Stat210B: Theoretical Statistics **Lecture Date: May 08, 2007**

Lower Bounds on Rate of Convergence

Lecturer: Michael I. Jordan *Scribe: Wei-Chun Kao*

1 Kernel Density Estimation

Consider kernel density estimator

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K(\frac{x-X_i}{h}), \tag{1}$$

with performance measure

$$\begin{aligned} \mathrm{MISE}_f(\hat{f}) &= \int E_f(\hat{f}(x) - f(x))^2 dx & (2)\\ &= \int \mathrm{Var}_f \hat{f}(x)dx + \int \mathrm{Bias}_f^2 \hat{f}(x)dx. & (3)\end{aligned}$$

We will show that

$$\begin{aligned} \int \mathrm{Var}_f \hat{f}(x)dx &= \Omega(\frac{1}{nh}) & (4)\\ \int \mathrm{Bias}_f^2 \hat{f}(x)dx &= \Omega(h^4). & (5)\end{aligned}$$

To get optimal MISE rate, we balance these two terms by choosing $h = n^{-1/5}$. And we get the optimal rate $\mathrm{MISE}^* = n^{-4/5}$.

Theorem 1. *(van der Vaart, 1998, Theorem 24.1)*

Assume that

- $X_1, ..., X_n$ *i.i.d. distributed according to density* f,
- f *twice continuously differentiable,*
- $\int |f''(x)|dx < \infty$,
- *kernel* K *s.t.* $\int yK(y)dy = 0$, $\int y^2K(y)dy < \infty$, *and* $\int K^2(y)dy < \infty$.

Then $\exists C_f$ *s.t. MISE* $\leq C_f(\frac{1}{nh} + h^4)$

Proof. We start by bounding the variance term.

$$\begin{aligned} \mathrm{Var}_f \hat{f}(x) &= \frac{1}{n}\mathrm{Var}_f \frac{1}{h}K(\frac{x-X_1}{h}) \quad \text{because } X_1, ..., X_n \text{ are i.i.d.} & (6)\\ &\leq \frac{1}{nh}E_f K^2(\frac{x-X_1}{h}) \quad \text{drop square of mean to get inequality} & (7)\\ &= \frac{1}{nh}\int K^2(y)f(x-hy)dy & (8)\end{aligned}$$

Integrate $\mathrm{Var}_f \hat{f}(x)$, we get

$$\int \frac{1}{nh} \int K^2(y) f(x-hy) dy dx = \frac{1}{nh} \int K^2(y) dy \tag{9}$$

For the bias term, by the following Taylor expansion:

$$f(x+h) - f(x) = hf'(x) + h^2 \int_0^1 f''(x-sh)(1-s) ds, \tag{10}$$

we have

$$\begin{aligned} E_f \hat{f}(x) - f(x) &= \int \frac{1}{h} K(\frac{x-t}{h}) f(t) dt - f(x) &(11)\\ &= \int_0^1 K(y)(f(x-hy) - f(x)) dy &(12)\\ &= \iint_0^1 K(y)[-hyf'(x) + (hy)^2 f''(x-shy)(1-s)] ds dy &(13) \end{aligned}$$

where the first term in the square brackets can be canceled due to the fact that K has mean 0. Use Cauchy-Schwartz on $Y \sim K(y)$ and $Y f''(x - ShY)(1 - S)$ with $S \sim \mathrm{Unif}(0, 1)$, we get

$$\mathrm{Bias}^2(x) \le \frac{1}{3} h^4 \left(\int K(y) y^2 dy \right)^2 \left(\int f''(x)^2 dx \right). \tag{14}$$

$\square$

More generally, assume that f has m-continous derivative, and K satisfies $\int K(y)dy > 1$, $\int yK(y)dy = \cdots = \int y^{m-1} K(y) dy = 0$, $\int y^2 K(y) dy < \infty$, $\int |y|^m K(y) < \infty$ and $\int K^2(y) dy < \infty$, then $\mathrm{Bias}^2 = \Omega(h^{2m})$, $h^* = \Omega(n^{\frac{1}{2m+1}})$, $\mathrm{MISE}^* = \Omega(n^{-\frac{2m}{2m+1}}) \to n^{-1}$ as $m \to \infty$ which is the parametric rate.

2 Rate Optimality

We use Assouad's lemma (c.f. Fano's lemma). We will concentrate our anallysis on a subset of functions $\mathcal{F}_n$ indexed by bit vectors $\theta \in \{0, 1\}^{r_n}$. where $r_n = \lfloor n^{\frac{1}{2m+1}} \rfloor$. $\mathcal{F}_n$ contains 2^{r_n} functions. Set $h_n = n^{-\frac{1}{2m+1}}$, let $X_{n,1}, ..., X_{n,m}$ be a grid of mesh with width $2h_n$. Define

$$f_{n,\theta}(x) = f(x) + h_n^m \sum_{j=1}^{r_n} \theta_j K(\frac{x - X_{n,j}}{h_n}), \tag{15}$$

for a kernel K with support $(0, 1)$. Also define Hamming distance

$$H(\theta, \theta') = \sum_{i=1}^{r_n} |\theta_i - \theta'_i|, \tag{16}$$

and some sort of variation

$$\|P \wedge Q\| = \int (P \wedge Q) d\mu \tag{17}$$

Lemma 2. *Assonad's Lemma (van der Vaart, 1998, Theorem 24.3)*

For an estimator T based on an observation from a model in the set $\{P_\theta : \theta \in \{0,1\}^r\}$ and any p,

$$\max_\theta 2^p E_\theta d^p(T, \psi(\theta)) \geq \min_{H(\theta,\theta')\geq 1} \frac{d^p(\psi(\theta), \psi(\theta'))}{H(\theta,\theta')} \frac{r}{2} \min_{H(\theta,\theta')=1} \|P_\theta \wedge P_{\theta'}\|. \tag{18}$$

We want to apply Assouad's lemma to the product measures resulting from the densities $f_{n,\theta}$.

First, define affinity

$$A(P,Q) = \int \sqrt{pq}\, d\mu, \tag{19}$$

we need the following lemma.

Lemma 3.

$$\|P^n \wedge Q^n\| \geq \frac{1}{2} A^2(P^n, Q^n) = \frac{1}{2}\left(1 - \frac{1}{2}H^2(P,Q)\right)^{2n} \tag{20}$$

Proof. First, note that

$$\begin{aligned} H^2(P,Q) &= \int (\sqrt{p} - \sqrt{q})^2 d\mu = 2 - 2A(P,Q), && (21)\\ pq &= (p \vee q)(p \wedge q). && (22) \end{aligned}$$

By definition of affinity,

$$\begin{aligned} A^2(P,Q) &= \left(\int \sqrt{pq}\, d\mu\right)^2 && (23)\\ &= \left(\int (p \vee q)^{1/2} (p \wedge q)^{1/2} d\mu\right)^2 && (24)\\ &\leq \left(\int (p + q)^{1/2} (p \wedge q)^{1/2} d\mu\right)^2 && (25)\\ &\leq \left(\int (p+q) d\mu\right)\left(\int (p \wedge q) d\mu\right) \text{ by Cauchy-Schwartz} && (26)\\ &= 2\int p \wedge q\, d\mu = 2\|p \wedge q\| && (27) \end{aligned}$$

By Fubini's theorem, we have $A(P^n, Q^n) = A(P,Q)^n$. Therefore,

$$\begin{aligned} \|P^n \wedge Q^n\| &\geq \frac{1}{2}A^2(P^n, Q^n) = \frac{1}{2}A(P,Q)^{2n} && (28)\\ &= \frac{1}{2}\left(1 - \frac{1}{2}H^2(P,Q)\right)^{2n}. && (29) \end{aligned}$$

□

Theorem 4. *(van der Vaart, 1998, Theorem 24.4)*

There exists a constant D such that for any density estimator $\hat{f}_n$

$$\sup_{f \in \mathcal{F}_m} E_f \int \left(\hat{f}_n(x) - f(x)\right)^2 dx \geq D\left(\frac{1}{n}\right)^{2m/(2m+1)} \tag{30}$$

Apply to $\{f_{n,\theta}\}$, we get

$$\begin{aligned}
\int (f_{n,\theta}^{1/2} - f_{n,\theta'}^{1/2})^2 dx &= \int \left(\frac{f_{n,\theta} - f_{n,\theta'}}{f_{n,\theta}^{1/2} + f_{n,\theta'}^{1/2}} \right)^2 dx && (31) \\
&\geq C \int (f_{n,\theta} - f_{n,\theta'})^2 dx && (32) \\
&= C h_n^{2m} \sum_{j=1}^{r_n} |\theta_j - \theta_j'|^2 \int K^2 (\frac{x - X_{n,j}}{h_n}) dx && (33) \\
&= C h_n^{2m+1} H(\theta, \theta') \int K^2(x) dx && (34)
\end{aligned}$$

For $H(\theta, \theta') = 1$, (34) is just a constant times $h_n^{2m+1} = n^{-\frac{2m+1}{2m+1}} = n^{-1}$. Thus,

$$\|P^n - Q^n\| \geq \frac{1}{2}(1 - \frac{1}{2} H^2(P, Q))^{2n} \geq \frac{1}{2}(1 - O(n^{-1}))^{2n}, \tag{35}$$

which is bounded. Plug (34), (35) into Assouad's lemma, we get

$$\begin{aligned}
\max_\theta 2^2 E_\theta \int (\hat{f}_n(x) - f_{n,\theta}(x))^2 dx &\geq h_n^{2m+1} \frac{r_n}{2} (1 - O(n^{-1}))^{2n} \text{ up to constants} && (36) \\
&\geq D n^{-\frac{2m}{2m+1}}. && (37)
\end{aligned}$$

References

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.