Week 6
Spring 2009

Lectures 11. Sharp Asymptotic Minimaxity for Nonparametric Regression – Lower Bound

Model:

$$y_i = \theta_i + \epsilon z_i,\ z_i \overset{i.i.d.}{\sim} N(0,1),\ \theta \in \Theta_M$$

where Θ_M is an ellipsoid in $l_2(\mathbb{N})$:

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

Pinsker's Theorem: Let $\Theta = \left\{ \theta : \sum a_i^2 \theta_i^2 \le M \right\}$ and $a_i \to \infty$, then

$$R_N(\Theta, \epsilon) \sim R_L(\Theta, \epsilon) \text{ as } \epsilon \to 0.$$

We will only prove this result for the following Sobolev ball,

$$\Theta_M = \left\{ \theta : \sum_i a_i^2 \theta_i^2 \le M,\ a_{2k} = a_{2k+1} = (2\pi k)^m \right\}.$$

Review of Linear Minimaxity. Recall that

$$R_L(\Theta, \epsilon) = \inf_c \sup_\theta \sum \mathbb{P}(c_i y_i - \theta)^2$$

is achieved at

$$c_i = \epsilon^2 (1 - a_i/\lambda_*)_+$$

where λ_* is determined by the following equation

$$\epsilon^2 \sum a_i (\lambda_* - a_i)_+ = M.$$

which gives

$$\lambda_* \sim \left[\frac{(m+1)(2m+1)}{m} \pi M \right]^{\frac{m}{2m+1}} \epsilon^{-\frac{2m}{2m+1}}$$

and suggests the least favorable prior should be

$$\theta_i \sim N\left(0, \tau_i^2\right) \text{ with } \tau^2 = \epsilon^2 (\lambda_*/a_i - 1)_+$$

since $\tau_i^2 / \left(\tau_i^2 + \epsilon^2\right) \cdot y_i = (1 - a_i/\lambda_*)_+ y_i$ which is the optimal linear estimator. Note that for that prior

$$r(G_\epsilon, \delta_{G_\epsilon}) = \sum_{i=1} \tau_i^2 \epsilon^2 / \left(\tau_i^2 + \epsilon^2\right) = \epsilon^2 \sum_{i=1} (1 - a_i/\lambda_*)_+ = R_L(\Theta, \epsilon).$$

Strategy. Find a sequence of priors Q_ϵ supported in Θ such that

$$R_L(\Theta, \epsilon) = (1 + o(1))\, r(Q_\epsilon, \delta_{Q_\epsilon}).$$

Then

$$R_N(\Theta, \epsilon) \sim R_L(\Theta, \epsilon)$$

since

$$r(G_\epsilon, \delta_{G_\epsilon}) \le R_N(\Theta, \epsilon) \le R_L(\Theta, \epsilon).$$

It is natural to define

$$G_\epsilon = \prod_i N\left(0, \epsilon^2 (\lambda_*/a_i - 1)_+\right).$$

But

$$\mathbb{E} \sum a_i^2 \theta_i^2 = \sum \epsilon^2 a_i^2 (\lambda_*/a_i - 1)_+ = \epsilon^2 \sum a_i (\lambda_* - a_i)_+ = M,$$

then G_ϵ is not supported in Θ!

What about $\gamma_\epsilon G_\epsilon$ with $\gamma_\epsilon \nearrow 1$? Can we show $\gamma_\epsilon G_\epsilon$ is concentrated on Θ? Define

$$G_\epsilon^* = \prod_i N\left(0, \gamma_\epsilon^2 \epsilon^2 (\lambda_*/a_i - 1)_+\right)$$

where $\gamma_\epsilon^2 < 1$ and increase to 1 with a certain rate, then $r\left(G_\epsilon^*, \delta_{G_\epsilon^*}\right) \sim r(G_\epsilon, \delta_{G_\epsilon})$. We hope that G_ϵ^* is concentrated on Θ by choosing γ_ϵ appropriately, although not supported in Θ. In other words, G_ϵ^* is "kind of" a prior with support in Θ.

Define $Q_\epsilon = G_\epsilon^*(\cdot|\Theta)$, then

$$r\left(G_\epsilon^*, \delta_{G_\epsilon^*}\right) = \int \mathbb{E} \left\| \delta_{G_\epsilon^*} - \theta \right\|^2 G_\epsilon^*(d\theta) \le \int \mathbb{E} \left\| \delta_{Q_\epsilon} - \theta \right\|^2 G_\epsilon^*(d\theta)$$

and

$$r(Q_\epsilon, \delta_{Q_\epsilon}) \le R_N(\Theta, \epsilon).$$

If we can show G_ϵ^* is concentrated on Θ, then G_ϵ^* must be very close to Q_ϵ, so $r\left(G_\epsilon^*, \delta_{G_\epsilon^*}\right)$ is very close to $r(Q_\epsilon, \delta_{Q_\epsilon})$. We have rigorous arguments as follows,

$$\begin{aligned}
r\left(G_\epsilon^*, \delta_{G_\epsilon^*}\right) &\le \int \mathbb{E} \left\| \delta_{Q_\epsilon} - \theta \right\|^2 G_\epsilon^*(d\theta) \\
&= \int_\Theta \mathbb{E} \left\| \delta_{Q_\epsilon} - \theta \right\|^2 G_\epsilon^*(d\theta) + \int_{\Theta^c} \mathbb{E} \left\| \delta_{Q_\epsilon} - \theta \right\|^2 G_\epsilon^*(d\theta) \\
&\le r(Q_\epsilon, \delta_{Q_\epsilon}) \int_\Theta G_\epsilon^*(d\theta) + CM \int_{\Theta^c} G_\epsilon^*(d\theta)
\end{aligned}$$

where we will show later

$$\begin{aligned}
\int_{\Theta} G_{\epsilon}^{*}(d\theta) &\rightarrow 1 \\
\int_{\Theta^{c}} G_{\epsilon}^{*}(d\theta) &= o\left(R_{L}(\Theta, \epsilon)\right).
\end{aligned}$$

Let $t=\left(\frac{1-\gamma_{\epsilon}^{2}}{\gamma_{\epsilon}^{2}}\right) M$ and pick γ_{ϵ} such that $\left(\frac{1-\gamma_{\epsilon}^{2}}{\gamma_{\epsilon}^{2}}\right) M=\sum\left(a_{i}^{2} \tau_{i}^{2}\right)^{2} / \sup_{i}\left(a_{i}^{2} \tau_{i}^{2}\right)$ (which is less than M). The lemma below implies

$$\begin{aligned}
\mathbb{P}\left(\sum a_{i}^{2} \theta_{i}^{2} \geq M\right) &= \mathbb{P}\left(\sum a_{i}^{2}\left(\theta_{i}^{2}-\gamma_{\epsilon}^{2} \tau_{i}^{2}\right) \geq\left(1-\gamma_{\epsilon}^{2}\right) M\right) \\
&= \mathbb{P}\left(\sum a_{i}^{2} \tau_{i}^{2}\left(\left(\frac{\theta_{i}}{\gamma_{\epsilon} \tau_{i}}\right)^{2}-1\right) \geq\left(\frac{1-\gamma_{\epsilon}^{2}}{\gamma_{\epsilon}^{2}}\right) M\right) \\
&\leq \exp \left(-\frac{t^{2}}{8 \sum\left(a_{i}^{2} \tau_{i}^{2}\right)^{2}}\right)=\exp \left(-\frac{\sum\left(a_{i}^{2} \tau_{i}^{2}\right)^{2}}{8\left[\sup_{i}\left(a_{i}^{2} \tau_{i}^{2}\right)\right]^{2}}\right).
\end{aligned}$$

It is easy to see $\frac{\sum\left(a_{i}^{2} \tau_{i}^{2}\right)^{2}}{8\left[\sup_{i}\left(a_{i}^{2} \tau_{i}^{2}\right)\right]^{2}} \asymp \lambda_{*}^{1/m}$, thus $\mathbb{P}\left(\sum a_{i}^{2} \theta_{i}^{2} \geq M\right)$ decays to 0 exponentially fast.

Lemma. Suppose that $Z_{i} \stackrel{i.i.d.}{\sim} N(0,1)$, then

$$\mathbb{P}\left(\sum \alpha_{i}\left(Z_{i}^{2}-1\right) \geq t\right) \leq \exp \left(-\frac{t^{2}}{8\|\alpha\|^{2}}\right), t \leq \frac{\|\alpha\|^{2}}{\|\alpha\|_{\infty}}.$$

Proof: Let $Y_{i}=Z_{i}^{2}-1$, then for $s \alpha_{i} \leq 1/4$,

$$\mathbb{E} \exp \left(\sum s \alpha_{i} Y_{i}\right)=\prod_{i} \mathbb{E} \exp \left(s \alpha_{i} Y_{i}\right)=\prod_{i} \frac{e^{-s \alpha_{i}}}{\left(1-2 s \alpha_{i}\right)^{1/2}} \leq \prod_{i} \exp \left(2 s^{2} \alpha_{i}^{2}\right)$$

The second equality is due to the moment generating function for χ^{2}, and the inequality follows from the fact $\log(1-v) \geq -v-v^{2}/2$ for $v \leq 1/2$. Then

$$\mathbb{P}\left(\sum \alpha_{i}\left(Z_{i}^{2}-1\right) \geq t\right) \leq \exp \left(2 s^{2}\|\alpha\|^{2}-s t\right), \text{ for } s \leq \frac{1}{4\|\alpha\|_{\infty}}$$

i.e.,

$$\mathbb{P}\left(\sum \alpha_{i}\left(Z_{i}^{2}-1\right) \geq t\right) \leq \exp \left(-\frac{t^{2}}{8\|\alpha\|^{2}}\right), \text{ for } t=4 s\|\alpha\|^{2} \leq \frac{\|\alpha\|^{2}}{\|\alpha\|_{\infty}}.$$

Lecture 12. Sharp Asymptotic Minimaxity for Nonparametric Regression – Adaptive Estimation

Model:

$$y_i = \theta_i + \epsilon z_i,\; z_i \overset{i.i.d.}{\sim} N(0,1),\; \theta \in \Theta_M$$

where Θ_M is an ellipsoid in $l_2(\mathbb{N})$

$$\Theta(m,M) = \left\{ f : \sum_{i=1}^{\infty} a_i^2 \theta_i^2 \le M,\; a_{2k} = a_{2k+1} = (2\pi k)^m \right\}$$

In this lecture, we apply blockwise James-Stein estimator to achieve the sharp linear adaptive minimaxity.

Lemma. Let $X \sim N_d(\mu, \epsilon I)$

$$R\left(\widehat{\mu}^{JS}, \mu\right) \le 2\epsilon^2 + \frac{(d-2)\,\epsilon^2 \|\mu\|^2}{(d-2)\,\epsilon^2 + \|\mu\|^2} \le 2\epsilon^2 + \frac{d\epsilon^2 \|\mu\|^2}{d\epsilon^2 + \|\mu\|^2}$$

Proof: Without loss of generality, we assume that $\epsilon = 1$. Since $\|X\|^2$ can be seen as a mixture of χ^2_{d+2N} and $N \sim$Poisson$\left(\|\mu\|^2/2\right)$, and

$$R\left(\widehat{\mu}^{JS}, \mu\right) = d - (d-2)^2\, \mathbb{E}\, \|X\|^{-2},$$

then

$$R\left(\widehat{\mu}^{JS}, \mu\right) = d - (d-2)^2\, \mathbb{E} \frac{1}{d-2+2N} \underset{\text{Jensen}}{\le} d - (d-2)^2 \frac{1}{d-2+\|\mu\|^2}.$$

Remark. For $\epsilon = 1$, this Lemma implies

$$R\left(\widehat{\mu}^{JS}, \mu\right) \le 2 + R\left(\widehat{\mu}^{IS}, \mu\right).$$

where $R\left(\widehat{\mu}^{IS}, \mu\right) = \inf_c R(\widehat{\mu}_c, \mu)$.

Proof of adaptive minimaxity. Define

$$B_b = \left\{ i : \left[a^{b-1}\right] \le i < \left[a^b\right],\; a = 1 + 1/\log n \right\}.$$

Set L such that $a^{L-1} > 3\log n + 1$, for instance, $L = 4$. Starting from the L-th block, we apply James-Stein estimator to the observations in each block. Then we have

$$R_\epsilon\left(\widehat{\theta}^{BJS}, \theta\right) \le \left(2\log_a n - 2L + a^L\right)\frac{1}{n} + \sum_{b=L}^{\log_a n} R\left(\widehat{\theta}^{IS}_{(b)}, \theta_{(b)}\right) + \sum_{i \ge n} \theta_i^2$$

Let L be finite. It is easy to see

$$\left(2\log_a n - 2L + a^L\right)\frac{1}{n} + \sum_{i\ge n} \theta_l^2 = O\left(n^{-2m}\right) = o\left(n^{-2m/(2m+1)}\right).$$

Note that

$$\sup_{\Theta(m,M)} \sum_{b=L}^{\log_a n} R\left(\widehat{\theta}_{(b)}^{IS}, \theta_{(b)}\right) \leq \sup_{\sum_b (\pi a)^{2(b-1)m} \|\theta\|_{(b)}^2 \leq M} \sum_{b=L}^{\log_a n} R\left(\widehat{\theta}_{(b)}^{IS}, \theta_{(b)}\right)$$

We will show later

$$\sup_{\sum_b (\pi a)^{2(b-1)m} \|\theta\|_{(b)}^2 \leq M} \sum_{b=L}^{\log_a n} R\left(\widehat{\theta}_{(b)}^{IS}, \theta_{(b)}\right) \sim R_L\left(\Theta\left(m, M\right), \epsilon\right).$$

Recall that

$$R_L\left(\Theta, \epsilon\right) = \epsilon^2 \sum_{i=1} \left(1 - a_i/\lambda_*\right)_+.$$

where λ_* is determined by the following equation

$$\epsilon^2 \sum a_i \left(\lambda_* - a_i\right)_+ = M.$$

In our setting, similarly we have

$$R_{BL}\left(\Theta, \epsilon\right) = \epsilon^2 \left|B_b\right| \sum_{b=L}^{B} \left(1 - \frac{\left(\pi a\right)^{(b-1)m}}{\lambda_{*B}}\right)_+$$

where λ_* is uniquely determined by the equation

$$\left|B_b\right| \epsilon^2 \sum_{b=L}^{\log_a n} \left(\pi a\right)^{(b-1)m} \left(\lambda_{*B} - \left(\pi a\right)^{(b-1)m}\right)_+ = M$$

Note that for $i \in B_b$ we have $a_i \sim \left(\pi a\right)^{(b-1)m}$. Equivalently we write

$$\epsilon^2 \sum b_i \left(\lambda_{*B} - b_i\right)_+ = M$$

where $b_i = \left(\pi a\right)^{(b-1)m} = \left(1 + o\left(1\right)\right) a_i$ for $i \in B_b$, and

$$R_{BL}\left(\Theta, \epsilon\right) = \epsilon^2 \sum \left(1 - \frac{b_i}{\lambda_{*B}}\right)_+.$$

Thus we have $R_{BL}\left(\Theta, \epsilon\right) \sim R_L\left(\Theta, \epsilon\right)$.