Week 8
Spring 2009

Lecture 13. Assouad's Lemma

The Assouad's Lemma will be applied to study functional regression and large covariance matrices estimation. We introduce this lemma by considering rate optimality of the classical Gaussian sequence model.

Example 1 *Gaussian sequence model. We consider a general setting*

$$y_i = \theta_i + \frac{1}{\sqrt{n}} z_i,\ z_i \overset{i.i.d.}{\sim} N(0,1),\ \theta \in \Theta$$

where Θ *is an ellipsoid,*

$$\Theta = \left\{ \theta : \sum_i i^{2\alpha} \theta_i^2 \le M \right\}.$$

It is very easy to show that there is an estimator $\hat{\theta}$ *such that*

$$\sup_{\Theta} \mathbb{E} \left\| \hat{\theta} - \theta \right\|^2 \le C_{\alpha,M} n^{-2\alpha/(2\alpha+1)}.$$

How to show that

$$\inf_{\hat{\theta}} \sup_{\Theta} \mathbb{E} \left\| \hat{\theta} - \theta \right\|^2 \ge c_{\alpha,M} n^{-2\alpha/(2\alpha+1)}\,?$$

It is enough to show

$$\inf_{\hat{\theta}} \sup_{\Theta_{sub}} \mathbb{E} \left\| \hat{\theta} - \theta \right\|^2 \ge c_{\alpha,M} n^{-2\alpha/(2\alpha+1)}$$

where

$$\Theta_{sub} = \left\{ \theta : \theta_i = \gamma_i \frac{a}{\sqrt{n}},\ \gamma_i = 0 \text{ or } 1,\ \text{for } 1 \le i \le n^{\frac{1}{1+2\alpha}};\ \theta_i = 0, \text{ for } i > n^{\frac{1}{1+2\alpha}} \right\}$$

with $0 < a \le M^{1/2}$*. Our intuition tells us it is hard for us to test* $\gamma_i = 0$ *or* 1 *if* a *is small. Then we expect a total error of* $cn^{\frac{1}{1+2\alpha}} \cdot \frac{1}{n} = cn^{-2\alpha/(2\alpha+1)}$ *to estimate* θ *under the squared error loss. This intuition can be justified rigorously by the Assouad's lemma.*

The Assouad's lemma gives a lower bound for the maximum risk over the parameter set $\Lambda = \{0,1\}^r$, in an abstract form, applicable to the problem of estimating an arbitrary quantity $\psi(\gamma)$, belonging to a metric space space with metric d. It can be seen as an extension of Le Cam's method for which r is

1. Let $H(\gamma,\gamma') = \sum_{i=1}^{r} |\gamma_i - \gamma_i'|$ be the Hamming distance on $\{0,1\}^r$, which counts the number of positions at which γ and γ' differ. In the next lecture we will apply this lemma to the functional linear regression.

Assouad's Lemma. For any estimator T based on an observation in the experiment $\{\mathbb{P}_\gamma, \gamma \in \Lambda\}$, and any $p > 0$

$$\max_\gamma 2^p \mathbb{E}_\gamma d^p(T, \psi(\gamma)) \geq \min_{H(\gamma,\gamma')\geq 1} \frac{d^p(\psi(\gamma), \psi(\gamma'))}{H(\gamma,\gamma')} \frac{r}{2} \min_{H(\gamma,\gamma')=1} \|\mathbb{P}_\gamma \wedge \mathbb{P}_{\gamma'}\|$$

Proof. Let

$$\alpha = \min_{H(\gamma,\gamma')\geq 1} \frac{d^p(\psi(\gamma), \psi(\gamma'))}{H(\gamma,\gamma')}.$$

(α satisfies $d^p(\psi(\gamma), \psi(\gamma')) \geq \alpha H(\gamma,\gamma')$ for all γ,γ'). Define an estimator S, taking its value in $\Lambda = \{0,1\}^r$, by letting $S = \gamma$ if $\gamma' \to d^p(T, \psi(\gamma'))$ is minimal over Λ at $\gamma' = \gamma$, i.e.,

$$S = \arg\min_{\gamma\in\Lambda} d^p(T, \psi(\gamma))$$

(If the minimum is not unique, choose a point of minimum in a consistent way.) The triangle inequality gives

$$\alpha \mathbb{E}_\gamma H(S,\gamma) \leq \mathbb{E}_\gamma d^p(\psi(S), \psi(\gamma)) \leq \mathbb{E}_\gamma [d(\psi(S), T) + d(T, \psi(\gamma))]^p \leq \mathbb{E}_\gamma [2d(\psi(T), \psi(\gamma))]^p.$$

Now its enough to show

$$\max_\gamma \mathbb{E}_\gamma H(S,\gamma) \geq \frac{r}{2} \min_{H(\gamma,\gamma')=1} \|\mathbb{P}_\gamma \wedge \mathbb{P}_{\gamma'}\|.$$

Note that

$$\begin{aligned}
\max_\gamma \mathbb{E}_\gamma H(S,\gamma) &= \max_\gamma \mathbb{E}_\gamma \sum_{i=1}^{r} |S_i - \gamma_i| \\
&\geq \frac{1}{2^r} \sum_\gamma \sum_{i=1}^{r} |S_i - \gamma_i| \\
&= \frac{1}{2} \sum_{i=1}^{r} \left[\frac{1}{2^{r-1}} \sum_{\gamma,\gamma_i=0} \int S_i d\mathbb{P}_\gamma + \frac{1}{2^{r-1}} \sum_{\gamma,\gamma_i=1} \int (1-S_i)\, d\mathbb{P}_\gamma \right] \\
&= \frac{1}{2} \sum_{i=1}^{r} \left[\int S_i d\left(\frac{1}{2^{r-1}} \sum_{\gamma,\gamma_i=0} \mathbb{P}_\gamma \right) + \int (1-S_i)\, d\left(\frac{1}{2^{r-1}} \sum_{\gamma,\gamma_i=1} \mathbb{P}_\gamma \right) \right] \\
&\geq \frac{1}{2} \sum_{i=1}^{r} \left\| \overline{\mathbb{P}}_{0,j} \wedge \overline{\mathbb{P}}_{1,j} \right\|,
\end{aligned}$$

where $\overline{\mathbb{P}}_{0,j} = \frac{1}{2^{r-1}} \sum_{\gamma,\gamma_i=0} \mathbb{P}_\gamma$, and $\overline{\mathbb{P}}_{1,j} = \frac{1}{2^{r-1}} \sum_{\gamma,\gamma_i=1} \mathbb{P}_\gamma$. The 2^{r-1} terms $\mathbb{P}_\gamma$ and $\mathbb{P}_{\gamma'}$ can be arranged such that each γ and γ' differs only in their j-th coordinate. Then the lemma below immediately implies

$$\max_{\gamma} \mathbb{E}_{\gamma} H(S, \gamma) \geq \frac{1}{2} \sum_{i=1}^{r} \min_{H(\gamma,\gamma')=1} \|\mathbb{P}_{\gamma} \wedge \mathbb{P}_{\gamma'}\| = \frac{r}{2} \min_{H(\gamma,\gamma')=1} \|\mathbb{P}_{\gamma} \wedge \mathbb{P}_{\gamma'}\|$$

■

Lemma:

$$\left\|\overline{\mathbb{P}}_m \wedge \overline{\mathbb{Q}}_m\right\| \geq \frac{1}{m} \sum \|\mathbb{P}_i \wedge \mathbb{Q}_i\| .$$

This lemma is due to the trivial fact that

$$\min \left\{ \frac{1}{m} \sum_{i=1}^{m} a_i, \frac{1}{m} \sum_{i=1}^{m} b_i \right\} \geq \frac{1}{m} \sum_{i=1}^{m} \min \{a_i, b_i\} .$$

Lecture 14. Assouad's Lemma and minimax lower bound for functional linear regression

The Assouad's lemma gives a lower bound for the maximum risk over the parameter set $\Lambda = \{0,1\}^r$, in an abstract form, applicable to the problem of estimating an arbitrary quantity $\psi(\gamma)$, belonging to a metric space space with metric d. Let $H(\gamma,\acute{\gamma}) = \sum_{i=1}^{r} |\gamma_i - \acute{\gamma_i}|$ be the Hamming distance on $\{0,1\}^r$, which counts the number of positions at which γ and $\acute{\gamma}$ differ. In this lecture we will apply this lemma to the functional linear regression.

Assouad's Lemma. For any estimator T based on an observation in the model $\{\mathbb{P}_\gamma, \gamma \in \Lambda\}$, and any $p > 0$

$$\max_{\gamma} 2^p \mathbb{E}_\gamma d^p(T, \psi(\gamma)) \geq \min_{H(\gamma,\acute{\gamma})\geq 1} \frac{d^p(\psi(\gamma), \psi(\acute{\gamma}))}{H(\gamma,\acute{\gamma})} \frac{r}{2} \min_{H(\gamma,\acute{\gamma})=1} \|\mathbb{P}_\gamma \wedge \mathbb{P}_{\acute{\gamma}}\|.$$

Functional linear regression

Assume that data pairs $(Y_i, X_i(t))$ for $i = 1, 2, \cdots, n$ are i.i.d. with

$$Y_i = a + \int_0^1 b(t)X_i(t)dt + \xi_i \quad 1 \leq i \leq n \tag{1}$$

where $X_i(t)$'s are i.i.d. Gaussian processes and $\xi_i \sim N(0,1)$. The main task is to estimate the slope function $b(t)$.

The distribution of a gaussian process $X(t)$ is uniquely determined by its mean process $\mu(t) = \mathbb{E}X(t)$ and covariance kernel $K(s,t) = \mathbb{E}Z(s)Z(t)$, where $Z(t) = X(t) - \mu(t)$. If the covariance kernel K is in $L^2([0,1]^2)$, it has a L^2-spectral decomposition,

$$K(s,t) = \sum_{j=1}^{+\infty} \theta_j \phi_j(s)\phi_j(t) \tag{2}$$

By convention, the eigenvalues are arranged in decreasing order, $\theta_1 \geq \theta_2 \geq \cdots \geq 0$. The eigenfunctions $\phi_1, \phi_2, \cdots$ form a complete orthonormal basis of $L^2([0,1])$ of real-valued functions that are square integrable with respect to Lebesgue measure on $[0,1]$. Note that the contribution from $\mu(t)$ can be absorbed into the intercept, so that (1) becomes

$$Y_i = b_0 + \int_{\mathcal{T}} b(t)Z_i(t)dt + \xi_i, \quad \text{with} \quad b_0 = a + \int_{\mathcal{T}} b(t)\mu(t)dt. \tag{3}$$

Condition 2 *Let $\beta > 0$ and $M_i > 0$ for $i = 0, 1$. Define the function class for b by*

$$b = \sum_{j=1}^{\infty} b_j\,\phi_j, \ \text{with } |b_j| \leq M_1\, j^{-\beta}, \text{ for all } j = 1, 2, ... \tag{4}$$

We can interpret this as a "smoothness class" of functions, where the functions become "smoother" (measured in the sense of generalized Fourier expansions in the basis $\left\{\phi_j\right\}$) as β increases. We shall also assume the eigenvalues satisfy

$$M_0^{-1}\, j^{-\alpha} \leq \theta_j \leq M_0\, j^{-\alpha} \tag{5}$$

Let $\mathcal{F}\left(\alpha, \beta, M_0, M_1\right)$ denote the set of distributions F of (X, Y) that satisfies (4) and (5).

The key idea for the upper bound.

Under some assumptions in addition to condition 2 Hall and Horowitz (2006) obtained a rate of convergence to estimate b,

$$\sup_{F\in\mathcal{F}(\alpha,\beta,M_0,M_1)} \mathbb{E}\int_{\mathcal{T}} (\hat{b}(t)-b(t))^2 dt \leq C n^{-(2\beta-1)/(\alpha+2\beta)}.$$

Now we give an explanation of this strange rate by assuming that K is known. Write

$$Y_i = b_0 + \sum_{j=1}^{\infty} b_j Z_{ij} + \xi_i$$

where $Z_{ij} = \int Z_i(t)\phi_j\left(t\right) dt$. Note that

$$\begin{aligned} \mathbb{E}Z_{ij}^2 &= \int\int K(s,t)\phi_j(s)\phi_j(t)dsdt = \theta_j \sim j^{-\alpha} \\ \mathbb{E}Z_{ij}Z_{ij'} &= \int\int K(s,t)\phi_j(s)\phi_{j'}(t)dsdt = 0. \end{aligned}$$

Let

$$\hat{b}_j = \frac{\theta_j^{-1}}{n}\sum_{i=1}^{n} Y_i Z_{ij}.$$

It is easy to show

$$\mathbb{E}\hat{b}_j = b_j.$$

Note that $\mathbb{E}\left[\mathbb{V}ar\left(\hat{b}_j|Z_{ij}\right)\right] = \frac{\theta_j^{-1}}{n}$ which may suggest $\mathbb{V}ar\left(\hat{b}_j\right) \sim j^{\alpha}/n$. Define

$$\hat{b}\left(t\right) = \sum_{j=1}^{k} \hat{b}_j \phi_j(t)$$

then

$$\mathbb{E}\int_{\mathcal{T}} (\hat{b}(t)-b(t))^2 dt \leq C\frac{k^{\alpha+1}}{n} + Ck^{-2\beta+1}.$$

The rate $n^{-(2\beta-1)/(\alpha+2\beta)}$ is the obtained by setting $k = n^{\frac{1}{\alpha+2\beta}}$.

Theorem 3 *Under the condition above we have*

$$\inf_{\hat{b}} \sup_{F\in\mathcal{F}(\alpha,\beta,M_0,M_1)} \mathbb{E}\int_{\mathcal{T}} (\hat{b}(t)-b(t))^2 dt \geq cn^{-(2\beta-1)/(\alpha+2\beta)}$$

for some c depending on α, β, M_0 *and* M_1.

Proof. We first define a subset $\mathcal{F}_n$ of $\mathcal{F}(\alpha,\beta,M_0,M_1)$. Let $a=0$, $\mu(t)\equiv 0$, and the covariance kernel $K_0(s,t)=\sum_{j\geq 1}\theta_j\phi_j(s)\phi_j(t)$ and the eigenvalues $\theta_j=j^{-\alpha}$, for $j\geq 1$. Let

$$b_\gamma(t)=\sum_{L_n<j\leq 2L_n} c_1 j^{-\beta}\gamma_j\phi_j(t).$$

where $L_n=c_0n^{1/(\alpha+2\beta)}$. Note that there is a one to one correspondence between $\mathcal{F}_n$ and $\mathcal{W}_n=[0,1]^{L_n}$. It is easy to verify $\mathcal{F}_n\subset\mathcal{F}$.

By Assouad's lemma with $p=2$, it follows that

$$\max_{\gamma\in\mathcal{W}_n}\mathbb{E}\int_{\mathcal{T}}(\hat{b}(t)-b(t))^2dt\geq\frac{c_1^2}{2}\min_{h(\gamma,\gamma')\geq 1}\frac{\sum_{j\in W_n}j^{-2\beta}(\gamma_j-\gamma_j')^2}{h(\gamma,\gamma')}\cdot L_n\cdot\min_{h(\gamma,\gamma')=1}||\mathbb{P}_\gamma\wedge\mathbb{P}_{\gamma'}|| \tag{6}$$

It is easy to see

$$\min_{h(\gamma,\gamma')\geq 1}\frac{\sum_{j\in W_n}j^{-2\beta}(\gamma_j-\gamma_j')^2}{h(\gamma,\gamma')}=\min_{h(\gamma,\gamma')\geq 1}\frac{\sum_{j\in W_n}j^{-2\beta}(\gamma_j-\gamma_j')^2}{\sum_{j\in W_n}(\gamma_j-\gamma_j')^2}\geq(2L_n)^{-2\beta} \tag{7}$$

If we can show that

$$\min_{h(\gamma,\gamma')=1}||\mathbb{P}_\gamma\wedge\mathbb{P}_{\gamma'}||\geq c_2 \tag{8}$$

then

$$\max_{\gamma\in\mathcal{W}_n}\mathbb{E}\int_{\mathcal{T}}(\hat{b}(t)-b(t))^2dt\geq c{L_n}^{-(2\beta-1)}=cn^{-(2\beta-1)/(\alpha+2\beta)}.$$

We know

$$||\mathbb{P}_\gamma\wedge\mathbb{P}_{\gamma'}||\geq\frac{1}{2}\alpha_2^2(\mathbb{P}_\gamma,\mathbb{P}_{\gamma'})=\left(1-\frac{1}{2}\mathbf{H}^2(\mathbb{Q}_\gamma,\mathbb{Q}_{\gamma'})\right)^{2n}$$

where $\mathbb{Q}_\gamma$ is the joint distribution of one single copy of (Y,X) with parameter γ. Note that

$$\begin{aligned}\mathbf{H}^2(\mathbb{Q}_\gamma,\mathbb{Q}_{\gamma'}) &\leq c_3\int_0^1\int_0^1[b_\gamma(s)-b_{\gamma'}(s)]\,[b_\gamma(t)-b_{\gamma'}(t)]\,K_0(s,t)dsdt\\ &= c_4j^{-\alpha-2\beta}=c_4/n,\end{aligned}$$

then equation (8) follows immediately.

Remark 4 *This approach can be applied to many functional regression models such as generalized functional linear regression and single index model.*