Lecturer: Prof. McDonald Scribe: Kiran Kumar

Procedure for proving a estimate is minimax

- Derive an upper bound:

$$\exists\widehat{\theta}, \quad s.t., \quad R(\theta,\widehat{\theta}) \leq r_{upper}$$
$$\forall\theta \in \Theta$$

- Derive a lower bound

$$\forall\widehat{\theta}, \exists\theta \quad s.t., \quad R(\theta,\widehat{\theta}) \geq r_{lower}$$

If r_{upper} and r_{lower} are the same then we found the minimax estimate with respect to our risk function.

Theorem 22.1 (minimax risk is looser then worst-case bayes risk)**.**

$$R^* \geq R_B^* := \sup_{\pi\in M(\theta)} \inf_{\theta} E_\pi \left[E_{p_\theta}[l(\theta,\widehat{\theta}\right]]$$

Example

$$X \sim N(\theta, 1)$$
$$l(\theta,\widehat{\theta}) = \left(\theta - \widehat{\theta}\right)^2$$
$$\widehat{\theta} = X$$

Let,

$$\inf_{\widehat{\theta}} \sup_{\theta\in\Theta} E_\theta \left[l(\theta,\widehat{\theta})\right] \leq 1$$
$$\Pi(\theta) \sim N(0,\sigma^2)$$

Then, we can write,

$$E_\pi \left[E_{p(\theta)}[l(\theta,\widehat{\theta}]\right] = \sup_\pi \inf_{\widehat{\theta}} \left(\frac{\sigma^2}{\sigma^2+1}\right) \leq 1 \qquad ;\forall\sigma^2 > 0$$
$$if\sigma^2 \to \infty \implies \frac{\sigma^2}{\sigma^2+1} \to 1$$
$$R^*_{Bayes} \geq 1$$

Therefore, we can say that the minimax Risk is one.

Another perspective

$$R^* \geq R*_{Bayes} = \underbrace{\sup_{\pi \in M(\theta)} R^*_\pi}_{\text{Dual of } R^*}$$

This inequality is a weak duality, if we can show strong duality that would give us an equality.

Suppose Θ is a finite set, $|\Theta| < \infty$, then

$$R^* = \min_{\widehat{\theta}} \max_{\theta} \ E_\theta \left[l(\theta, \widehat{\theta}) \right]$$

is a convex function if l is convex. We can find a dual by minimization.

$$R^* = \min_{\widehat{\theta}, t} \ t \qquad s.t., E_\theta \left[l(\theta, \widehat{\theta}) \right] \leq t; \quad \forall \theta \in \Theta$$

Let $u \geq 0$, we can write the Lagrangian as

$$\begin{aligned} L(\widehat{\theta}, t, u) &= t + u^T \left[E_\theta [l(\theta, \widehat{\theta}) - t] \right] \\ &= (1 - u^T \mathbb{1})t + u^T E_\theta \left[l(\theta, \widehat{\theta}) \right] \end{aligned}$$

Here, L is ∞ unless the term $u^T \mathbb{1} = 1$. u is a probability distribution on Θ.

The dual problem can be written as

$$\begin{aligned} &\max_{\substack{u \geq 0 \\ u^T \mathbb{1} = 1}} \ \min_{\widehat{\theta}, t} \ L(\widehat{\theta}, t, u) \\ &= \max_{\pi \in M(\Theta)} \ \min_{\widehat{\theta}} \ R_\pi(\widehat{\theta}) \\ &= \max_{\pi \in M(\Theta)} R^*_\pi \end{aligned}$$

Maximum Likelihood

For parametric models MLE's are asymptotically minimax (under some conditions).

Consider the square error loss
$$R(\theta, \widehat{\theta}) = Var(\widehat{\theta}) + Bias^2(\widehat{\theta})$$
. This bias variance decomposition has an MLE which is bound as

$$Bias = O(n^{-2}), Var = O(n^{-1})$$

Under right regularity conditions (for fisher information $\mathbb{I}_\theta$)

$$Var(\widehat{\theta}_{MLE}) = \frac{C}{n\mathbb{I}_\theta}$$

As $n \to \infty$, variance dominates and for large n $MSE \approx Var$ which is the Cramer-Rao lower bound. Therefore the MSE will provide an efficient estimator or a minimax estimator.

The Hodges Estimator

Let $x_1, \ldots, x_n$ be iid $N(\theta, 1)$,

$$\widehat{\theta}_{MLE} = \overline{X}$$

Let $J_n = \left[-\frac{1}{n^{\frac{1}{4}}}, \frac{1}{n^{\frac{1}{4}}}\right]$

$$\tilde{\theta} = \begin{cases} \overline{X} & if\ \overline{X} \notin J_n \\ 0 & if\ \overline{X} \in J_n \end{cases}$$

There are two cases possible: Case 1, suppose $\theta \neq 0$, then choose $\epsilon > 0$ such that $I = (\theta - \epsilon, \theta + \epsilon)$ does not contain zero. By LLN $p(\overline{X} \in I) \to 1$. At the same time J_n shrinks with high probability.

Case 2, suppose $\theta = 0$ then $P(\overline{X} \in J - n) = P(|\overline{X}| \leq n^{-\frac{1}{4}}) = P(\sqrt{n}|\overline{X}| \leq n^{\frac{1}{4}}) = P(|N(0,1)| \leq n^{\frac{1}{4}}) \to 1$

For large n, $\tilde{\theta} = 0 = \theta$ with high probability.

James-Stein Estimator

Let $X \sim N_p(\theta, I - p)$, $\widehat{\theta}(x0 = x$ is minimax for $l(\theta, \widehat{\theta}) = \|\theta - \widehat{\theta}\|_2^2$

$X = \widehat{\theta}(x) = \underset{a}{\operatorname{argmin}}\|a - x\|_2^2$. This is for the minimax case, we look at the other case below.

$\widehat{\theta}_{JS}$ has the property that $\sup_\theta E\left[\|\theta - \widehat{\theta}_{js}\|_2^2\right] = \sup_\theta E\left[\|\theta - x\|_2^2\right]$

But for almost all θ $E\left[\|\theta - \widehat{\theta}_{js}\|_2^2\right] = E\left[\|\theta - x\|_2^2\right]$.

We can write $\widehat{\theta}_{js}(x) = \left(1 - \frac{p-2}{\|x\|_2^2}\right) x$

This gives us better risk everywhere as long as $p \geq 3$. Further, as $\|\theta\|_2^2 \to \infty$, $R(\theta, \widehat{\theta}_{js})$ converges upward to $R(\theta, x)$

We can shrink the estimator to any value $v \in R^p$ as below:

$$\widehat{\theta}_{js} = \left(1 - \frac{p-2}{\|x - v\|_2^2}\right)(x - v) + v$$