / Decision	$\theta \in \Omega_0$	$\theta \in \Omega_1$
Reject H_0	1 Type I error (More severe)	0 (Good)
Accept H_0	0 (Good)	1 Type II error

Terminologies

- ⋆ Test function/critical function : $\phi(x) \in [0, 1]$

$$\phi(x) = P\left(\delta_\phi(x, u) = \text{Reject } H_0 | x\right)$$

where u is a uniform random variable independent of X.

- ⋆ Power function of a test ϕ is $\beta(\theta) = \mathbb{E}_\theta\left(\phi(X)\right) = P_\theta(\text{Reject } H_0)$.

Note:

If $\theta_0 \in \Omega_0$, then $\beta(\theta_0) = R(\theta_0, \delta_\phi)$ = Type I error.

For $\theta_1 \in \Omega_1$, then $\beta(\theta_1) = 1 - R(\theta_1, \delta_\phi)$ = 1-Type II error.

Our "ideal" optimality goal is to minimize $\beta(\theta_0)$ uniformly for all $\theta_0 \in \Omega_0$ and maximize $\beta(\theta_1)$ uniformly for all $\theta_1 \in \Omega_1$.

Neyman-Pearson Framework

Control the level of significance

$$\sup_{\theta_0 \in \Omega_0} \mathbb{E}_{\theta_0} \phi(X) = \sup_{\theta_0 \in \Omega_0} \beta(\theta_0) \leq \alpha$$

where $\sup_{\theta_0 \in \Omega_0} \beta(\theta_0)$ is called the size of the test.

Optimality Goal: Find a level α test that maximizes the power $\beta(\theta_1) = \mathbb{E}_{\theta_1}(\phi(X))$ for each $\theta_1 \in \Omega_1$. Such a test is called a uniformly powerful (UMP) test.

MP for the ”simple-vs-simple” case

Definition 7 *A hypothesis H_0 is called **simple** if $|\Omega_0| = 1$, otherwise it is called **composite**. This applies to H_1 as well.*

Hence a simple-vs-simple test:

$$H_0 : X \sim p_0 \;\; (p_0 = P_{\theta_0})$$
$$H_1 : X \sim p_1 \;\; (p_1 = P_{\theta_1})$$

Our goal is to find ϕ:

$$\max_{\phi} \mathbb{E}_{p_1}(\phi(X)) \quad \text{subject to} \;\; \mathbb{E}_{p_0}(\phi(X)) \leq \alpha$$

Lemma 1 (Neyman-Pearson Lemma) .

(i) Existence. For testing $H_0 : p_0$ vs $H_1 : p_1$, there exists a test $\phi(X)$ and a constant k such that

(a) $E_{p_0}(\phi(X)) = \alpha$ (size = level)

(b)

$$\phi(x) = \begin{cases} 1, & \textit{if } \frac{p_1(x)}{p_0(x)} > k \;\; [\textit{Rejection}] \\ 0, & \textit{otherwise} \;\; [\textit{Acceptance}] \end{cases}$$

such a test is called a likelihood ratio test.

(ii) Sufficiency: If a test satisfies (a) and (b) for some constant k, it is most powerful for testing $H_0 : p_0$ vs $H_1 : p_1$ at level α.

(iii) Necessity: If a test ϕ is MP at level α, then it satisfies (b) for some k, and it also satisfies (a) unless there exists a test if $size < \alpha$ with power 1.

Example 2 *Let $X_1, ..., X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ with σ^2 known. We want to test:*

$$H_0 : \mu = 0 \quad vs \quad H_1 : \mu = \mu_1$$

where μ_1 is given. We calculate the likelihood ratio:

$$r(x) = \frac{p_1(x)}{p_0(x)} = \cdots = \exp\left(\frac{1}{\sigma^2}\mu_1 \sum_{i=1}^{n} x_i - \frac{n\mu_1^2}{2\sigma^2}\right)$$

Suppose

$$
\begin{aligned}
r(x) > k &\iff \mu_1 \frac{\sum_{i=1} x_i}{\sigma^2} - \frac{n\mu_1^2}{2\sigma^2} > \log k \\
&\iff \mu_1 \sum_{i=1}^{n} x_i > k' \\
&\iff \begin{cases} \sum_{i=1}^{n} x_i > k'', & \textit{if } \mu_1 > 0 \\ \sum_{i=1}^{n} x_i < k''', & \textit{if } \mu_1 < 0 \end{cases}
\end{aligned}
$$

Let's focus on the case where $\mu_1 > 0$,

$$
\begin{aligned}
r(x) > k &\iff \sum_{i=1}^{n} x_i > k'' \\
&\iff \frac{\sqrt{n}\bar{x}}{\sigma} > k''''
\end{aligned}
$$

To calculate k''''(critical value), we need to evaluate

$$
\mathbb{E}_{p_0}(\phi(X)) = \alpha = P_{\mu=0}\left(\frac{\bar{X}}{\sigma/\sqrt{n}} > k''''\right)
$$

where $\frac{\bar{X}}{\sigma/\sqrt{n}}$ is normally distributed with zero mean.