

Lower Bounds on Rate of Convergence

Lecturer: Michael I. Jordan

Scribe: Wei-Chun Kao

1 Kernel Density Estimation

Consider kernel density estimator

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

with performance measure

$$\text{MISE}_f(\hat{f}) = \int E_f(\hat{f}(x) - f(x))^2 dx \quad (2)$$

$$= \int \text{Var}_f \hat{f}(x) dx + \int \text{Bias}_f^2 \hat{f}(x) dx. \quad (3)$$

We will show that

$$\int \text{Var}_f \hat{f}(x) dx = \Omega\left(\frac{1}{nh}\right) \quad (4)$$

$$\int \text{Bias}_f^2 \hat{f}(x) dx = \Omega(h^4). \quad (5)$$

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

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

Assume that

- X_1, \dots, 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. $\text{MISE} \leq C_f\left(\frac{1}{nh} + h^4\right)$

Proof. We start by bounding the variance term.

$$\text{Var}_f \hat{f}(x) = \frac{1}{n} \text{Var}_f \frac{1}{h} K\left(\frac{x - X_1}{h}\right) \quad \text{because } X_1, \dots, X_n \text{ are i.i.d.} \quad (6)$$

$$\leq \frac{1}{nh} E_f K^2\left(\frac{x - X_1}{h}\right) \quad \text{drop square of mean to get inequality} \quad (7)$$

$$= \frac{1}{nh} \int K^2(y) f(x - hy) dy \quad (8)$$

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

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

For the bias term, by the following Taylor expansion:

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

we have

$$E_f \hat{f}(x) - f(x) = \int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t) dt - f(x) \quad (11)$$

$$= \int_0^1 K(y) (f(x - hy) - f(x)) dy \quad (12)$$

$$= \iint_0^1 K(y) [-hyf'(x) + (hy)^2 f''(x - shy)(1 - s)] ds dy \quad (13)$$

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 $S \sim \text{Unif}(0, 1)$, we get

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

□

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

2 Rate Optimality

We use Assouad's lemma (c.f. Fano's lemma). We will concentrate our analysis 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 $X_{n,1}, \dots, X_{n,m}$ be a grid of mesh with width $2h_n$. Define

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

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

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

and some sort of variation

$$\|P \wedge Q\| = \int (P \wedge Q) d\mu \quad (17)$$

Lemma 2. *Assouad'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 ,

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

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

First, define affinity

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

we need the following lemma.

Lemma 3.

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

Proof. First, note that

$$H^2(P, Q) = \int (\sqrt{p} - \sqrt{q})^2 d\mu = 2 - 2A(P, Q), \quad (21)$$

$$pq = (p \vee q)(p \wedge q). \quad (22)$$

By definition of affinity,

$$A^2(P, Q) = \left(\int \sqrt{pq} d\mu\right)^2 \quad (23)$$

$$= \left(\int (p \vee q)^{1/2} (p \wedge q)^{1/2} d\mu\right)^2 \quad (24)$$

$$\leq \left(\int (p + q)^{1/2} (p \wedge q)^{1/2} d\mu\right)^2 \quad (25)$$

$$\leq \left(\int (p + q) d\mu\right) \left(\int (p \wedge q) d\mu\right) \text{ by Cauchy-Schwartz} \quad (26)$$

$$= 2 \int p \wedge q d\mu = 2\|p \wedge q\| \quad (27)$$

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

$$\|P^n \wedge Q^n\| \geq \frac{1}{2} A^2(P^n, Q^n) = \frac{1}{2} A(P, Q)^{2n} \quad (28)$$

$$= \frac{1}{2} \left(1 - \frac{1}{2} H^2(P, Q)\right)^{2n}. \quad (29)$$

□

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

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

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