Fast Nonparametric Maximum Likelihood Density Deconvolution Using Bernstein Polynomials

Zhong Guan

Indiana University South Bend

Supplementary Material

Additional Simulation Results

Table 1: The square root multiplied by 100 of the mean integrated squared error. $\hat{f}_{\mathrm{P}}$, the parametric estimator; $\hat{f}_{\mathrm{B}}$, the proposed estimator; $\hat{f}_{\mathrm{F}}$, the inverse Fourier estimator; $\tilde{f}_{\mathrm{K}}$, the kernel density estimator based on the uncontaminated data, based on 1000 Monte Carlo runs with $x_1, \ldots, x_n$ being generated from normal and mixture normal distributions and errors $\varepsilon_1, \ldots, \varepsilon_n$ from $\mathrm{N}(0, \sigma_0^2)$ and $\mathrm{L}(0, \sigma_0)$. In the parametric models the variances are assumed to be known. $\mathcal{M} = \{10, 11, \ldots, 100\}$ and $n = 200$.

	$X \sim \mathrm{N}(0,1)$					$X \sim 0.6\mathrm{N}(-2,1) + 0.4\mathrm{N}(2, 0.8^2)$				
σ_0	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
					$\varepsilon \sim \mathrm{N}(0, \sigma_0^2)$					
$\hat{f}_{\mathrm{P}}$	2.72	2.93	3.17	3.48	3.68	4.09	4.31	4.75	5.45	6.35
$\hat{f}_{\mathrm{B}}$	3.81	4.07	4.36	4.75	5.08	5.71	6.05	6.60	7.45	8.51
$\hat{f}_{\mathrm{F}}$	7.47	8.93	11.04	13.80	16.63	6.96	11.56	13.44	15.37	17.20
$\tilde{f}_{\mathrm{K}}$	6.23	6.27	6.31	6.16	6.24	6.23	6.36	6.23	6.26	6.34
					$\varepsilon \sim \mathrm{L}(0, \sigma_0)$					
$\hat{f}_{\mathrm{P}}$	2.78	3.14	3.62	3.78	4.51	4.26	4.71	5.46	6.38	7.56
$\hat{f}_{\mathrm{B}}$	3.87	4.30	5.00	5.44	6.38	5.88	6.41	7.28	8.34	9.64
$\hat{f}_{\mathrm{F}}$	11.69	17.39	22.97	27.59	31.42	10.32	15.03	18.19	22.40	25.94
$\tilde{f}_{\mathrm{K}}$	6.23	6.09	6.12	6.33	6.23	6.34	6.24	6.34	6.31	6.26

Framingham Data

The Framingham data is from a study on coronary heart disease (Carroll et al., 2006) and consist of measurements of systolic blood pressure (SBP) in 1,615 males, Y_1 taken at an examination and Y_2 at an 8-year follow-up

Table 2: The square root multiplied by 100 of the mean integrated squared error. $\hat{f}_{\mathrm{P}}$, the parametric estimator; $\hat{f}_{\mathrm{B}}$, the proposed estimator; $\hat{f}_{\mathrm{F}}$, the inverse Fourier estimator; $\tilde{f}_{\mathrm{K}}$, the kernel density estimator based on the uncontaminated data, based on 1000 Monte Carlo runs with $x_1, \ldots, x_n$ being generated from the nearly normal distribution NN(4) and errors $\varepsilon_1, \ldots, \varepsilon_n$ from the normal $\mathrm{N}(0, \sigma_0^2)$ and Laplace $\mathrm{L}(0, \sigma_0)$. We assume the normal distribution $\mathrm{N}(\mu, \sigma^2)$ with known variance $\sigma^2 = 1/48$ as the parametric model. $\mathcal{M} = \{2, 3, \ldots, 100\}$ and $n = 200$.

	$X \sim \mathrm{NN}(4)$, $\varepsilon \sim \mathrm{N}(0, \sigma_0^2)$					$X \sim \mathrm{NN}(4)$, $\varepsilon \sim \mathrm{L}(0, \sigma_0)$				
$\sqrt{3}\sigma_0$	0.05	0.10	0.15	0.20	0.25	0.05	0.10	0.15	0.20	0.25
$\hat{f}_{\mathrm{P}}$	8.34	8.61	9.13	9.94	10.60	8.50	9.34	10.13	11.15	11.73
$\hat{f}_{\mathrm{B}}$	16.49	19.71	24.11	29.13	34.43	17.30	22.41	27.72	33.08	37.86
$\hat{f}_{\mathrm{F}}$	20.09	49.60	80.42	95.17	95.80	21.47	26.65	31.80	37.09	41.22
$\tilde{f}_{\mathrm{K}}$	15.90	15.89	15.94	15.92	15.69	16.04	16.28	16.01	15.96	16.28

Table 3: The square root multiplied by 100 of the mean integrated squared error. $\hat{f}_{\mathrm{P}}$, the parametric estimator; $\hat{f}_{\mathrm{B}}$, the proposed estimator; $\hat{f}_{\mathrm{F}}$, the inverse Fourier estimator; $\tilde{f}_{\mathrm{K}}$, the kernel density estimator based on the uncontaminated data, based on 1000 Monte Carlo runs with $x_1, \ldots, x_n$, $n = 100$, being generated from the beta distribution with shapes $(3.5, 5.5)$ and errors $\varepsilon_1, \ldots, \varepsilon_n$ from the normal $\mathrm{N}(0, \sigma_0^2)$ and Laplace $\mathrm{L}(0, \sigma_0)$. We used the method of moment estimators to obtain $\hat{f}_{\mathrm{P}}$. $\mathcal{M} = \{2, 3, \ldots, 100\}$

	$X \sim \mathrm{beta}(3.5, 5.5)$, $\varepsilon \sim \mathrm{N}(0, \sigma_0^2)$					$X \sim \mathrm{beta}(3.5, 5.5)$, $\varepsilon \sim \mathrm{L}(0, \sigma_0)$				
σ_0/σ	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
$\hat{f}_{\mathrm{P}}$	16.55	19.25	21.60	24.72	27.17	15.42	17.04	18.81	21.80	29.56
$\hat{f}_{\mathrm{B}}$	13.62	15.61	17.40	21.12	25.49	13.68	15.10	19.77	26.65	34.60
$\hat{f}_{\mathrm{F}}$	23.42	54.53	83.99	88.71	88.92	24.51	27.66	31.63	35.33	38.48
$\tilde{f}_{\mathrm{K}}$	20.35	20.85	20.46	20.38	20.94	20.69	20.50	20.07	20.85	20.22

examination after the first. At the ith examination, the SBP was measured twice, Y_{i1} and Y_{i2} $(i = 1, 2)$, for each individual. Assuming normal error ε_i with mean zero for each individual, then ε_i and $\tilde{\varepsilon}_i = (Y_{i1} - Y_{i2})/\sqrt{2}$ have the same distribution. Q-Q plots suggest that the mixture normal models $\lambda_i N(0, \sigma_{i1}^2) + (1 - \lambda_i)N(0, \sigma_{i2}^2)$ fit better than single normals. After fitting $\tilde{\varepsilon}_i$ with this mixture normal model we obtained $\lambda_1 = 0.6592$, $(\sigma_{11}, \sigma_{12}) = (5.45, 10.67)$, $\lambda_2 = 0.8227$, and $(\sigma_{21}, \sigma_{22}) = (6.40, 12.55)$. We estimated the densities of Y_i based $\bar{Y}_i = (Y_{i1}+Y_{i2})/2 = X_i+\bar{\varepsilon}_i$, where $\bar{\varepsilon}_i = (\varepsilon_{i1}+\varepsilon_{i2})/2$ has a population error distribution $\lambda_i N(0, \sigma_{i1}^2/2) + (1-\lambda_i)N(0, \sigma_{i2}^2/2)$, $i = 1, 2$. The Bernstein polynomial density estimates are obtained on interval $[a, b] = [70, 270]$ using the optimal degree $\hat{m} = 35$ selected from $\mathbb{M} = \{5, 6, \ldots, 100\}$. The kernel density estimate $\hat{f}_{\mathrm{F}}$ is produced by R package decon (Wang and Wang, 2011). The parametric estimate $\hat{f}_{\mathrm{P}}$ was obtained by maximum likelihood method using the log-normal model with the estimated mixture normal error distributions. We also calculated kernel density estimate $\tilde{\psi}_{\mathrm{K}}$ by ignoring measurement errors. Figure 1 shows that the difference between $\hat{f}_{\mathrm{B}}$, $\hat{f}_{\mathrm{F}}$, and $\tilde{\psi}_{\mathrm{K}}$ are noticeable.

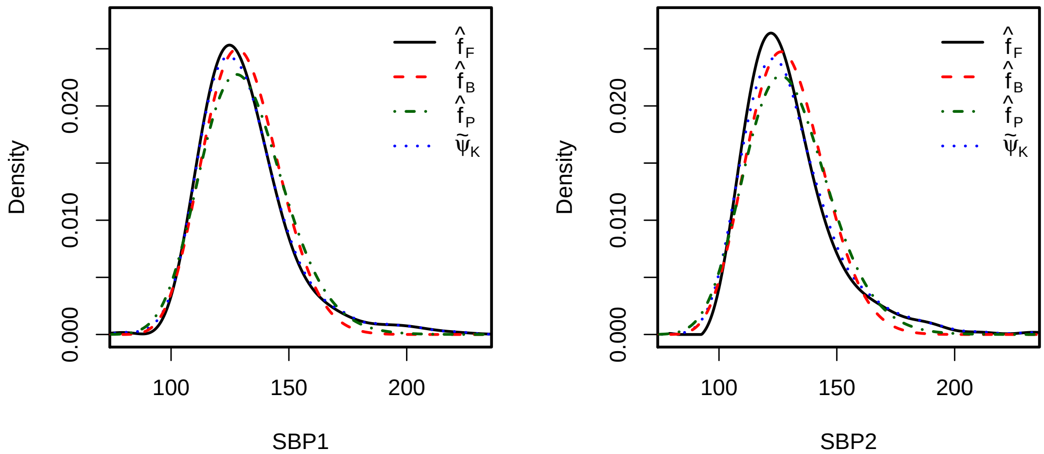

Figure 1: Left (right) panel: Density deconvolution of the systolic blood pressure at the first(second) examination based on Framingham data, $\hat{f}_{\mathrm{F}}$ is the inverse Fourier transform estimate(solid); $\hat{f}_{\mathrm{B}}$ is the proposed estimate using Bernstein polynomial with $m = 35$(dashed); $\hat{f}_{\mathrm{P}}$ is the parametric estimate using lognormal model; $\tilde{\psi}_{\mathrm{K}}$ is the kernel estimate ignoring measurement errors (dotted).

Proof of Proposition 1

Proof. By Cauchy-Schwarz inequality, for any function p (not necessarily density)

$$[p * g(y)]^2 \leq f * g(y) \int \frac{p^2(x)}{f(x)} g(y-x)dx. \tag{S0.1}$$

Applying (S0.1) with $p = h - f$

$$\chi^2(h * g \| f * g) = \int \frac{[(h-f) * g(y)]^2}{f * g(y)} dy \leq \iint \frac{[h(x) - f(x)]^2}{f(x)} g(y-x)dxdy$$
$$= \int \frac{[h(x) - f(x)]^2}{f(x)} dx = \chi^2(h \| f).$$

Thus, part (i) is true. For part (ii) we have $\chi^2(h * g \| f * g) = 0$ iff $h * g(y) = f * g(y)$ almost everywhere. Then, the characteristic functions of h and f are identical. This means that h and f are identical almost everywhere. □

Proof of Theorem 1.

Proof. By Theorem 1 of Lorentz (1963) we have $f_0(x) - P_m(x) = R_m(x)$, where $P_m(x)$ is a polynomial with positive coefficients and $|R_m(x)| \leq C_0(f)$ $m^{-(r+\alpha)/2}$, $0 \leq x \leq 1$. So $f(x) - Q_{\tilde{m}}(x) = R_{\tilde{m}}(x)$, where $Q_{\tilde{m}}(x) = x^a(1-x)^b P_m(x) = \sum_{i=0}^{\tilde{m}} a_i \cdot \beta_{\tilde{m}i}(x)$ is a polynomial of degree $\tilde{m} = m+a+b$ with positive coefficients, $R_{\tilde{m}}(x) = x^a(1-x)^b R_m(x)$, and $|R_{\tilde{m}}(x)| \leq C_0(f)m^{-(r+\alpha)/2}$, $0 \leq x \leq 1$. For large m, $\rho_{\tilde{m}} := \int_0^1 R_{\tilde{m}}(x)dx \leq C_0(d, f)m^{-(r+\alpha)/2} < c_0 < 1$. Because $f(x)$ and $\beta_{\tilde{m}i}(x)$ are densities on $[0, 1]$, $\sum_{i=0}^{\tilde{m}} a_i = 1 - \rho_{\tilde{m}} > 0$. Normalizing a_i we obtain $f_{\tilde{m}}(x; \boldsymbol{p}_0) = Q_{\tilde{m}}(x)/(1 - \rho_{\tilde{m}}) = \sum_{i=0}^{\tilde{m}} p_{0i}\beta_{\tilde{m}i}(x)$, where $p_{0i} = a_i/(1 - \rho_{\tilde{m}})$. Noticing that $f_0(x) \geq b_0 > 0$, we have

$$\begin{aligned}|f_{\tilde{m}}(x; \boldsymbol{p}_0) - f(x)|/f(x) &= (1 - \rho_{\tilde{m}})^{-1} |R_{\tilde{m}}(x)/f(x) + \rho_{\tilde{m}}| \\ &= (1 - \rho_{\tilde{m}})^{-1} |R_m(x)/f_0(x) + \rho_{\tilde{m}}| \\ &\leq (1 - c_0)^{-1} C_0(f)(1/\delta_0 + 1)m^{-(r+\alpha)/2}.\end{aligned}$$

Therefore, (A.2.) is implied. (A.1.) is implied by (A.2). The proof is complete. □

Proof of Theorem 2.

Proof. The approximate Bernstein log likelihood is

$$\ell(f_m) = \ell(\boldsymbol{p}) = \sum_{i=1}^{n} \log[\psi_m(y_i; \boldsymbol{p})].$$

Define the log-likelihood ratio $\mathcal{R}(\boldsymbol{p}) = \ell(f) - \ell(\boldsymbol{p})$, where

$$\ell(f) = \sum_{i=1}^{n} \log \psi(y_i) = \sum_{i=1}^{n} \log(f * g)(y_i).$$

Because $\log(1+z) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{z^k}{k}$, $|z| < 1$, we have

$$\log(x) \;=\; \log a + \sum_{k=1}^{2}(-1)^{k+1}\frac{1}{k}\left(\frac{x-a}{a}\right)^k + R_2, \quad |x-a| < a, \tag{S0.2}$$

where $R_2 = \sum_{k=3}^{\infty}(-1)^{k+1}\frac{1}{k}\left(\frac{x-a}{a}\right)^k$. Clearly,

$$|R_2| = \mathcal{O}\left(\left|\frac{x-a}{a}\right|^3\right) = o\left(\left|\frac{x-a}{a}\right|^2\right), \quad |x-a|/a \to 0.$$

Consider subset $\mathcal{A}_m(\epsilon_n)$ of $\mathbb{S}_m$ so that, for all $y \in R$, $|\psi_m(y; \boldsymbol{p}) - \psi(y)|/\psi(y) \le \epsilon_n < 1$, $0 < \epsilon_n \searrow 0$ slowly, as $n \to \infty$, e.g., $\epsilon_n = 1/\log(n+2)$. Clearly $\boldsymbol{p}_0 \in \mathcal{A}_m(\epsilon_n)$ for large m, $\mathcal{A}_m(\epsilon_n)$ is nonempty. By (S0.2) we have

$$\mathcal{R}(\boldsymbol{p}) \;=\; -\sum_{i=1}^{n}\left[Z_i(\boldsymbol{p}) - \frac{1}{2}Z_i^2(\boldsymbol{p})\right] + o(R_{mn}(\boldsymbol{p})),\ a.s.,$$

where $R_{mn}(\boldsymbol{p}) = \sum_{i=1}^{n} Z_i^2(\boldsymbol{p})$, and $Z_i(\boldsymbol{p}) = [\psi_m(y_i; \boldsymbol{p}) - \psi(y_i)]/\psi(y_i)$, $i \in \mathbb{I}_1^n$. Because $\mathrm{E}[Z_i(\boldsymbol{p})] = 0$, $\sigma^2[Z_i(\boldsymbol{p})] = \mathrm{E}[Z_i^2(\boldsymbol{p})] = D^2(\boldsymbol{p})$, by the law of iterated logarithm(LIL) we have, for all $\boldsymbol{p} \in \mathcal{A}_m(\epsilon_n)$,

$$\sum_{i=1}^{n} Z_i(\boldsymbol{p})/\sigma[Z_i(\boldsymbol{p})] = \mathcal{O}(\sqrt{n \log\log n}), \quad a.s..$$

By the Kolmogorov's strong law of large numbers we have, for all $\boldsymbol{p} \in \mathcal{A}_m(\epsilon_n)$,

$$\mathcal{R}(\boldsymbol{p}) = \frac{n}{2}D^2(\boldsymbol{p}) - \mathcal{O}(D(\boldsymbol{p})\sqrt{n\log\log n}) + o(nD^2(\boldsymbol{p})),\ a.s.. \tag{S0.3}$$

If $D^2(\boldsymbol{p}) = r_n = \log n/n$ for some $\boldsymbol{p} \in \mathcal{A}_m(\epsilon_n)$ then, by (S0.3), there is an $\eta > 0$ such that $\mathcal{R}(\boldsymbol{p}) \geq \eta \log n$, *a.s.*. At $\boldsymbol{p} = \boldsymbol{p}_0$, if $m = Cn^{1/k}$ we have $D^2(\boldsymbol{p}_0) = \chi^2(\psi_m(\cdot; \boldsymbol{p}_0) \| \psi) = \mathcal{O}(m^{-k}) = \mathcal{O}(n^{-1})$. By (S0.3) again we have $\mathcal{R}(\boldsymbol{p}_0) = \mathcal{O}(\sqrt{\log \log n})$, a.s.. Therefore, similar to the proof of Lemma 1 of Qin and Lawless (1994), we have

$$D^2(\hat{\boldsymbol{p}}) = \int_{S_\psi} \frac{[\psi_m(y; \hat{\boldsymbol{p}}) - \psi(y)]^2}{\psi(y)} dy < \frac{\log n}{n}, a.s.,$$

and $\hat{\boldsymbol{p}} \in \mathcal{A}_m(\epsilon_n)$. The proof is complete. □

In order to prove Theorem 3 we need the following lemma.

Lemma 1. *If $f \in C^{(r)}[0,1]$ and $f(x) \geq b_0 > 0$ on [0,1], then, for m large enough, there exists $f_m(x; \boldsymbol{p}_0)$ that fulfills both (A1) and (A2) and with coefficients satisfying* $0 < c_0 < (m+1)p_{i0} < c_1 < \infty$.

Proof. Lorentz (1963) proved that, under the conditions of his Theorem 1, for $r = 0, 1, 2, \ldots$, there exist polynomials of the form, using his notations,

$$Q_{nr}^f(x) = \sum_{k=0}^{n} \left\{ f(\tfrac{k}{n}) + \sum_{i=2}^{r} f^{(i)}(\tfrac{k}{n}) \frac{1}{n^i} \tau_{ri}(x, n) \right\} b_{nk}(x) \tag{S0.4}$$

such that for each function with first r continuous derivatives

$$|f(x) - Q_{nr}^f(x)| \leq C_r' \Delta_n^r \omega_r(\Delta_n);$$

C_r' depends only upon r. The $\tau_{ri}(x, n)$ are some polynomials in x and n, independent of f, in x of degree i, in n of degree $[i/2] = \lfloor i/2 \rfloor$.

Assuming that $f(x) \geq b_0 > 0$, Lorentz (1963) then proved that $Q_{nr}^f(x)$ is a polynomial with positive coefficients of degree $n + r$ (see Remark (a) of Lorentz, 1963).

Assuming that $f^{(i)}(\frac{k}{n})$ are all bounded as in Theorem 1 of Lorentz (1963), we know that, for $r \geq 2$, $\tilde{Q}_{kr}(x, n^{-1}) := \sum_{i=2}^{r} f^{(i)}(\frac{k}{n}) \frac{1}{n^i} \tau_{ri}(x, n)$ is a polynomial of degree r with coefficients $c_{ki} = \mathcal{O}(n^{-1})$, $i = 0, \ldots, r$. By Remark (a) of Lorentz (1963), for large n,

$$f(\tfrac{k}{n}) + \tilde{Q}_{kr}(x, n^{-1}) = \sum_{j=0}^{r} [f(\tfrac{k}{n}) + a_{kj}] b_{rj}(x),$$

where, uniformly in $k \in \{0, \ldots, n\}$,

$$a_{kj} = \binom{r}{j}^{-1} \sum_{i=0}^{j} c_{ki} \binom{r-i}{j-i} = \mathcal{O}(n^{-1}), \quad j = 0, \ldots, r.$$

Thus, for large n, we have

$$
\begin{aligned}
Q_{nr}^{f}(x) &= \sum_{k=0}^{n}\left\{f(\tfrac{k}{n})+\sum_{j=0}^{r}a_{kj}b_{rj}(x)\right\}b_{nk}(x)\\
&= \sum_{k=0}^{n}f(\tfrac{k}{n})b_{nk}(x)+\sum_{k=0}^{n}\sum_{j=0}^{r}a_{kj}b_{rj}(x)b_{nk}(x),
\end{aligned}
$$

where

$$
\sum_{k=0}^{n}f(\tfrac{k}{n})b_{nk}(x)=\sum_{k=0}^{n+r}\alpha_{n+r,k}b_{n+r,k}(x)
$$

with coefficients

$$
\alpha_{n+r,j}=\sum_{i=0}^{n}\frac{\binom{n}{i}\binom{r}{j-i}}{\binom{n+r}{j}}f(\tfrac{i}{n})\geq\min_{0\leq x\leq 1}f(x),\quad j=0,\ldots,n+r. \tag{S0.5}
$$

Let V_{nrj} be a random variable having hypergeometric distribution. Then

$$
\alpha_{n+r,j}=\mathrm{E}\Big[f(\tfrac{V_{nrj}}{n})\Big],\quad j=0,\ldots,n+r.
$$

Thus

$$
\begin{aligned}
\sum_{k=0}^{n}\sum_{j=0}^{r}a_{kj}b_{rj}(x)b_{nk}(x) &= \sum_{k=0}^{n}\sum_{j=0}^{r}a_{kj}\binom{r}{j}\binom{n}{k}x^{j+k}(1-x)^{n+r-j-k}\\
&= \sum_{k=0}^{n}\sum_{j=0}^{r}\frac{a_{kj}\binom{r}{j}\binom{n}{k}}{\binom{n+r}{j+k}}b_{n+r,j+k}(x)\\
&\overset{l=j+k}{=\!=\!=} \sum_{l=0}^{n+r}\sum_{k=0}^{n}\frac{a_{k,l-k}\binom{r}{l-k}\binom{n}{k}}{\binom{n+r}{l}}b_{n+r,l}(x).
\end{aligned}
$$

Consequently

$$
Q_{nr}^{f}(x)=\sum_{k=0}^{n+r}c_{n+r,k}b_{n+r,k}(x), \tag{S0.6}
$$

where

$$\begin{aligned}c_{n+r,j} &= \alpha_{n+r,j} + \sum_{k=0}^{n} \frac{a_{k,j-k}\binom{r}{j-k}\binom{n}{k}}{\binom{n+r}{j}} \\ &= \sum_{i=0}^{n} \frac{\binom{n}{i}\binom{r}{j-i}}{\binom{n+r}{j}} f(\tfrac{i}{n}) + \sum_{k=0}^{n} \frac{a_{k,j-k}\binom{r}{j-k}\binom{n}{k}}{\binom{n+r}{j}} \\ &= \sum_{i=0}^{n} \frac{\binom{n}{i}\binom{r}{j-i}}{\binom{n+r}{j}} [f(\tfrac{i}{n}) + a_{i,j-i}], \quad j = 0, \ldots, n+r.\end{aligned}$$

Because

$$\sum_{i=0}^{n} \frac{\binom{n}{i}\binom{r}{j-i}}{\binom{n+r}{j}} = 1$$

we have, for $j = 0, \ldots, n+r$,

$$\begin{aligned}c_{n+r,j} &\le \sum_{i=0}^{n} \frac{\binom{n}{i}\binom{r}{j-i}}{\binom{n+r}{j}} f(\tfrac{i}{n}) + \max_{0\le i\le n} |a_{i,j-i}| \\ &= \alpha_{n+r,j} + \mathcal{O}(n^{-1}),\end{aligned}$$

$$\min_{x\in[0,1]} f(x) + \mathcal{O}(n^{-1}) \le c_{n+r,j} \le \max_{x\in[0,1]} f(x) + \mathcal{O}(n^{-1}).$$

Therefore, for all large n and some $\eta > 0$,

$$\begin{aligned}c_{n+r,j} &\ge \alpha_{n+r,j} + \mathcal{O}(n^{-1}) \\ &\ge \min_{0\le x\le 1} f(x) + \mathcal{O}(n^{-1}) \\ &\ge \eta \min_{0\le x\le 1} f(x) > 0, \quad j = 0, \ldots, n+r.\end{aligned}$$

Combining the above with the proof of Theorem 1, we can easily see that $p_{0i}(m+1) = c_{n+r,j}/(1-\rho_{n+r}) \ge \eta \min_{0\le x\le 1} f(x)/(1-\rho_{n+r}) := c_0 > 0$, $m = n+r$. Similarly, $p_{0i}(m+1) \le c_1$.

□

Proof of Theorem 3.

Proof. The following identities are useful in this proof that, for $a \ne 0$ and $x \ne 0$,

$$\frac{1}{x} = \frac{1}{a} + \frac{a-x}{ax} = \frac{1}{a}\Big(1 + \frac{a-x}{a}\Big) + \frac{1}{x}\Big(\frac{a-x}{a}\Big)^2. \tag{S0.7}$$

Define an empirical Fisher information matrix

$$\widehat{\mathcal{J}}_m(\boldsymbol{p}) = [\widehat{J}_m^{ij}(\boldsymbol{p})] = -\Big[\frac{\partial S_{mn}^{(i)}(\boldsymbol{p})}{\partial p_j}\Big],$$

where

$$\widehat{J}_m^{ij}(\boldsymbol{p}) = \frac{1}{n}\sum_{l=1}^{n}\frac{\psi_{mi}(y_l)\psi_{mj}(y_l)}{\psi_m^2(y_l;\boldsymbol{p})}, \quad i,j\in\mathbb{I}_0^m.$$

In order to estimate the order of $\mathbb{R}_n = \mathcal{I}_m(f,g)(\hat{\boldsymbol{p}}-\boldsymbol{p}_0)$, we need to estimate $\bar{\mathbb{R}}_n = \widehat{\mathcal{J}}_m(\boldsymbol{p}_0)(\hat{\boldsymbol{p}}-\boldsymbol{p}_0)$ and $\bar{\mathbb{R}}_n - \mathbb{R}_n$.

Define $V_{mi}(Y) = \psi_{mi}(Y)/\psi(Y)$. Then $\mathrm{E}[V_{mi}(Y)] = 1$ and $\sigma_{mi}^2 \equiv \sigma^2\left[V_{mi}(Y)\right] \equiv \tau_{mi}^2 - 1$. Thus, the LIL ensures that, for all $i\in\mathbb{I}_0^m$,

$$\Sigma_{mn}^{(i)} \equiv n^{-1}\sum_{j=1}^{n} V_{mi}(y_j) = 1 + \mathcal{O}(\sigma_{mi}\sqrt{\log\log n/n}),\ a.s. \tag{S0.8}$$

It is necessary and sufficient (Redner and Walker, 1984) for $\hat{\boldsymbol{p}}$ to maximize $\ell(\boldsymbol{p})$ that $S_{mn}^{(i)}(\hat{\boldsymbol{p}}) \le 1$, $i\in\mathbb{I}_0^m$, with equality when $\hat{p}_i > 0$, where $S_{mn}^{(i)}(\boldsymbol{p}) = n^{-1}\sum_{j=1}^{n}\psi_{mi}(y_j)/\psi_m(y_j;\boldsymbol{p})$. Thus we have, for all $i\in\mathbb{I}_0^m$,

$$\hat{p}_i S_{mn}^{(i)}(\hat{\boldsymbol{p}}) = \hat{p}_i. \tag{S0.9}$$

Estimation of $\bar{\mathbb{R}}_n$: First, we estimate the differences $\Sigma_{mn}^{(i)} - S_{mn}^{(i)}(\boldsymbol{p}_0)$ and $S_{mn}^{(i)}(\boldsymbol{p}_0) - 1$.

For each $i\in\mathbb{I}_0^m$, by (A.1.) and the first equation of (S0.7)

$$S_{mn}^{(i)}(\boldsymbol{p}_0) - \Sigma_{mn}^{(i)} = -\frac{1}{n}\sum_{j=1}^{n}\frac{\psi_{mi}(y_j)Z_j(\boldsymbol{p}_0)}{\psi_m(y_j;\boldsymbol{p}_0)} = \mathcal{O}(n^{-1/2})S_{mn}^{(i)}(\boldsymbol{p}_0), \tag{S0.10}$$

Then, we have, for all $i\in\mathbb{I}_0^m$,

$$\begin{aligned}\tilde{R}_{ni}^* &\equiv S_{mn}^{(i)}(\boldsymbol{p}_0) - \Sigma_{mn}^{(i)} = \mathcal{O}(\Sigma_{mn}^{(i)} n^{-1/2})\\ &= \mathcal{O}(n^{-1/2}) + \mathcal{O}(\sigma_{mi}\sqrt{\log\log n}/n),\ a.s.\end{aligned} \tag{S0.11}$$

Combining (S0.8) through (S0.11) we obtain, a.s., for all $i\in\mathbb{I}_0^m$,

$$\tilde{R}_{ni} \equiv S_{mn}^{(i)}(\boldsymbol{p}_0) - 1 = \mathcal{O}(\sigma_{mi}\sqrt{\log\log n/n}) + \mathcal{O}(n^{-1/2}). \tag{S0.12}$$

Secondly, we obtain an asymptotic expression for $\bar{\mathbb{R}}_n$ and use it to get an estimate. By (S0.9) and (S0.7) again, we have, for all $i\in\mathbb{I}_0^m$,

$$\hat{p}_i \;=\; \hat{p}_i S_{mn}^{(i)}(\boldsymbol{p}_0) - \hat{p}_i\sum_{j=0}^{m}\widehat{J}_m^{ij}(\boldsymbol{p}_0)(\hat{p}_j - p_{j0}) + \hat{R}_{ni}, \tag{S0.13}$$

where, by (3),

$$\hat{R}_{ni} = \frac{1}{n}\sum_{j=1}^{n} \frac{\hat{p}_i\psi_{mi}(y_j)}{\psi_m(y_j;\hat{\boldsymbol{p}})} \frac{[\psi_m(y_j;\boldsymbol{p}_0) - \psi_m(y_j;\hat{\boldsymbol{p}})]^2}{\psi_m^2(y_j;\boldsymbol{p}_0)} = \mathcal{O}(\log n/n),\ a.s. \tag{S0.14}$$

Combining (S0.12) and (S0.13), we have, for all $i \in \mathbb{I}_0^m$,

$$\hat{p}_i \sum_{j=0}^{m} \widehat{J}_m^{ij}(\boldsymbol{p}_0)(\hat{p}_j - p_{j0}) = \hat{p}_i\tilde{R}_{ni} + \hat{R}_{ni},\ a.s. \tag{S0.15}$$

Defining

$$\boldsymbol{\Delta}_n = \mathrm{diag}\Big\{\sum_{j=0}^{m} \widehat{J}_m^{0j}(\boldsymbol{p}_0)(\hat{p}_j - p_{j0}), \ldots, \sum_{j=0}^{m} \widehat{J}_m^{mj}(\boldsymbol{p}_0)(\hat{p}_j - p_{j0})\Big\}, \tag{S0.16}$$

$\boldsymbol{\Sigma}_n = (\Sigma_0, \ldots, \Sigma_m)^{\mathrm{T}}$, $\tilde{\boldsymbol{R}}_n = (\tilde{R}_{n0}, \ldots, \tilde{R}_{nm})^{\mathrm{T}}$, and $\hat{\boldsymbol{R}}_n = (\hat{R}_{n0}, \ldots, \hat{R}_{nm})^{\mathrm{T}}$, we have, in matrix form,

$$\Pi_0\widehat{\mathcal{J}}_m(\boldsymbol{p}_0)(\hat{\boldsymbol{p}} - \boldsymbol{p}_0) = \hat{\Pi}\tilde{\boldsymbol{R}}_n + \hat{\boldsymbol{R}}_n - \boldsymbol{\Delta}_n(\hat{\boldsymbol{p}} - \boldsymbol{p}_0), \tag{S0.17}$$

where $\Pi_0 = \mathrm{diag}(p_{00}, \ldots, p_{m0})$ and $\hat{\Pi} = \mathrm{diag}(\hat{p}_0, \ldots, \hat{p}_m)$. Because $f(x) \geq b_0 > 0$, by Lemma 1, $0 < c_0 < (m+1)p_{i0} < c_1 < \infty$. The ith component of $\boldsymbol{\Delta}_n$ can be written as

$$\begin{aligned}\Delta_{mn}^{(i)} &= \sum_{j=0}^{m} \widehat{J}_m^{ij}(\boldsymbol{p}_0)(\hat{p}_j - p_{j0}) \\ &= \frac{1}{n}\sum_{l=1}^{n} \frac{\psi_{mi}(y_l)}{\psi_m^2(y_l;\boldsymbol{p}_0)}[\psi_m(y_l;\hat{\boldsymbol{p}}) - \psi_m(y_l;\boldsymbol{p}_0)].\end{aligned}$$

Thus, we have

$$\begin{aligned}|\Delta_{mn}^{(i)}| &\leq p_{i0}^{-1}\frac{1}{n}\sum_{l=1}^{n} \frac{|\psi_m(y_l;\hat{\boldsymbol{p}}) - \psi_m(y_l;\boldsymbol{p}_0)|}{\psi_m(y_l;\boldsymbol{p}_0)} \\ &\leq p_{i0}^{-1}\left(\frac{1}{n}\sum_{l=1}^{n} \frac{|\psi_m(y_l;\hat{\boldsymbol{p}}) - \psi_m(y_l;\boldsymbol{p}_0)|^2}{\psi_m^2(y_l;\boldsymbol{p}_0)}\right)^{1/2} \\ &= \mathcal{O}(p_{i0}^{-1}\sqrt{\log n/n}).\end{aligned}$$

Therefore, we have an asymptotic expression $\bar{\mathbb{R}}_n = \Pi_0^{-1}[\hat{\Pi}\tilde{\boldsymbol{R}}_n + \hat{\boldsymbol{R}}_n - \boldsymbol{\Delta}_n(\hat{\boldsymbol{p}} - \boldsymbol{p}_0)]$ and

$$\|\bar{\mathbb{R}}_n\|^2 = \mathcal{O}\Big(\sum_{i=0}^{m} \frac{\hat{p}_i^2}{p_{i0}^2}\sigma_{mi}^2 \frac{\log\log n}{n}\Big) + \mathcal{O}\Big(\frac{1}{n}\sum_{i=0}^{m}\frac{\hat{p}_i^2}{p_{i0}^2}\Big) + \mathcal{O}\Big(\sum_{i=0}^{m} p_{i0}^{-2}\Big(\frac{\hat{p}_i}{p_{i0}} - 1\Big)^2 \frac{\log n}{n}\Big). \quad \text{(S0.18)}$$

Because $0 < c_0 < (m+1)p_{i0} < c_1 < \infty$, $\sum_{i=0}^{m} p_{i0}^{-2}(\hat{p}_i/p_{i0} - 1)^2 = \mathcal{O}(m^4) = \mathcal{O}(n^{4/k})$. By (2), we have

$$\|\bar{\mathbb{R}}_n\|^2 = \mathcal{O}\Big(m^2\sum_{i=0}^{m}\sigma_{mi}^2\frac{\log\log n}{n}\Big) + \mathcal{O}\Big(\frac{\log n}{n^{1-4/k}}\Big) = \mathcal{O}\Big(\frac{\log n}{n^{1-4/k}}\Big). \quad \text{(S0.19)}$$

Estimation of $\bar{\mathbb{R}}_n - \mathbb{R}_n$: First, we find an asymptotic expression of $\mathbb{R}_n$.

It is easy to show that there is a constant $c > 0$ such that $\psi_m(y; \boldsymbol{p}_0) \ge c\psi(y)$ and

$$\begin{aligned} |\hat{I}_m^{ij} - \hat{J}_m^{ij}(\boldsymbol{p}_0)| &= \left|\frac{1}{n}\sum_{l=1}^{n}\frac{\psi_{mi}(y_l)\psi_{mj}(y_l)}{\psi^2(y_l)} - \frac{1}{n}\sum_{l=1}^{n}\frac{\psi_{mi}(y_l)\psi_{mj}(y_l)}{\psi_m^2(y_l;\boldsymbol{p}_0)}\right| \\ &\le \frac{1+c}{c^2}\frac{1}{n}\sum_{l=1}^{n}\frac{\psi_{mi}(y_l)\psi_{mj}(y_l)|\psi_m(y_l;\boldsymbol{p}_0) - \psi(y_l)|}{\psi^3(y_l)} \\ &\le \frac{1+c}{c^2}\frac{1}{n}\sum_{l=1}^{n}\frac{\psi_{mi}(y_l)\psi_{mj}(y_l)}{\psi^2(y_l)}\mathcal{O}(n^{-1/2}) \\ &= \hat{I}_m^{ij}\mathcal{O}(n^{-1/2}), \end{aligned}$$

where $\hat{I}_m^{ij} = n^{-1}\sum_{l=1}^{n}\psi_{mi}(y_l)\psi_{mj}(y_l)/\psi^2(y_l)$. We have

$$\bar{R}_{ni}^{(1)} = \hat{I}_m^{ij} - \hat{J}_m^{ij}(\boldsymbol{p}_0) = \hat{I}_m^{ij}\mathcal{O}(n^{-1/2}).$$

Let $W_{ij}(Y) = \frac{\psi_{mi}(Y)\psi_{mj}(Y)}{\psi^2(Y)}$. Then $\hat{I}_m^{ij} = n^{-1}\sum_{l=1}^{n} W_{ij}(y_l)$ and $\mathcal{I}_m(f,g) = \mathrm{E}[\hat{\mathcal{I}}_m(f,g)] = [\mathrm{E}\{W_{ij}(Y)\}]$. Define

$$\varrho_{ij}^2 = \sigma^2\{W_{ij}(Y)\} = \int \frac{\psi_{mi}^2(y)\psi_{mj}^2(y)}{\psi^3(y)}dy - [I_m^{ij}(f,g)]^2.$$

By the LIL,

$$\bar{R}_{ni}^{(2)} = \hat{I}_m^{ij} - I_m^{ij} = \mathcal{O}\left(\varrho_{ij}\sqrt{\log\log n/n}\right).$$

We have

$$
\begin{aligned}
\hat{J}^{ij}_m(\boldsymbol{p}_0) &= \hat{I}^{ij}_m[1+\mathcal{O}(n^{-1/2})]\\
&= [I^{ij}_m + \mathcal{O}(\varrho_{ij}\sqrt{\log\log n/n})][1+\mathcal{O}(n^{-1/2})]\\
&= I^{ij}_m[1+\mathcal{O}(n^{-1/2})] + \mathcal{O}(\varrho_{ij}\sqrt{\log\log n/n})[1+\mathcal{O}(n^{-1/2})]\\
&= I^{ij}_m + \mathcal{O}(I^{ij}_m n^{-1/2}) + \mathcal{O}(\varrho_{ij}\sqrt{\log\log n/n}).
\end{aligned}
$$

Replacing $\widehat{\mathcal{J}}_m(\boldsymbol{p}_0)$ by $\mathcal{I}_m$ in (S0.17) we get

$$
\Pi_0\mathbb{R}_n = \Pi_0\mathcal{I}_m(\hat{\boldsymbol{p}}-\boldsymbol{p}_0) = \hat{\Pi}\tilde{\boldsymbol{R}}_n + \hat{\boldsymbol{R}}_n - \boldsymbol{\Delta}_n(\hat{\boldsymbol{p}}-\boldsymbol{p}_0) - \Pi_0\bar{\boldsymbol{R}}_n, \tag{S0.20}
$$

where $\bar{\boldsymbol{R}}_n = (\bar{R}_{n0},\ldots,\bar{R}_{nm})^{\mathrm{T}}$ and

$$
\begin{aligned}
|\bar{R}_{in}| &\le \sum_{j=0}^m [\mathcal{O}(I^{ij}_m n^{-1/2}) + \mathcal{O}(\varrho_{ij}\sqrt{\log\log n/n})]|\hat{p}_j - p_{j0}|\\
&= \mathcal{O}(\sum_{j=0}^m I^{ij}_m|\hat{p}_j - p_{j0}|n^{-1/2}) + \mathcal{O}(\sum_{j=0}^m \varrho_{ij}|\hat{p}_j - p_{j0}|\sqrt{\log\log n/n}).
\end{aligned}
$$

Thus, we have $\bar{\boldsymbol{R}}_n = \bar{\mathbb{R}}_n - \mathbb{R}_n$.

Secondly, we estimate $\bar{\boldsymbol{R}}_n$. By the Cauchy-Schwarz inequality

$$
\begin{aligned}
\sum_{j=0}^m I^{ij}_m|\hat{p}_j - p_{j0}| &\le \int \frac{\psi_{mi}(y)[|\psi_m(y;\hat{\boldsymbol{p}})| + |\psi_m(y;\boldsymbol{p}_0)|]}{\psi(y)}dy\\
&\le 2 + \int \frac{\psi_{mi}(y)[|\psi_m(y;\hat{\boldsymbol{p}}) - \psi(y)| + |\psi_m(y;\boldsymbol{p}_0) - \psi(y)|]}{\psi(y)}dy\\
&\le 2 + \left[\int \frac{\psi^2_{mi}(y)}{\psi(y)}dy \int \frac{[\psi_m(y;\hat{\boldsymbol{p}}) - \psi(y)]^2}{\psi(y)}dy\right]^{1/2}\\
&\quad + \left[\int \frac{\psi^2_{mi}(y)}{\psi(y)}dy \int \frac{[\psi_m(y;\boldsymbol{p}_0) - \psi(y)]^2}{\psi(y)}dy\right]^{1/2}\\
&= 2 + \tau_{mi}\mathcal{O}(\sqrt{\log n/n}).
\end{aligned}
$$

Because $\beta_{mi}(x) \le m+1$ and $f \ge b_0 > 0$,

$$
\psi_{mi}(u) \le (m+1)\int_0^1 g(y-x)dx \le \frac{m+1}{b_0}\int_0^1 f(x)g(y-x)dx = \frac{m+1}{b_0}\psi(y).
$$

Thus

$$
\varrho^2_{ij} < \int \frac{\psi^2_{mi}(y)\psi^2_{mj}(y)}{\psi^3(y)}dy \le \left(\frac{m+1}{b_0}\right)^2 I^{ij}_m(f,g),
$$

$$\Big[\sum_{j=0}^m \varrho_{ij}|\hat{p}_j - p_{j0}|\Big]^2 \le \sum_{j=0}^m \varrho_{ij}^2(\hat{p}_j + p_{j0}) \le \left(\frac{m+1}{b_0}\right)^2 \sum_{j=0}^m I_m^{ij}(\hat{p}_j + p_{j0})$$
$$= \mathcal{O}(m^2) + \mathcal{O}(m^2\tau_{mi}\sqrt{\log n/n}).$$

Therefore, we have

$$\bar{R}_{in} = \mathcal{O}(n^{-1/2}) + \mathcal{O}(\tau_{mi}\sqrt{\log n/n}) + \mathcal{O}\Big(m(\log n/n)^{1/4}\sqrt{\tau_{mi}\log\log n/n}\Big)$$
$$+\mathcal{O}\Big(m\sqrt{\log\log n/n}\Big),$$

and

$$\|\bar{\boldsymbol{R}}_n\|^2 = \mathcal{O}(mn^{-1}) + \mathcal{O}\Big(\sum_{i=0}^m \tau_{mi}^2 \log n/n\Big) + \mathcal{O}\Big(m^2\sum_{i=0}^m \tau_{mi}\sqrt{\log n}\log\log n/n^{3/2}\Big)$$
$$+\mathcal{O}\Big(m^3\log\log n/n\Big).$$

By (2) we have

$$\|\bar{\boldsymbol{R}}_n\|^2 = \mathcal{O}(n^{-1+1/k}) + \mathcal{O}\Big(\log n/n^{1-2/k}\Big) + \mathcal{O}\Big(\sqrt{\log n}\log\log n/n^{3/2-7/2k}\Big)$$
$$+\mathcal{O}\Big(\log\log n/n^{1-3/k}\Big) = \mathcal{O}\Big(n^{-1+3/k}\log\log n\Big).$$

Finally, combining this with (S0.19) and $\|\mathbb{R}_n\|^2 \le 2\|\bar{\mathbb{R}}_n\|^2 + 2\|\bar{\boldsymbol{R}}_n\|^2$ we get (6). For the integrated squared error, we have, a.s.,

$$(\hat{\boldsymbol{p}} - \boldsymbol{p}_0)^{\mathrm{T}}\mathcal{I}_m(1,\delta)(\hat{\boldsymbol{p}} - \boldsymbol{p}_0) = \mathbb{R}_n^{\mathrm{T}}\tilde{\Omega}_m(f,g)\mathbb{R}_n.$$

The proof is complete. □

Proof of Theorem 4.

Proof. Because the largest eigenvalue λ_m of $\Omega_m(f,g)$ is also the largest eigenvalue of $\mathcal{I}_m^{-2}(f,g)\mathcal{I}_m(f,\delta)$. Let $\boldsymbol{w}$ be the associated eigenvector satisfying $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} = 1$. Then, we have

$$\lambda_m = \frac{\boldsymbol{w}^{\mathrm{T}}\mathcal{I}_m(f,\delta)\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\mathcal{I}_m^2(f,g)\boldsymbol{w}}. \tag{S0.21}$$

From $f(x) = f_0(x) \ge b_0 > 0$, it follows that $I_m^{ij}(f,\delta) \le b_0^{-1}(m+1)$, $\forall\ i,j \in \mathbb{I}_0^m$. Thus, by the Cauchy-Schwarz inequality we have

$$\boldsymbol{w}^{\mathrm{T}}\mathcal{I}_m(f,\delta)\boldsymbol{w} \le b_0^{-1}(m+1)^2. \tag{S0.22}$$

Denote, $\boldsymbol{v} = \mathcal{I}_m(f,g)\boldsymbol{w} = (v_0,\ldots,v_m)^{\mathrm{T}}$. Then $\boldsymbol{w}^{\mathrm{T}}\mathcal{I}_m^2(f,g)\boldsymbol{w} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = \sum_{i=0}^m v_i^2 \ge (m+1)^{-1}(\sum_{i=0}^m v_i)^2 = (m+1)^{-1}(\mathbf{1}^{\mathrm{T}}\mathcal{I}_m(f,g)\boldsymbol{w})^2$. Define function of $\boldsymbol{x} = (x_0,\ldots,x_m)^{\mathrm{T}}$,

$$H(\boldsymbol{x}) = \int \frac{\psi_0(y)\psi_m(y;\boldsymbol{x})}{\psi(y)}dy,$$

where $\psi_0(y) = (1*g)(y) = \int_0^1 g(y-x)dx$ and $\psi_m(y;\boldsymbol{x}) = \sum_{i=0}^m w_i\psi_{mi}(y) = \sum_{i=0}^m x_i\int_0^1 \beta_{mi}(x)g(y-x)dx$. By binomial theorem we have

$$\boldsymbol{w}^{\mathrm{T}}\mathcal{I}_m^2(f,g)\boldsymbol{w} \ge (m+1)H^2(\boldsymbol{w}). \tag{S0.23}$$

Clearly $H^2(\boldsymbol{x})$ attains its minimum on the unit sphere at some $\boldsymbol{x}_0$ satisfying $x_{i0} = H(\boldsymbol{e}_i)/H(\boldsymbol{x}_0)$, $i \in \mathbb{I}_0^m$, where $\boldsymbol{e}_i$ denotes the vector with a 1 in the ith coordinate and 0's elsewhere. Because $H(\boldsymbol{e}_i) > 0$ for all $i \in \mathbb{I}_0^m$, we can assume that all x_{i0}'s are positive. Since g is nonvanishing, nonincreasing on $(0,\infty)$ and nondecreasing on $(-\infty,0)$, for all $x \in [0,1]$,

$$g(y-x) \ge \begin{cases} \min\{g(-1),g(1)\}, & \text{if } y \in (0,1); \\ g(y-1), & \text{if } y \le 0; \\ g(y), & \text{if } y \ge 1. \end{cases}$$

There exists a constant $C_1 > 0$ so that $\psi(y) \le C_1$ for all $y \in (-\infty,\infty)$. Hence we have, for all $i \in \mathbb{I}_0^m$, $\int \frac{\psi_0(y)\psi_{mi}(y)}{\psi(y)}dy \ge C_0/C_1$, where

$$C_0 = \min\{g^2(-1),g^2(1)\} + \int_{-\infty}^{-1} g^2(y)dy + \int_1^\infty g^2(y)dy > 0.$$

Consequently,

$$H(\boldsymbol{x}_0) \ge \frac{C_0}{C_1}\sum_{i=0}^m x_{i0} \ge \frac{C_0}{C_1}\sum_{i=0}^m x_{i0}^2 = \frac{C_0}{C_1}.$$

Combining this with (S0.21) through (S0.23) we obtain

$$\lambda_m \le \frac{m+1}{b_0H^2(\boldsymbol{x}_0)} \le \frac{C_1^2(m+1)}{b_0C_0^2} = \mathcal{O}(m).$$

Similarly $\tilde{\lambda}_m = \mathcal{O}(m)$. These combined with (4), (5), and (A.2.) ensure (7) and (8). The proof is complete. $\square$

References

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman Hall, New York, 2nd edition.

Lorentz, G. G. (1963). The degree of approximation by polynomials with positive coefficients. *Mathematische Annalen* **151**, 239–251.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* **26**, 195–239.

Wang, X.-F. and Wang, B. (2011). Deconvolution estimation in measurement error models: The R package decon. *Journal of Statistical Software* **39**, 1–24.