Week 11
April 10 – April 12


Let $P$ and $Q$ be two probability measures with densities $p$ and $q$ w.r.t. measure $\mu$. Define

$$\|P \land Q\| = \int p \land q d\mu.$$  

Lemma:

$$\inf_f P_0 f + P_1 (1 - f) = \|P_0 \land P_1\|, \ 0 \leq f \leq 1.$$  

Proof: Let $p_0$ and $p_1$ be probability densities of $P$ and $Q$ respectively. The result follows from the following equation

$$\int (p_0 - p_1) (f - I (p_0 < p_1)) d\mu \geq 0$$

i.e.,

$$\int p_0 f + p_1 (1 - f) d\mu \geq \int p_0 I (p_0 < p_1) + p_1 d\mu.$$  

A corollary for this lemma is

$$\inf_{f \geq 0, g \geq 0, f + g \geq 1} P_0 f + P_1 g \geq \|P_0 \land P_1\|$$

Remark: Neyman-Pearson test. Let $f$ be any rejection region such that $P_0 f \leq \alpha$. Find $c$ such that $P_0 I (p_0 < cp_1) = \alpha$, then

$$\int (p_0 - cp_1) (f - I (p_0 < cp_1)) d\mu \geq 0$$

which implies

$$0 \geq P_0 (f - I (p_0 < cp_1)) d\mu \geq cP_1 (f - I (p_0 < cp_1))$$

An application: Le Cam’s method.

Example: Show that the minimax rate in estimating $\theta$ for i.i.d. $U (0, \theta)$ is $1/n$ for the squared error loss, where $\theta \in [a, b]$ with $a < b$.

Let $\hat{\theta}$ an estimator of $\theta$. We need to show for some $c > 0$

$$\sup_{\theta} \mathbb{E} (\hat{\theta} - \theta)^2 \geq c \frac{1}{n^2}$$

We know

$$\sup_{\theta} \mathbb{E} (\hat{\theta} - \theta)^2 \geq \sup_{\theta \in \{\theta_1, \theta_2\}} \mathbb{E} (\hat{\theta} - \theta)^2$$

$$\geq \frac{1}{2} \mathbb{E}_{X|\theta_1} (\hat{\theta} - \theta_1)^2 + \frac{1}{2} \mathbb{E}_{X|\theta_2} (\hat{\theta} - \theta_2)^2.$$
Since 
\[ (\hat{\theta} - \theta_1)^2 + (\hat{\theta} - \theta_2)^2 \geq \frac{1}{2} (\theta_1 - \theta_2)^2 \]
i.e.,
\[ \frac{1}{2} (\theta_1 - \theta_2)^2 + \frac{1}{2} (\theta_1 - \theta_2)^2 \geq 1, \]
we have
\[ \sup_{\theta} E \left( \hat{\theta} - \theta \right)^2 \geq \frac{(\theta_1 - \theta_2)^2}{4} \| P_{\theta_1} \land P_{\theta_2} \| . \]

Let \( \theta_1 = 1 \) and \( \theta_2 = \theta_1 + \frac{1}{n} \). By the lemma below it is easy to show
\[ \| P_{\theta_1} \land P_{\theta_2} \| \geq \frac{1}{2} \left( \prod_{i=1}^{n} \int_{[0,1]} (x) \frac{1}{1+1/n} I_{[0,1+\frac{1}{n}]} (x) \, dx \right)^2 = \frac{1}{2} \left( 1 + \frac{1}{n} \right)^{-2n} \rightarrow \frac{1}{2} e^{-2}. \]

**Lemma**
\[ \left( \int \sqrt{pq} \right)^2 \leq 2 \int (p \land q). \]

Proof. It is easy to see
\[ pq \leq (p \lor q) (p \land q) \leq (p + q) (p \land q) \]
then the Cauchy-Schwarz inequality implies
\[ \left( \int \sqrt{pq} \right)^2 \leq \left( \int (p + q) (p \land q) \right)^2 \leq \int (p + q) \int (p \land q) = 2 \int (p \land q). \]

**Homework** Let \( Y_1, Y_2, \ldots, Y_n \) i.i.d. on \([0,1]\) with density \( f \in \mathcal{F} \), \( \mathcal{F} = \{ f, \int (f^{(m)} (x))^2 \, dx \leq M \} \) with \( m = 2 \). Define
\[ \hat{f}_h (x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K \left( \frac{Y_i - x}{h} \right) \]
where \( K (y) = I (-1/2 \leq y < 1/2) \). Let \( h = n^{-1/(2m+1)} \). Show that
\[ \sup_{f \in \mathcal{F}} E \int \left( \hat{f}_h (x) - f \right)^2 \leq C_M n^{-2m/(2m+1)} \]
Hint: show that
\[ E \hat{f}_n (x) - f (x) = \int K (y) [f (x + hy) - f (x)] \, dy = h^2 \int \int_0^1 K (y) y^2 f^2 (x + sy) (1 - s) \, ds \, dy \]
\[ \text{var} \left( \hat{f}_n (x) \right) \leq \frac{1}{nh^2} E \left[ K \left( \frac{Y_i - x}{h} \right) \right]^2 = \frac{1}{nh} \int K^2 (y) f (x + hy) \, dy. \]

Model: Let $Y_1, Y_2, \ldots, Y_n$ i.i.d. on $[0,1]$ with density $f \in \mathcal{F}$, $\mathcal{F} = \left\{ f, \int (f^{(m)}(x))^2 \, dx \leq M \right\}$.

It can be shown that there is a kernel estimator $\hat{f}_n$ such that

$$\sup_{f \in \mathcal{F}} \int E \left( \hat{f}_n(x) - f(x) \right)^2 \leq C_n^{-2m/(2m+1)}.$$

**Question:** Is this rate optimal? Can you improve this rate?

**Idea.** The answer for the question is No. We will show

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \int E \left( \hat{f}_n(x) - f(x) \right)^2 \geq C_1 n^{-2m/(2m+1)},$$

for some $C_1 > 0$.

Because it is hard to analyze the whole ball $\mathcal{F}$, we will construct a simpler ball $\mathcal{F}_1 \subset \mathcal{F}$ such that

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{F}_1} \int E \left( \hat{f}_n(x) - f(x) \right)^2 \geq C_1 n^{-2m/(2m+1)}.$$

For $h_n = n^{-1/(2m+1)}$ and $r_n = 1/h_n$, let $x_{n,1} < x_{n,2} < \ldots < x_{n,r_n}$ with $x_{n,i} = -1 + (2i-1)h_n$. The meshwidth is $2h_n$. For a fixed probability density $f_0$ and a fixed $K$ with support on $(-1,1)$, define

$$\mathcal{F}_1 = \left\{ f_{n,\theta}(x) : f_{n,\theta}(x) = f_0(x) + h_n^m \sum_{j=1}^{r_n} \theta_j K \left( \frac{x - x_{n,j}}{h_n} \right), \theta \in \{0,1\}^{r_n} \right\}.$$

Let $f_0(x) = \frac{1}{2} I_{[-1/2,1/2]}(x)$ is bounded away from 0. Assume that $|K|$ is bounded and $\int K = 0$, then $\int f_{n,\theta}(x) = 1$, and $f_{n,\theta}(x) \geq 0$ as $n$ sufficiently large, and

$$\int \left[ f_{n,\theta}^{(m)}(x) \right]^2 \leq h_n r_n \int \left[ K^{(m)}(x) \right]^2 = \int \left[ K^{(m)}(x) \right]^2,$$

which could very small by replacing $K$ with $cK$, $c$ small.

We will show

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{F}_1} \int E \left( \hat{f}_n(x) - f(x) \right)^2 \geq C_1 n^{-2m/(2m+1)}$$

or write

$$\inf_{\hat{f}_n} \sup_{\theta \in \{0,1\}^{r_n}} \int E \left( \hat{f}_n(x) - f_{n,\theta}(x) \right)^2 \geq C_1 n^{-2m/(2m+1)}.$$

**Proof.**
Let \( H(\theta, \theta') = \sum_{i=1}^{r} |\theta_i - \theta'_i| \) be the Hamming distance on \( \{0, 1\}^r \), which counts the number of positions at which \( \theta \) and \( \theta' \) differ. The Lemma below implies

\[
\int \left[ \prod_{i=1}^{n} f_{n, \theta}(x_i) \wedge \prod_{i=1}^{n} f_{n, \theta'}(x_i) \right] \geq \frac{1}{2} \left( \prod_{i=1}^{n} \int \sqrt{f_{n, \theta}(x_i) f_{n, \theta'}(x_i)} \right)^2
\]

\[
= \frac{1}{2} \left( 1 - \frac{1}{2} \int \left( \sqrt{f_{n, \theta}(x_i) - \sqrt{f_{n, \theta'}(x_i)}} \right)^2 \right) 2^n
\]

and it is easy to see

\[
\int (f_{n, \theta}(x) - f_{n, \theta'}(x))^2 = h_n^{2m} \sum_{j=1}^{r_n} \int |\theta_j - \theta'_j| K^2 \left( \frac{x - x_{n,j}}{h_n} \right) = h_n^{2m+1} H(\theta, \theta') \int K^2,
\]

which implies

\[
\min_{H(\theta, \theta') = 1} \int \left[ \prod_{i=1}^{n} f_{n, \theta}(x_i) \wedge \prod_{i=1}^{n} f_{n, \theta'}(x_i) \right] \geq C_2, C_2 > 0
\]

This suggests, of course without a proof, that

\[
\inf_{f_n} \sup_{f \in F_1} E \left[ \int (\hat{f}_n(x) - f_{n, \theta}(x))^2 \right] \geq c r_n \int_{H(\theta, \theta') = 1} \left[ \prod_{i=1}^{n} f_{n, \theta}(x_i) \wedge \prod_{i=1}^{n} f_{n, \theta'}(x_i) \right].
\]

For any estimator \( \hat{f}_n \), define

\[
S = \arg \min_{\theta \in \Theta} \int (\hat{f}_n(x) - f_{n, \theta}(x))^2
\]

then we have

\[
4E \int (\hat{f}_n(x) - f_{n, \theta}(x))^2 \geq 2E \int (\hat{f}_n(x) - f_{n, S}(x))^2 + 2E \int (\hat{f}_n(x) - f_{n, \theta}(x))^2
\]

\[
\geq E \int (f_{n, \theta}(x) - f_{n, S}(x))^2
\]

\[
= h_n^{2m+1} \int K^2 E(\theta, S)
\]

We will show that

\[
\max_{\theta} E_{\theta} H(S, \theta) \geq \frac{r_n}{2} \min_{H(\theta, \theta') = 1} \| P_\theta \wedge P_{\theta'} \|
\]

where

\[
P_\theta(dx) = \prod_{i=1}^{n} f_{n, \theta}(x_i)
\]
Note that

\[
\max_\theta E_\theta H (S, \theta) = \max_\theta E_\theta \sum_{i=1}^r |S_i - \theta_i|
\]

\[
\geq \frac{1}{2^r} \sum_\theta \sum_{i=1}^r |S_i - \theta_i|
\]

\[
= \frac{1}{2} \sum_{i=1}^r \left[ \frac{1}{2^{r-1}} \sum_{\theta, \theta_i=0} \int S_i dP_\theta + \frac{1}{2^{r-1}} \sum_{\theta, \theta_i=1} \int (1 - S_i) dP_\theta \right]
\]

\[
= \frac{1}{2} \sum_{i=1}^r \left[ \int S_i d\left( \frac{1}{2^{r-1}} \sum_{\theta, \theta_i=0} P_\theta \right) + \int (1 - S_i) \left( \frac{1}{2^{r-1}} \sum_{\theta, \theta_i=1} P_\theta \right) \right]
\]

\[
\geq \frac{1}{2} \sum_{i=1}^r \| P_{0,j} \wedge P_{1,j} \|
\]

\[
\geq \frac{1}{2} \sum_{i=1}^r \min_{H(\theta, \theta') = 1} \| P_\theta \wedge P_{\theta'} \|
\]

where \( P_{0,j} = \frac{1}{2^{r-1}} \sum_{\theta, \theta_i=0} P_\theta \), and \( P_{1,j} = \frac{1}{2^{r-1}} \sum_{\theta, \theta_i=1} P_\theta \). The last inequality is due to the fact that

\[
\| P_m \wedge Q_m \| \geq \frac{1}{m} \sum_i \| P_i \wedge Q_i \|
\]

and the \( 2^{r-1} \) terms \( P_\theta \) and \( P_{\theta'} \) can be arranged such that each \( \theta \) and \( \theta' \) differs only in their \( j \)th coordinate.