Supplementary Material

0.1 Technical Lemmas

Lemma 1. Assume B1 and B2 hold. Define \( H_\sqrt{f/f^*} \in \mathcal{M}_q, \mathcal{H}(h, h^*) \leq \epsilon \). Then

\[
N \left( H_\sqrt{f/f^*} \right) \leq \left( C_1 \epsilon \right)^{10} \left( C_2 \delta^{\frac{1}{q+1}} + 1 \right)
\]

for all \( q \geq q^* \) and \( \delta/\epsilon \leq 1 \).

Proof. We prove the lemma by extending the arguments in [1]. Note that \( f/f^* = h/h^* \).

We denote by \( \| \cdot \|_p \) the \( L^p(h^*d\mu) \)-norm, i.e., \( \| g \|_p^0 = \int |g|^p h^*d\mu \). Denote \( \eta_i^* = (x, \beta_i^*) \) and \( \eta_j = (x, \beta_j^*), i = 1, \ldots, q^*, j = 1, \ldots, q \). It follows from B2 that we can find a partition of \( \mathbb{R} \), \( A_0, A_1, \ldots, A_{q^*} \), such that each bounded set \( A_i, i = 1, \ldots, q^* \), contains precisely one component \( \eta_i^* = (X, \beta_i^*) \) and the unbounded set \( A_0 = \mathbb{R}^M \setminus (A_1 \cup \cdots \cup A_{q^*}) \) contains no component. Let \( f \in \mathcal{M}_q \), so that we can write \( f = \sum_{i=1}^q \pi_i f_0(y - \eta_i) \). Then,

\[
\frac{f - f^*}{f^*} = \sum_{j: \eta_j \in A_0} \pi_j \frac{f_0(y - \eta_j)}{f^*} + \sum_{i=1}^{q^*} \left( \sum_{j: \eta_j \in A_i} \pi_j - \pi_i^* \right) \frac{f_0(y - \eta_i^* - \eta_j)}{f^*} + \sum_{j: \eta_j \in A_i} \frac{f_0(y - \eta_j) - f_0(y - \eta_i^*)}{f^*}.
\]

Taylor expansion gives

\[
f_0(y - \eta_j) - f_0(y - \eta_i^*) = \tilde{f}_0(y - \eta_i^*)(\eta_j - \eta_i^*) + \frac{1}{2} \tilde{f}_0(y - \eta)(\eta_j - \eta_i^*)^2.
\]
Using Assumption B2, we find that
\[
\| f - f^* \|_1 \geq c^* \left[ \sum_{j : \eta_j \in A_0} \pi_j + \sum_{i=1}^{q^*} \left( \left| \sum_{j : \eta_j \in A_i} \pi_j - \pi_i^* \right| + \left| \sum_{j : \eta_j \in A_i} \pi_j(\eta_j - \eta_i^*) \right| \right) + \frac{1}{2} \sum_{j : \eta_j \in A_i} \pi_j(\eta_j - \eta_i^*)^2 \right].
\]

On the other hand, it follows from Theorem 3.10 of [1] that there exists a constant \(c^*\) such that
\[
\| f - f^* \|_1 \leq \frac{1}{c^*}(H_0 + H_1 + H_2).
\]

In addition, using \(|\sqrt{x} - 1| \leq |x - 1|\), we find
\[
\frac{|\sqrt{h/h^*} - 1|}{H(h, h^*)} \leq \frac{|h/h^* - 1|}{\frac{1}{2}||h/h^* - 1||_1} = \frac{|f/f^* - 1|}{\frac{1}{2}||f/f^* - 1||_1} \leq 2S.
\]

Similar to Lemma 3.15 of [1], for any \(f \in \mathcal{M}\), we have
\[
\frac{|\sqrt{h/h^*} - 1|}{H(h, h^*)} = \frac{h/h^* - 1}{\sqrt{\chi^2(h/h^*)}} \leq (4S^2S + 2S^3)H(h, h^*),
\]

where the chi-square divergence is defined as \(\chi^2(h/h^*) = \int (h/h^* - 1)^2 h^* d\mu\). This allows us to make further approximation based on Lemma 3.16 of [1]. Let \(\alpha > 0\), and for every \(f \in \mathcal{M}_q\) such that \(H(h, h^*) \leq \alpha\). Define
\[
\tilde{\ell} = \sum_{i=1}^{q^*} \left\{ a_i \frac{f_0(y - \eta_i^*)}{f^*} + b_i \frac{f_0(y - \eta_i^*)}{f^*} + e_i \frac{f_0(y - \eta_i^*)}{f^*} \right\} + \sum_{j=1}^{q} \gamma_j \frac{f_0(y - \eta_j)}{f^*},
\]

where
\[
\sum_{i=1}^{q^*} |a_i| \leq \frac{1}{c^*} + \frac{1}{\sqrt{c^*} \alpha}, \quad \sum_{i=1}^{q^*} |b_i| \leq \frac{1}{c^*} + \frac{2}{\sqrt{c^*} \alpha}, \quad \sum_{i=1}^{q^*} e_i^2 \leq \frac{1}{c^*}, \quad \sum_{j=1}^{q} \gamma_j \leq \frac{1}{\sqrt{c^*} \alpha} + c^*.
\]

(2)
0.1. TECHNICAL LEMMAS

We have

\[ \frac{|h/h^* - 1|}{\sqrt{\chi^2(h^2)}} - \bar{t} \leq \frac{\sqrt{2}}{2(c^*)^{5/4}}(\|H_3\|_2 S + H_3)\alpha^{1/4}. \]

Define

\[ d_f = \frac{\sqrt{f/f^*} - 1}{\|\sqrt{f/f^*} - 1\|_2}, \]

and

\[ D_q = \{ d_f : f \in M_q, f \neq f^* \}, \quad D_{q,\alpha} = \{ d_f : f \in M_q, f \neq f^*, \mathcal{H}(h, h^*) \leq \alpha \}. \]

Then,

\[ N_{|\mathcal{H}|}(D_q, \delta) \leq N_{|\mathcal{H}|}(D_{q,\alpha}, \delta) + N_{|\mathcal{H}|}(D_q \setminus D_{q,\alpha}, \delta). \]

We estimate both bracketing numbers separately. Define a family of functions

\[ \tilde{L}_{q,\alpha} = \left\{ \sum_{i=1}^{q} \left\{ a_i \frac{f_0(y - \eta_i)}{f^*} + b_i \frac{f_0(y - \eta_i)}{f^*} + c_i \right\} \right\}, \]

where \((a, b, c, \gamma) \in \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^q\) satisfies the constraint \( \| \mathbf{e} \| = 1 \). For \( \bar{t}, \bar{\bar{t}} \in \tilde{L}_{q,\alpha}, \)

\[ |\bar{t} - \bar{\bar{t}}| \leq H_0 \sum_{i=1}^{q} |a_i - a'_i| + H_1 \sum_{i=1}^{q} |b_i - b'_i| + H_0 \sum_{i=1}^{q} |\gamma_i - \gamma'_i| \]

\[ + \frac{1}{\sqrt{c^* \wedge c^*}} H_1 \sum_{i=1}^{q} |(x, \beta_i - \beta'_i)| + \frac{2H_2}{\sqrt{c^*}} \sum_{i=1}^{q} |e_i - e'_i|. \]

Note that

\[ \mathbb{E}_X \left( \sum_{i=1}^{q} |(x, \beta_i - \beta'_i)| \right)^2 \leq q \sum_{i=1}^{q} \mathbb{E}_X |(x, \beta_i - \beta'_i)|^2 = q \sum_{i=1}^{q} \sum_{k=1}^{\infty} \rho_k (g_{ik} - g'_{ik})^2 \]

\[ \leq q \sum_{i=1}^{q} \sum_{k=1}^{M} \rho_k (g_{ik} - g'_{ik})^2 + M^{-2} \sigma_M(1). \]

We may choose \( M = c_1 \delta^{-1/r} \), so that

\[ |\bar{t} - \bar{\bar{t}}| \leq V \|(a, b, \gamma, g, e) - (a', b', \gamma', g', e')\|_{\text{norm}} + o(\delta), \]

where \( V = 3(H_0 + H_1 + H_2) \) and \( \| \cdot \|_{\text{norm}} \) is the norm on \( \mathbb{R}^{3q^* + c_2 q(\delta^{-1/r} + 1)} \) defined by

\[ \|(a, b, \gamma, g, e)\|_{\text{norm}} = \sum_{i=1}^{q} |a_i| + \sum_{i=1}^{q} |b_i| + \sum_{i=1}^{q} |\gamma_i| + \sum_{i=1}^{q} \rho_k g_{ik} + \frac{2}{\sqrt{c^*}} \sum_{i=1}^{q} |e_i|. \]
Using the standard fact of the covering number for the Euclidean ball we obtain

\[ N(\tilde{L}_{q,\alpha}, \delta) = \left( \frac{c_3 + \delta}{\delta} \right)^{3q^* + c_2 q (\delta^{-1/r} + 1)}. \]

Since \( q \geq q^* \) and \( \delta \leq 1 \), we therefore obtain

\[ N(D_{q,\alpha}, \delta) \leq \left( \frac{c_4 \delta^{-1/r}}{\delta} \right)^{c_5 q (\delta^{-1/r} + 1)}. \]

**Lemma 2.** Assume B1 and B2 hold.

(a).

\[
P \left[ \sup_{f \in \mathcal{M}_q} \frac{1}{n} \sum_{i=1}^{n} \log \frac{f(Y_i|X_i)}{f^*(Y_i|X_i)} \geq \alpha \right] \leq C_3 e^{-\alpha/C_3}
\]

for all \( \alpha \geq C_4 q^{2r/(2r+1) n^{1/(2r+1)}} \).

(b).

\[
\lim_{n \to \infty} \sup_{f \in \mathcal{M}_q} \frac{1}{n} \sum_{i=1}^{n} \log \frac{f(Y_i|X_i)}{f^*(Y_i|X_i)} < 0, \ a.s.
\]

**Proof.** Let \( h(y, x) = f(y|x)f_X(x) \) and \( h^*(y, x) = f^*(y|x)f_X(x) \) for \( f \in \mathcal{M}_q \) and \( f^* \in \mathcal{M}_{q^*} \).

Denote by \( \tilde{h} = (h + h^*)/2 \). We first have the inequalities \( \mathcal{K}(h, h^*) \geq \mathcal{H}^2(h, h^*) \), and

\[
\sum_{i=1}^{n} \log \frac{\tilde{h}}{h^*} \leq 2 \sqrt{n} \nu_n(\log \frac{\tilde{h}}{h^*}) - 2n \mathcal{K}(\tilde{h}, h^*). \tag{3}
\]

Therefore,

\[
P \left[ \sup_{\tilde{h}} \sum_{i=1}^{n} \log \frac{\tilde{h}}{h^*} \geq \alpha \right] \leq P \left[ \sup_{\tilde{h}} \sqrt{n} \nu_n(\log \frac{\tilde{h}}{h^*}) - n \mathcal{H}^2(\tilde{h}, h^*) \geq \alpha \right]
\]

\[
\leq \sum_{s=0}^{S} P \left[ \sup_{\tilde{h} \in \mathcal{B}_s} \nu_n(\log \frac{\tilde{h}}{h^*}) \geq \frac{\alpha 2^{s-1}}{\sqrt{n}} \right]
\]

\[
= \sum_{s=0}^{S} P \left[ \sup_{\tilde{h} \in \mathcal{B}_s} \nu_n(\sqrt{\frac{\tilde{h}}{h^*}}) \geq \frac{\alpha 2^{s-2}}{\sqrt{n}} \right].
\]
where $\mathcal{G}_0 = \{ \tilde{h} : nH^s(\tilde{h}, h^*) \leq \alpha \}$, $\mathcal{G}_s = \{ \tilde{h} : \alpha 2^{s-1} < nH^2(\tilde{h}, h^*) \leq \alpha 2^s \}$, $1 \leq s \leq S$, $S = \min \{ s : \alpha 2^s > 2n \}$. We need to find the bracketing number for $H_q(\epsilon) = \{ \sqrt{\tilde{h}/h^*} : H(\tilde{h}, h^*) \leq \epsilon \}$.

This can be easily deduced from Lemma 1 such that

$$N_1[H_q(\epsilon), \delta] \leq \left( \frac{2\sqrt{2C_1}}{\delta} \right) 10(C_2 \delta^{-\frac{1}{2}+1})^{q+1}. \tag{4}$$

Note that $\int_0^\epsilon \log N_1[H_q(\epsilon), \delta]d\delta \leq c\sqrt{q} \epsilon^{1-1/(2r)}$. It requires that $c\sqrt{q} \epsilon^{1-1/(2r)} \leq \sqrt{n} \epsilon^2$ such as $\epsilon \geq c_1(q/n)^{r/(2r+1)}$. Next, apply Theorem 7.4 of [2], together with $\sqrt{\alpha} 2^{s/2+2}/\sqrt{n} \geq c_1(q/n)^{r/(2r+1)}$, i.e., $\alpha \geq c_2 q^{2r/(2r+1)} n^{1/(2r+1)}$. We obtain

$$\sum_{s=0}^S \mathbb{P}\left[ \sup_{h \in \mathcal{G}_s} \nu_n(\log \sqrt{\frac{h}{h^*}}) \geq \frac{\alpha 2^{s-2}}{\sqrt{n}} \right] \leq \sum_{s=0}^S C_3 e^{-\frac{c_2 q^{2r}}{C_3 2^s}} \leq C_4 e^{-\alpha/C_3}.$$

This finishes the proof of Part (a).

To prove Part (b), it follows from (3) that it is enough to show that

$$\lim_{n \to \infty} \sup_{f \in M_q} n^{-1/2} \nu_n(\log \sqrt{\frac{h}{h^*}}) = 0, \text{ a.s.}$$

As in the proof of Part (a), we have

$$\mathbb{P}\left[ \sup_{f \in M_q} n^{-1/2} \nu_n(\log \sqrt{\frac{h}{h^*}}) \geq \alpha \right] \leq C_4 e^{-\alpha n^2/c_4}$$

for every $\alpha > 0$ such that $c_2 q^{2r/(2r+1)} n^{1/(2r+1)} \leq \alpha \sqrt{n} \leq 32 \sqrt{n}$. Hence,

$$\sum_{n=1}^{\infty} \mathbb{P}\left[ \sup_{f \in M_q} n^{-1/2} \nu_n(\log \sqrt{\frac{h}{h^*}}) \geq \alpha \right] \leq \infty$$

for $0 < \alpha \leq 32$. Part b follows from Borel-Cantelli.

\[ \square \]

**Lemma 3.** Assume B1 and B2 hold. Define

$$\Delta_n(q,q^*) = \sup_{f \in M_q} \left( t_n(f) - \lambda_{q,n} \| \beta \|^2_K \right) \sup_{f \in M_{q^*}} \left( t_n(f) - \lambda_{q^*,n} \| \beta \|^2_K \right),$$

where $\lambda_{q,n} = C_5 q^{2r/(2r+1)} n^{1/(2r+1)}$ and $\lambda_{q^*,n} = \tilde{C}_5 (q^*)^{2r/(2r+1)} n^{1/(2r+1)}$. Then,

$$\lim_{n \to \infty} \sup_{q > q^*} \frac{\Delta_n(q,q^*)}{q^{2r/(2r+1)} n^{1/(2r+1)}} \leq C_6, \text{ a.s.} \tag{5}$$
Proof. Using the fact that

$$\sup_{f \in \mathcal{M}_{q^*}} \left( \ell_n(f) - \lambda_{q^*,n} \| \beta^* \|_K^2 \right) \geq \ell_n(f^*) - \lambda_{q^*,n} \| \beta^* \|_K^2$$

yield

$$\Delta_n(q, q^*) \leq \sup_{f \in \mathcal{M}_q} \sum_{i=1}^{n} \log \frac{f(Y_i|X_i)}{f^*(Y_i|X_i)} + \lambda_{q^*,n} \| \beta^* \|_K^2.$$ 

Furthermore, from Part (a) of Lemma 2, for $\alpha \geq C_4$,

$$\mathbb{P}\left[ \sup_{q \geq q^*} \frac{1}{q^{2qd+1}n^{qd+1}} \sup_{f \in \mathcal{M}_q} \sum_{i=1}^{n} \log \frac{f(Y_i|X_i)}{f^*(Y_i|X_i)} \geq \alpha \right] \leq \sum_{q=q^*}^{\infty} C_3 e^{-\alpha q^{2qd+1}n^{qd+1}/C_3} \leq \frac{c_1}{n^2}.$$ 

The lemma follows easily using the Borel-Cantelli. 

0.2 Proofs of Theorems

0.2.1 Proof of Theorem 1

The constants $c_i$ used in the proof are all generic positive constants. First realize that any lower bound for a specific case yields immediately a lower bound for the general case. In the following, consider a special case where $q^*$ and $(\pi_{1}^*, \ldots, \pi_{q^*}^*)$ are known. $f_0$ is a fixed known density function.

Direct calculation of the Kullback-Leibler divergence between $h^* = f^*(y|x)f_X(x)$ and $h = f(y|x)f_X(x)$ where $f^*(y|x), f(y|x) \in \mathcal{M}_{q^*}$ yields

$$K(h, h^*) = \mathbb{E}_{h^*} \left\{ \log \frac{h^*}{h} \right\} = \mathbb{E}_{h^*} \left\{ \log \frac{\sum_{k=1}^{q^*} \pi_k^* h_k^*}{\sum_{k=1}^{q^*} \pi_k^* h_k} \right\} = \mathbb{E}_{h^*} \left\{ \log \sum_{k=1}^{q^*} \tau_k \frac{h_k^*}{h_k} \right\} \geq \mathbb{E}_{h^*} \left\{ \sum_{k=1}^{q^*} \tau_k \log \frac{h_k^*}{h_k} \right\} = \sum_{k=1}^{q^*} \pi_k^* \mathbb{E}_{h_k^*} \left\{ \frac{h_k}{h} \log \frac{h_k^*}{h_k} \right\},$$

(6)

where $\tau_k = \pi_k^* h_k^*/h$ for $k = 1, \ldots, q^*$ and the inequality is due to the Jensen’s Inequality. Note that $\pi_k^* h_k/h = \pi_k^* f_0(y - \eta_k)/\sum_{i=1}^{q^*} \pi_i^* f_0(y - \eta_i)$ is the probability of a data point coming from
the $k$-th component. We assume this probability bound below away from zero. Therefore, $h_k/h$ is bounded below away from zero and above by a positive constant. Hence,

$$K(h, h^*) = \sum_{k=1}^{q^*} \pi_k E_{h_k^*} \left\{ \frac{h_k/h}{h_k^*/h^*} \log \frac{h_k^*}{h_k} \right\}$$

$$= \sum_{k=1}^{q^*} \pi_k \int_{h_k^* > h_k} \frac{h_k/h}{h_k^*/h^*} \log \frac{h_k^*}{h_k} d\mu + \sum_{k=1}^{q^*} \pi_k \int_{h_k^* \leq h_k} \frac{h_k/h}{h_k^*/h^*} \log \frac{h_k^*}{h_k} d\mu, \geq c_3 \sum_{k=1}^{q^*} \pi_k \int_{h_k^* > h_k} h_k^* \log \frac{h_k^*}{h_k} d\mu + c_4 \sum_{k=1}^{q^*} \pi_k \int_{h_k^* \leq h_k} h_k^* \log \frac{h_k^*}{h_k} d\mu \geq c_5 \sum_{k=1}^{q^*} \pi_k K(h_k, h_k^*).$$

Using the fact $\log y - \log x = \frac{1}{2} (y - x) + \frac{1}{2} (y - x)^2$ for $0 \leq y \leq \sqrt{e}$,

$$K(h_k, h_k^*) = E_{h_k^*} \left\{ \log \frac{h_k^*}{h_k} \right\} \geq \frac{1}{2c_6} E_{h_k^*} \left\{ f_0 (Y - \eta(X, \beta_k)) - f_0 (Y - \eta(X, \beta_k^*)) \right\}^2$$

$$= \frac{1}{2c_6} E_{h_k^*} \left\{ \tilde{f}_0^2 (Y - \tilde{\eta}_k)(\eta(X, \beta_k - \beta_k^*))^2 \right\} \geq c_7 E \left\{ (\eta(X, \beta_k - \beta_k^*))^2 \right\} = c_7 \| \beta_k - \beta_k^* \|^2_C,$$

where $\tilde{\eta}_k$ is a point between $\eta(X, \beta_k^*)$ and $\eta(X, \beta_k)$. Therefore, $K(h, h^*)$ is bounded below by $\| \beta - \beta^* \|^2_C$ up to a constant, where $\| \beta - \beta^* \|^2_C = \sum_{k=1}^{q^*} \| \beta_k - \beta_k^* \|^2_C$. A similar calculation also yields that $K(h, h^*)$ is bounded above by $\| \beta - \beta^* \|^2_C$ up to a constant.

In the following, we adopt the results from [3] to establish the lower bound which is based upon testing multiple hypotheses. In particular, we can find a subset $\{ \beta^{(0)}, \ldots, \beta^{(N)} \} \subset F^*$ with $N$ increasing with $n$, such that for some positive constant $c$ and all $0 \leq i < j \leq N$,

$$\| \beta^{(i)} - \beta^{(j)} \|^2_C \geq 2c_\gamma \frac{2r}{n/q^*} (n/q^*)^{-\frac{2r}{n/q^*}}.$$

(7)
\[
\frac{1}{N} \sum_{j=1}^{N} K(h^{(j)}, h^{(0)}) \leq \gamma \log N, \tag{8}
\]
then we can conclude according to Theorem 2.5 of [3] that,

\[
\inf_{\beta} \sup_{\beta^* \in \mathcal{F}^{q*}} P\left( \|\beta - \beta^*\|_C \geq c (n/q^*)^{-\frac{2r}{2r+1}} \right) \geq \frac{\sqrt{N}}{1 + \sqrt{N}} (1 - 2\gamma - \sqrt{2\gamma \log N}),
\]
which yields

\[
\lim_{a \to 0} \lim_{n \to \infty} \inf_{\hat{\beta}} \sup_{\beta^* \in \mathcal{F}^{q*}} P\left( \|\hat{\beta} - \beta^*\|_C \geq a (n/q^*)^{-\frac{2r}{2r+1}} \right) \geq 1.
\]

Hence Theorem 1 will be proved.

Next, we construct the subset \{\beta^{(0)}, ..., \beta^{(N)}\} \subset \mathcal{F}^{q*}, k = 1, ..., q^*. Let \beta^{(j)} = (\beta^{(j)}_1, \beta^{(j)}_2, ..., \beta^{(j)}_{q^*}), j = 1, ..., N. We show that both (7) and (8) are satisfied. Let \widetilde{M} = \lfloor M/q^* \rfloor for some large number \(M\) to be decided later. Consider the function space

\[
\mathcal{H}^* = \left\{ \beta = \sum_{k=M+1}^{2M} b_k M^{-1/2} L_{K^{1/2}} \varphi_k : (b_{M+1}, ..., b_{2M}) \in \{0, 1\}^M \right\}, \tag{9}
\]
where \{\varphi_k : k \geq 1\} are the orthonomal eigenfunctions of \(K^{1/2} C K^{1/2}\). For any \(\beta \in \mathcal{H}^*\), observe that

\[
\|\beta\|_K^2 = \left\| \sum_{k=M+1}^{2M} b_k M^{-1/2} L_{K^{1/2}} \varphi_k \right\|_K^2 \\
= \sum_{k=M+1}^{2M} b_k^2 M^{-1} \left\| L_{K^{1/2}} \varphi_k \right\|_K^2 \\
\leq \sum_{k=M+1}^{2M} M^{-1} \left\| L_{K^{1/2}} \varphi_k \right\|_K^2 \leq 1,
\]
which shows that \(\mathcal{H}^* \subset \mathcal{F}\). The Varshamov-Gilbert bound shows that for any \(M \geq 8\), there exists a set \(\mathcal{B} = \{b^{(0)}, b^{(1)}, ..., b^{(N)}\} \subset \{0, 1\}^M\) such that

1. \(b^{(0)} = (0, ..., 0)^t\);
2. \( H(b, b') > M/8 \) for any \( b \neq b' \in B \), where \( H(\cdot, \cdot) = \frac{1}{4} \sum_{i=1}^{M} (b_i - b'_i)^2 \) is the Hamming distance;

3. \( N \geq 2^{M/8} \).

The subset \( \{ \beta^{(0)}, \ldots, \beta^{(N)} \} \subset \mathcal{F}^{q^*} \) is chosen as \( \beta^{(i)} = \sum_{k=M+1}^{2M} b^{(i)}_{j,k-M} M^{-1/2} L_{K^{1/2}} \varphi_k, \ i = 0, \ldots, N, \ j = 1, \ldots, q^* \). For any \( 0 \leq i < j \leq N \), observe that

\[
\| \beta^{(i)} - \beta^{(j)} \|_C^2 = \sum_{l=1}^{q^*} \mathbb{E}_X \left( \left( \beta^{(i)}_l - \beta^{(j)}_l, X \right) \right)^2 = \sum_{l=1}^{q^*} \sum_{k=M+1}^{2M} (b^{(i)}_{l,k-M} - b^{(j)}_{l,k-M})^2 M^{-1} \rho_k.
\]

Therefore,

\[
\| \beta^{(i)} - \beta^{(j)} \|_C^2 \geq s_{2M} M^{-1} \sum_{l=1}^{q^*} \sum_{k=1}^{2M} (b^{(i)}_{l,k} - b^{(j)}_{l,k})^2 \geq \rho_{2M}/2 \simeq (M/q^*)^{-2r},
\]

and

\[
\| \beta^{(i)} - \beta^{(j)} \|_C^2 \leq s_{2M} M^{-1} \sum_{l=1}^{q^*} \sum_{k=1}^{M} (b^{(i)}_{l,k} - b^{(j)}_{l,k})^2 \leq \rho_{M} \simeq (M/q^*)^{-2r}.
\]

By taking \( M \) to be the smallest integer greater than \( c_2 \gamma^{-\frac{1}{2r+1}} \left( q^* \right)^{2r/(2r+1)} n^{\frac{1}{r+1}} \) with \( c_2 = (c_1 \cdot 8 \log 2)^{1/(1+2r)} \), the theorem is proved.

### 0.2.2 Proof of Theorem 2

For any \( q \geq q^* \), since \( \hat{f} \) is the maximum, we have

\[
-\ell(\hat{f}) + \lambda \| \hat{\beta} \|_K^2 \leq -\ell(f^*) + \lambda \| \beta^* \|_K^2,
\]

which gives

\[
-(\ell(\hat{f}) - \ell(f^*)) + \lambda \| \hat{\beta} \|_K^2 \leq \lambda \| \beta^* \|_K^2.
\]

Define

\[
d_h = \sqrt{h - 1} / H(h, h^*).
\]
Using \(\log(1 + x) \leq x\),
\[
\ell(f) - \ell(f^*) = \sum_{i=1}^{n} 2\log(1 + \mathcal{H}(h, h^*)d_h(Y, X_i)) \\
\leq \sum_{i=1}^{n} 2\mathcal{H}(h, h^*)d_h(Y, X_i) \\
= 2\sqrt{n} \nu_n(d_h)\mathcal{H}(h, h^*) - n \mathcal{H}^2(h, h^*),
\]
where \(\nu_n(g) = n^{-1/2}\sum_{i=1}^{n} (g(Y_i, X_i) - \mathbb{E}g(Y_i, X_i))\). So,
\[
-(\ell(f) - \ell(f^*)) \geq n \mathcal{H}^2(h, h^*) - 2\sqrt{n} \nu_n(d_h)\mathcal{H}(h, h^*).
\]
Combining this with (10),
\[
\lambda \|\hat{\beta}\|_K^2 + \lambda \|\beta^*\|_K^2 \leq \lambda \|\beta^*\|_K^2 + 2\sqrt{n} \nu_n(d_h)\mathcal{H}(\hat{h}, h^*) \\
= \lambda \|\beta^*\|_K^2 + 2\sqrt{n} (\nu_n(g_h) - \nu_n(g_{h^*})), 
\]
where \(g_h = \sqrt{h/h^*} = \sqrt{f/f^*}\). It is critical to investigate the behavior of \(|\nu_n(g_h) - \nu_n(g_{h^*})|\) as a function of \(\mathcal{H}(h, h^*)\).

It follows from Lemma 1 that the bracketing entropy of \(\mathcal{H}_q(\epsilon)\) is
\[
H_{1/2}(\epsilon) = \log N_{1/2}(\mathcal{H}_q(\epsilon), \epsilon) \leq (10(C_2\delta^{-\frac{1}{2}} + 1)q + 1) \log \left(\frac{C_1\epsilon}{\delta}\right).
\]
Then,
\[
\int_{0}^{\epsilon} H_{1/2}(\epsilon) d\epsilon \leq c \sqrt{q} \epsilon^{1 - \frac{r}{2}}.
\]
The remainder of proof is identical to that in Section 5.6 of [4]. We obtain
\[
\sup_{f \in \mathcal{M}_q} \frac{|\nu_n(g_h) - \nu_n(g_{h^*})|}{\sqrt{q} \mathcal{H}(h, h^*)^{1 - \frac{r}{2}} \sqrt{n}^{-\frac{r-1}{2r+1}}} = O_p(1). 
\]
This allows us to conclude that
\[
\sqrt{n} |\nu_n(g_h) - \nu_n(g_{h^*})| = \sqrt{n} \frac{|\nu_n(g_h) - \nu_n(g_{h^*})|}{\sqrt{q} \mathcal{H}(h, h^*)^{1 - \frac{r}{2}} \sqrt{n}^{-\frac{r-1}{2r+1}}} \mathcal{H}(\hat{h}, h^*)^{1 - \frac{1}{2r}} = O_p(\sqrt{n}) \sqrt{q} \mathcal{H}(\hat{h}, h^*)^{1 - \frac{1}{2r}}.
\]
Combining this with (11) yields that \(\mathcal{H}(\hat{h}, h^*) = O_p(n^{-r/(2r+1)})\) provided that \(\lambda\) is of order \(n^{1/(2r+1)}\). This finishes the proof of Theorem.
0.2.3 Proof of Theorem 3

First note that

\[
\lim_{n \to \infty} \sup_{q > q^*} \Delta_n(q, q^*) \leq \lim_{n \to \infty} \sup_{q > q^*} \frac{q^{\frac{2r}{r + 1} n^{\frac{1}{r + 1}}} \Delta_n(q, q^*)}{\text{pen}_n(q) - \text{pen}_n(q^*)} = 0.
\]

Therefore, for all \(q > q^*\),

\[
\sup_{f \in \mathcal{M}_q} \left( \ell_n(f) - \lambda_{q,n} \|\beta\|_K^2 \right) - \text{pen}_n(q) < \sup_{f \in \mathcal{M}_{q^*}} \left( \ell_n(f) - \lambda_{q^*,n} \|\beta\|_K^2 \right) - \text{pen}_n(q^*).
\]

This shows that \(\lim_{n \to \infty} \hat{q}_n \leq q^*\) a.s., which means that we do not asymptotically overestimate the order.

On the other hand, for any \(q < q^*\),

\[
\lim_{n \to \infty} \frac{1}{n} \Delta_n(q, q^*) \leq \lim_{n \to \infty} \sup_{f \in \mathcal{M}_q} \frac{1}{n} \sum_{j=1}^{n} \log \frac{f(Y_i|X_i)}{f^*(Y_i|X_i)} + \lim_{n \to \infty} \frac{\lambda_{q,n}}{n} \|\beta^*\|_K^2
\]

which is strictly negative based on Part (b) of Lemma 2. Since \(\text{pen}_n(q)/n \to 0\) as \(n \to \infty\) for \(q < q^*\), we have

\[
\lim_{n \to \infty} \frac{1}{n} \left\{ \Delta_n(q, q^*) - \text{pen}_n(q) + \text{pen}_n(q^*) \right\} < 0, \text{ a.s.}
\]

We obtain, for all \(q < q^*\),

\[
\sup_{f \in \mathcal{M}_q} \left( \ell_n(f) - \lambda_{q,n} \|\beta\|_K^2 \right) - \text{pen}_n(q) < \sup_{f \in \mathcal{M}_{q^*}} \left( \ell_n(f) - \lambda_{q^*,n} \|\beta\|_K^2 \right) - \text{pen}_n(q^*).
\]

This shows that \(\lim_{n \to \infty} \hat{q}_n \geq q^*\) a.s., which means that we do not asymptotically underestimate the order.
Bibliography


