Week 4

Lecture 7

Optimal rate of convergence in the sup-norm

Model: Let $Y_1, Y_2, \ldots, Y_n$ i.i.d. on $[0, 1]$ with density $f \in \mathcal{F}_\alpha(M)$, Hölder ball of order $\alpha$.

Minimax rate: It can be shown that
\[
\inf_{\tilde{f}} \sup_{\mathcal{F}_\alpha(M)} E \left\| \tilde{f} - f \right\|_\infty^2 \approx C \left( \frac{n}{\log n} \right)^{-2\alpha/(2\alpha+1)}.
\]

For simplicity we will only show the case $\alpha = 1$ to reveal the “mysterious” log term.

Upper bound: In this lecture we will show there is an estimator $\tilde{f}$ such that
\[
\sup_{\mathcal{F}_\alpha(M)} E \left\| \tilde{f} - f \right\|_\infty^2 \leq C \left( \frac{n}{\log n} \right)^{-2\alpha/(2\alpha+1)}
\]
where $\alpha = 1$.

A histogram estimator:
Let $r = [1/h]$. Define
\[
\tilde{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{I((r-1)h \leq Y_i < rh)}{h}, \quad (r-1)h \leq x < rh.
\]

Denote
\[
Z_i = \frac{1}{h} I((r-1)h \leq Y_i < rh)
\]
\[
\tilde{f}(x) = E Z_i = \frac{1}{h} \int_{(r-1)h}^{rh} f(x) \, dx, \quad (r-1)h \leq x < rh
\]

Bias:
\[
\sup_x \left| f(x) - \tilde{f}(x) \right|^2 = \sup_r \left| f(x) - \frac{1}{h} \int_{(r-1)h}^{rh} f(x) \, dx \right|^2 \{(r-1)h \leq x < rh\}
\]
\[
\leq C h^2.
\]

Variance:
Since
\[
\tilde{f}(x) - \tilde{f}(x) \approx N \left( 0, \frac{\tilde{f}(x)}{nh} \right)
\]
which implies

\[
E \sup_x \left| \hat{f}(x) - \tilde{f}(x) \right|^2 = E \sup_r \left| \hat{f}(x) - \frac{1}{h} \int_{(r-1)h}^{rh} f(x') dx' \right|^2 \{ (r-1)h \leq x < rh \} \\
\leq C \frac{1}{nh} \log r.
\]

(more details will be given in class).

A trade-off between bias and variance gives

\[
E \sup_x \left| \hat{f}(x) - f(x) \right|^2 \leq C \left( \frac{n}{\log n} \right)^{-2/3}.
\]

Questions:

- What about general $\alpha$ in the sup-norm? Let $\hat{f}(x)$ be a kernel estimator. We have

\[
E \sup_x \left| \hat{f}(x) - f(x) \right|^2 \leq 2E \sup_x \left| \hat{f}(x) - E\hat{f}(x) \right|^2 + 2\sup_x \left| E\hat{f}(x) - f(x) \right|^2 \\
\leq C_{\alpha,M} \left( \frac{1}{nh} \log h^{-1} + h^{2\alpha} \right)
\]

- Adaptive estimation in the sup-norm?
Lecture 8

Lower bound for the sup-norm

Le Cam’s method – Two-point argument

We will introduce Le Cam’s method to derive minimax lower bounds. The essential of this approach is the Neyman-Pearson Lemma.

Let $P$ and $Q$ be two probability measures with densities $p$ and $q$ w.r.t. a measure $\mu$. The affinity between $P$ and $Q$ is defined as

$$\alpha_1 (P, Q) = \int p \wedge q d\mu = 1 - \frac{1}{2} \int |p - q| d\mu.$$  

Lemma 1

$$\inf_{f \geq 0, g \geq 0, f + g \geq 1} P_0 f + P_1 g \geq \alpha_1 (P_0, P_1)$$

Observe

$$Y \sim P_\theta.$$  

Let $f_\theta$ be the corresponding density function. Consider a problem of finding a lower bound for the minimax risk

$$\sup_{\theta \in \{\theta_0, \theta_1\}} \mathbb{E} L \left( \hat{\theta}, \theta \right) \geq ?$$

where $L \left( \hat{\theta}, \theta \right) = \mathbb{E} \left\| \hat{f} - f_\theta \right\|_\infty^2$ for the sup-norm case in this lecture. Let

$$d \left( \theta_0, \theta_1 \right) = \inf_t \left[ L \left( t, \theta_0 \right) + L \left( t, \theta_1 \right) \right].$$

Lemma 2

$$\sup_{\theta \in \{\theta_0, \theta_1, \ldots, \theta_r\}} \mathbb{E} L \left( \hat{\theta}, \theta \right) \geq \frac{1}{2} d \left( \theta_0, \theta_1 \right) \cdot \alpha_1 \left( P_{\theta_0}, P_{\theta_1} \right).$$

Le Cam’s method – Multiple comparisons

Let

$$d \left( \theta_0, \theta_1 \right) = \inf_t \left[ L \left( t, \theta_0 \right) + L \left( t, \theta_1 \right) \right] \quad \text{and} \quad d_{\text{min}} = \inf_{1 \leq i \leq r} d \left( \theta_0, \theta_i \right)$$

and

$$g_0 = L \left( \hat{\theta}, \theta_0 \right) / d_{\text{min}} \quad \text{and} \quad g_a = \inf_i L \left( \hat{\theta}, \theta_i \right) / d_{\text{min}}.$$  

Note that there is an $i$ such that

$$g_0 + g_a = \frac{L \left( \hat{\theta}, \theta_0 \right) + L \left( \hat{\theta}, \theta_i \right)}{d_{\text{min}}} \geq 1.$$
Then
\[
\sup_{\theta(\theta_0, \theta_1, \ldots, \theta_r)} \mathbb{E} L(\hat{\theta}, \theta) \geq \frac{1}{2r} \sum_{i=1}^{r} \left[ \mathbb{P}_{\theta_0} L\left(\hat{\theta}, \theta_0\right) + \mathbb{P}_{\theta_i} L\left(\hat{\theta}, \theta_i\right) \right] \\
\geq \frac{d_{\text{min}}}{2} \cdot \frac{1}{r} \sum_{i=1}^{r} \left[ \mathbb{P}_{\theta_0} \frac{L\left(\hat{\theta}, \theta_0\right)}{d_{\text{min}}} + \mathbb{P}_{\theta_i} \frac{L\left(\hat{\theta}, \theta_i\right)}{d_{\text{min}}} \right] \\
\geq \frac{1}{2} d_{\text{min}} \left[ \mathbb{P}_{\theta_0} g_0 + \left( \frac{1}{r} \sum_{i=1}^{r} \mathbb{P}_{\theta_i} \right) g_a \right] \geq \frac{1}{2} d_{\text{min}} \cdot \alpha_1 \left( \mathbb{P}_{\theta_0}, \frac{1}{r} \sum_{i=1}^{r} \mathbb{P}_{\theta_i} \right)
\]

Lemma 3
\[
\sup_{\theta(\theta_0, \theta_1, \ldots, \theta_r)} \mathbb{E} L(\hat{\theta}, \theta) \geq \frac{1}{2} d_{\text{min}} \cdot \alpha_1 \left( \mathbb{P}_{\theta_0}, \frac{1}{r} \sum_{i=1}^{r} \mathbb{P}_{\theta_i} \right).
\]

**Application to lower bound under the sup-norm:**

Let \( h_n = \left( \frac{n}{\log n} \right)^{-1/3} \). For a fixed probability density \( f_0 \) and a fixed \( K \) with support on \((-1, 1)\), define
\[
f_r = \begin{cases} 
1 + h_n \frac{(x - (r-1/2)h)}{h_{n/2}} & (r-1)h \leq x < rh \\
1 & \text{otherwise}
\end{cases}
\]
where \(|G|\) is bounded and \( \int G = 0 \) (more properties need to be satisfied. What are they?)

Let \( f_0 \) be the density of uniform. Show
\[
\alpha_1 \left( \mathbb{P}_0, \frac{1}{r} \sum_{i=1}^{r} \mathbb{P}_i \right) \geq c, \text{ for some } c > 0
\]
where \( \mathbb{P}_i \) has density \( f_i \) (more details in class).

**Questions:**

- What about estimation at a point \( x_0 \)? Is it true
\[
\inf_{f} \sup_{f \in \mathcal{F}_n(M)} \mathbb{E} \left| \hat{f}(x_0) - f(x_0) \right|^2 \asymp n^{-2\alpha/(2\alpha+1)}?
\]

**Notation:**
\[
a_n \asymp b_n \iff \exists c_0, C > 0 \text{ s.t. } c_0 \leq a_n / b_n \leq C.
\]

- What about adaptive estimation at a point \( x_0 \)? Is it true
\[
\inf_{f} \sup_{f \in \mathcal{F}_n(M)} \mathbb{E} \left( \hat{f}(x_0) - f(x_0) \right)^2 \asymp \left( \frac{n}{\log n} \right)^{-2\alpha/(2\alpha+1)}?
\]

Lepski’s method will do for the upper bound. Should we introduce Brown-Low inequality for the lower bound?
Adaptive estimation in the sup-norm? Is it true

\[
\inf_{f} \sup_{f \in \cup_{a<n<b} \mathcal{F}_{n}(M)} \mathbb{E} \sup_{x} \left| \hat{f}(x) - f(x) \right|^2 \asymp \left( \frac{n}{\log n} \right)^{-2\alpha/(2\alpha+1)}
\]

Lepski’s method?

Next lecture other lower bound arguments such as Tsybakov’s lower bound to obtain the minimax rate?