Continue from last time.

Definition 20.1. *Minimax estimator $\widehat{\theta}$ satisfies*

$$\begin{aligned} a_n &\geq \sup_\theta \mathbb{E}\Big[L(\theta,\widehat{\theta})\Big] \\ &\geq \inf_{\widehat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}\Big[L(\theta,\widehat{\theta})\Big] \geq b_n \\ &=: \mathcal{R}_n(\theta). \end{aligned}$$

$\widehat{\theta}$ is asymptotically minimax if $\sup_\theta \mathbb{E}\Big[L(\theta,\widehat{\theta})\Big] \sim \mathcal{R}_n(\theta)$ where $a_n \sim b_n$ (i.e., $\frac{a_n}{b_n} = O(1)$). Also, we say $\widehat{\theta}$ is rate minimax if $\sup_\theta \mathbb{E}\Big[L(\theta,\widehat{\theta})\Big] \asymp \mathcal{R}_n(\theta)$ where $a_n \asymp b_n$ (i.e., $\frac{a_n}{b_n} = O(1)$ and $\frac{b_n}{a_n} = O(1)$).

20.0.1 Bayes Estimators

Given estimator $\widehat{\theta}$, posterior risk of $\widehat{\theta}$ as

$$r(\widehat{\theta}|x^n) = \int L(\theta,\widehat{\theta})\Pi(\theta|x^n)d\theta,$$

Bayes estimator is defined to

$$\widehat{\theta}_B = \operatorname*{argmin}_{\theta\in\Theta} r(\theta|x^n).$$

Example 20.2. *Let $X_1,\ldots,X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta,\sigma^2)$. Prior on $\theta \sim \mathcal{N}(a,b^2)$ w.r.t. squared error loss.*

$$\widehat{\theta}_B = \frac{b^2}{b^2+\sigma^2/n}\cdot\overline{X} + \frac{\sigma^2/n}{b^2+\sigma^2/n}\cdot a \ (= \mathbb{E}[\theta|x^n]),$$

for any $\Pi(\theta|x^n)$.

- *If L is $|\cdot|$, $\widehat{\theta}_N$ is median.*
- *If L is 0-1 loss, $\widehat{\theta}_N$ is mode.*

Theorem 20.3. *Let $\widehat{\theta}_B$ be the Bayes Estimator for some posterior Π if*

$$\mathcal{R}(\theta,\widehat{\theta}_B) \leq \inf_\theta \int \underbrace{\mathcal{R}(\theta,\widehat{\theta})\Pi(\theta|x^n)}_{B_\Pi(\widehat{\theta})} d\theta,$$

then $\widehat{\theta}_B$ is minimax and the prior led to Π is called the least favorable prior.

Theorem 20.4. *Suppose $\widehat{\theta}_B$ is Bayes Estimator for some prior. If $\mathcal{R}(\theta,\widehat{\theta}_B)$ is constant in θ, then $\widehat{\theta}_B$ is minimax.*

Proof. Suppose $\int \mathcal{R}(\theta,\widehat{\theta})\Pi(\theta|x^n)d\theta = C$. Then $\mathcal{R}(\theta,\widehat{\theta}) \leq C$. Applying previous theorem, this theorem is proved. ∎

Example 20.5. *Suppose $X_1, \ldots, X_n \sim \mathrm{Bern}(\theta)$. We define two estimators*

$$\widehat{\theta}_1 \doteq \overline{X},$$

and

$$\widehat{\theta}_2 \doteq \frac{n\overline{X} + \frac{\sqrt{n}}{4}}{n + \sqrt{n}},$$

for $\Pi(\theta) \sim \mathrm{Beta}(\frac{\sqrt{n}}{4}, \frac{\sqrt{n}}{4})$. For squared error loss, $\widehat{\theta}_2$ is Bayes estimator for $\Pi(\theta)$. We have $\mathcal{R}(\theta, \widehat{\theta}_2) = \frac{n}{4(n+\sqrt{n})^2}$, which means $\widehat{\theta}_2$ is minimax. If $L(\theta, \widehat{\theta}) = \frac{(\theta-\widehat{\theta})^2}{\theta(1-\theta)}$, we have $\mathcal{R}(\theta, \widehat{\theta}_1) = \frac{1}{n}$ which means $\widehat{\theta}_1$ is minimax.

Example 20.6. *Suppose $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$. Then $\overline{X_n}$ is minimax for many loss functions. Its risk is $\frac{1}{n}$. $\frac{c}{n}$ is called "parametric rate".*

Example 20.7. *Suppose $X_1, \ldots, X_n \sim f \in \mathcal{F}$ where $\mathcal{F} = \{f, \int f = 1, \int f^2 < c_0\}$. And $c \cdot n^{-4/5} \leq \mathcal{R}_n(f, \widehat{f}) \leq C \cdot n^{-4/5}$. Then the loss function can be defined as $L(f, \widehat{f}) = \int (f - \widehat{f})^2 dx$.*

Definition 20.8. *A loss function L is "bowl shaped" if $\{x : L(x) \leq c\}$ is convex and symmetric about 0.*

Theorem 20.9. *If $X \sim \mathcal{N}_p(\Theta, \Sigma)$ and L is bowl shaped, then X is the unique minimax estimator for Θ.*

Example 20.10. *Suppose $X \sim \mathcal{N}(\theta, 1)$. But $\theta \in [-m, m]$ where $m \in (0, 1)$. Then under squared error loss, the unique minimax estimator is*

$$\widehat{\theta}(n) = m\left(\frac{e^{mx} - e^{-mx}}{e^{mx} + e^{-mx}}\right).$$