Supplementary Material for "Nonparametric Empirical Bayes estimation on heterogeneous data"

This supplement contains additional theoretical results (Sections A and B) and additional numerical results (Section C).

APPENDIX A: EXPRESSIONS FOR COMMON MEMBERS OF THE EXPONENTIAL FAMILY

We observe $(x_1, \theta_1), \ldots, (x_n, \theta_n)$ with conditional distribution

$$f_{\theta_i}(x_i|\eta_i) = \exp\{\eta_i z_i - \psi(\eta_i)\}\, h_{\theta_i}(z_i), \tag{A.1}$$

where θ_i is a known nuisance parameter and η_i is an unknown parameter of interest. In addition to the Gaussian distribution, there are several common cases of (A.1).

Binomial:

$$f_{n_i}(x_i|\eta_i) = \frac{n_i!}{x_i!(n_i - x_i)!} p_i^{x_i}(1-p_i)^{n_i - x_i} = \exp\{\eta_i x_i - \psi(\eta_i)\}\, h_{n_i}(x_i),$$

where $\eta_i = \log\left(\frac{p_i}{1-p_i}\right), \theta_i = n_i, \psi(\eta_i) = n_i \log(1 + e^{\eta_i})$, and $h_{n_i}(x_i) = \frac{n_i!}{x_i!(n_i-x_i)!}$.

Negative Binomial:

$$f_{r_i}(x_i|\eta_i) = \frac{(x_i + r_i - 1)!}{x_i!(r_i - 1)!} p_i^{z_i}(1-p_i)^{r_i} = \exp\{\eta_i x_i - \psi(\eta_i)\}\, h_{r_i}(x_i),$$

where $\eta_i = \log p_i, \theta_i = r_i, \psi(\eta_i) = r_i \log(1 - e^{\eta_i})$, and $h_{r_i}(x_i) = \frac{(x_i+r_i-1)!}{z_i!(r_i-1)!}$.

Gamma:

$$f_{\alpha_i}(x_i|\eta_i) = \frac{1}{\Gamma(\alpha_i)} \beta_i^{\alpha_i} x_i^{\alpha_i - 1} \exp(-\beta_i x_i) = \exp\{\eta_i x_i - \psi(\eta_i)\}\, h_{\alpha_i}(x_i),$$

where $\eta_i = -\beta_i, \theta_i = \alpha_i, \psi(\eta_i) = -\alpha_i \log(-\eta_i)$, and $h_{\alpha_i}(x_i) = \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha-1}$.

Beta:

$$f_{\alpha_i}(z_i|\eta_i) = \frac{1}{B(\alpha_i,\beta_i)} x_i^{\alpha_i}(1-x_i)^{\beta_i-1} = \exp\{\eta_i z_i - \psi(\eta_i)\}\, h_{\beta_i}(z_i),$$

where $z_i = \log x_i, \eta_i = \alpha_i, \theta_i = \beta_i, \psi(\eta_i) = \log B(\eta_i,\beta_i)$ and $h_{\beta_i}(z_i) = (1-e^{z_i})^{\beta_i-1}$.

Hence, we can compute $l'_{h,\theta}(z)$ explicitly for these distributions.

- Binomial: $-l'_{h,n_i}(x_i) = \sum_{k=1}^{x_i}\frac{1}{k} + \sum_{k=1}^{n_i-x_i}\frac{1}{k} - 2\gamma$ where γ is the Euler-Mascheroni constant
- Negative Binomial: $-l'_{h,r_i}(x_i) = \begin{cases}\sum_{k=x_i+1}^{x_i+r_i-1}\frac{1}{k} & r_i>1\\ 0 & r_i=1\end{cases}$
- Gamma: $-l'_{h,\alpha_i}(x_i) = (1-\alpha_i)\frac{1}{x_i}$
- Beta: $-l'_{h,\alpha_i}(z_i) = (\beta_i-1)\frac{e^{z_i}}{1-e^{z_i}} = (\beta_i-1)\frac{x_i}{1-x_i}$.

Combining these expressions with (2.9) we can express $E_\theta(\eta|x)$ as follows:

- Binomial: $E_{n_i}\left(\log(\frac{p_i}{1-p_i})|x_i\right) = \sum_{k=1}^{x_i}\frac{1}{k} + \sum_{k=1}^{n_i-x_i}\frac{1}{k} - 2\gamma + l'_{f,n_i}(x_i)$
- Negative Binomial: $E_{r_i}(\log p_i|x_i) = l'_{f,r_i}(x_i) + \begin{cases}\sum_{k=x_i+1}^{x_i+r_i-1}\frac{1}{k} & r_i>1\\ 0 & r_i=1\end{cases}$
- Gamma: $E_{\alpha_i}(\beta_i|x_i) = (\alpha_i-1)\frac{1}{x_i} - l'_{f,\alpha_i}(x_i)$
- Beta: $E_{\beta_i}(\alpha_i|z_i) = (\beta_i-1)\frac{x_i}{1-x_i} + l'_{f,\beta_i}(z_i)$.

APPENDIX B: PROOF OF LEMMAS 2 TO 4

B.1. Proof of Lemma 2. We first argue in Section B.1.1 that it is sufficient to prove the result over the following domain

$$\mathbb{R}_x \coloneqq \{x : C_n - \log n \le x \le C_n + \log n\}. \tag{B.1}$$

This simplification can be applied to the proofs of other lemmas.

B.1.1. *Truncating the domain.* Our goal is to show that $(\hat{\delta} - \delta^\pi)^2$ is negligible on $\mathbb{R}_x^C$. Since $|\mu| \le C_n$ by Assumption 1, the oracle estimator is bounded:

$$\delta^\pi = \mathbb{E}(X|\mu, \sigma^2) = \frac{\int \mu \phi_\sigma(x-\mu) dG_\mu(\mu)}{\int \phi_\sigma(x-\mu) dG_\mu(\mu)} < C_n.$$

Let $C_n' = C_n + \log n$. Consider the truncated NEST estimator $\hat{\delta} \wedge C_n'$. The two intermediate estimators $\tilde{\delta}$ and $\bar{\delta}$ are truncated correspondingly without altering their notations. Let $\mathbb{1}_{\mathbb{R}_x}$ be the indicator function that is 1 on $\mathbb{R}_x$ and 0 elsewhere. Our goal is to show that

$$\int\int\int_{\mathbb{R}_x^C} (\hat{\delta} - \delta^\pi)^2 \phi_\sigma(x-\mu) dx dG_\mu(\mu) dG_\sigma(\sigma) = O(n^{-\kappa}) \tag{B.2}$$

for some small $\kappa > 0$. Note that for all $x \in \mathbb{R}_x^C$, the normal tail density vanishes exponentially: $\phi_\sigma(x-\mu) = O(n^{-\epsilon'})$ for some $\epsilon' > 0$. The desired result follows from the fact that $(\hat{\delta} - \delta^\pi)^2 = o(n^\eta)$ for any $\eta > 0$, according to the assumption on C_n.

B.1.2. *Proof of the lemma.* We first apply triangle inequality to obtain

$$(\bar{\delta} - \delta^\pi)^2 \le \sigma^4 \left\{ \frac{f_\sigma^{(1)}(x)}{f_\sigma(x)} \right\}^2 \left\{ \frac{f_\sigma(x)}{\bar{f}_\sigma(x)} \right\}^2 \left[\left\{ \frac{\bar{f}_\sigma^{(1)}(x)}{f_\sigma^{(1)}(x)} - 1 \right\}^2 + \left\{ \frac{\bar{f}_\sigma(x)}{f_\sigma(x)} - 1 \right\}^2 \right]^2.$$

Hence the lemma follows if we can prove the following facts for $x \in \mathbb{R}_x$.

(i) $f_\sigma^{(1)}(x)/f_\sigma(x) = O(C_n')$, where $C_n' = C_n + \log n$.
(ii) $\bar{f}_\sigma(x)/f_\sigma(x) = 1 + O(n^{-\varepsilon})$ for some $\varepsilon > 0$.
(iii) $\bar{f}_\sigma^{(1)}(x)/f_\sigma^{(1)}(x) = 1 + O(n^{-\varepsilon})$ for some $\varepsilon > 0$.

To prove (i), note that $\delta^\pi = O(C_n)$ as shown earlier, and $x = O(C_n')$ if $x \in \mathbb{R}_x$. The oracle estimator satisfies $\delta^\pi = x + \sigma^2 f_\sigma^{(1)}(x)/f_\sigma(x)$. By Assumption 2, G_σ has a finite support, so we claim that $f_\sigma^{(1)}(x)/f_\sigma(x) = O(C_n)$.

Now consider claim (ii). Let $\mathcal{A}_\mu \coloneqq \left\{ \mu : |\mu - x| \le \sqrt{\log(n)} \right\}$. Following similar arguments to the previous sections, we apply the normal tail bounds to claim that $\phi_{\nu\bar{\sigma}}(\mu - x) = O\{n^{-1/(2\sigma^2+1)}\}$. Similar arguments apply to $f_\sigma(x)$ when $\mu \in \mathcal{A}_\mu$. Therefore

$$\frac{\bar{f}_\sigma(x)}{f_\sigma(x)} = \frac{\int_{\mu \in \mathcal{A}_\mu} \phi_{\nu\bar{\sigma}}(x-\mu) dG_\mu(\mu)}{\int_{\mu \in \mathcal{A}_\mu} \phi_\sigma(x-\mu) dG_\mu(\mu)} \left\{ 1 + O(n^{-\kappa_1}) \right\} \tag{B.3}$$

for some $\kappa_1 > 0$. Next, we evaluate the ratio in the range of $\mathcal{A}_\mu$:

$$\frac{\phi_{\nu\bar{\sigma}}(\mu - x)}{\phi_\sigma(\mu - x)} = \frac{\sigma}{(\nu\bar{\sigma})} \exp\left[-\frac{1}{2}(\mu - x)^2 \left\{\frac{1}{(\nu\bar{\sigma})^2} - \frac{1}{\sigma^2}\right\}\right] = 1 + O(n^{-\kappa_2}) \tag{B.4}$$

for some $\kappa_2 > 0$. This result follows from our definition of $\bar{\sigma}$, which is in the range of $[\sigma - L_n, \sigma + L_n]$ for some $L_n \sim n^{-\eta_l}$. Since the result (B.4) holds for all μ in $\mathcal{A}_\mu$, we have

$$\begin{aligned}
\int_{\mu\in\mathcal{A}_\mu} \phi_{\bar{\sigma}}(x-\mu)dG_\mu(\mu) &= \int_{\mu\in\mathcal{A}_\mu} \phi_\sigma(x-\mu)\frac{\phi_{\nu\bar{\sigma}}(\mu - x)}{\phi_\sigma(\mu - x)}dG_\mu(\mu) \\
&= \int_{\mu\in\mathcal{A}_\mu} \phi_{\bar{\sigma}}(x-\mu)dG_\mu(\mu)\left\{1 + O(n^{-\kappa_2})\right\}.
\end{aligned}$$

Together with (B.3), claim (ii) holds true.

To prove claim (iii), we first show that

$$\begin{aligned}
f_\sigma^{(1)}(x) &= \int \phi_\sigma(x-\mu)\frac{\mu - x}{\sigma^2}dG_\mu(\mu) \\
&= \int_{\mu\in\mathcal{A}_\mu} \phi_\sigma(x-\mu)\frac{\mu - x}{\sigma^2}dG_\mu(\mu)\left\{1 + O(n^{-\kappa_2})\right\}
\end{aligned}$$

for some $\kappa > 0$. The above claim holds true by using similar arguments for normal tails (as the term $(x - \mu)$ essentially has no impact on the rate). We can likewise argue that

$$\begin{aligned}
\bar{f}_\sigma^{(1)}(x) &= \int_{\mathcal{A}_\mu} \frac{\sigma^2}{(\nu\bar{\sigma})^2}\frac{\phi_{\nu\bar{\sigma}}(\mu - x)}{\phi_\sigma(\mu - x)}\phi_\sigma(\mu - x)\frac{\mu - x}{\sigma^2}dG_\mu(\mu) \\
&= f_\sigma^{(1)}(x)\{1 + O(n^{-\varepsilon})\}
\end{aligned}$$

for some $\epsilon > 0$. This proves (iii) and completes the proof of the lemma. Note that the proof is done without using the truncated version of $\bar{\delta}$. Since the truncation will always reduce the MSE, the result holds for the truncated $\bar{\delta}$ automatically.

B.2. Proof of Lemma 3. It is sufficient to prove the result over $\mathbb{R}_x$ defined in (B.1). Begin by defining $R_1 = \tilde{f}_\sigma^{(1)}(x) - \bar{f}_\sigma^{(1)}(x)$ and $R_2 = \tilde{f}_\sigma(x) - \bar{f}_\sigma(x)$. Then we can represent the squared difference as

$$(\tilde{\delta} - \bar{\delta})^2 = O\left(\left\{\frac{R_1}{\bar{f}_\sigma(x) + R_2}\right\}^2 + \left[\frac{R_2\bar{f}_\sigma^{(1)}(x)}{\bar{f}_\sigma(x)\{\bar{f}_\sigma(x) + R_2\}}\right]^2\right). \tag{B.5}$$

Consider L_n defined in the previous section. We first study the asymptotic behavior of R_2.

$$R_2 = \sum_{\sigma_j \in \mathcal{A}_\sigma} w_j \left\{ f_{\sigma_j}(x) - f_{\nu\bar{\sigma}}(x) \right\} + K_n(\sigma), \tag{B.6}$$

where the last term can be calculated as

$$K_n(\sigma) = \sum_{\sigma_j \in \mathcal{A}_\sigma^C} w_j \left\{ f_{\sigma_j}(x) - f_{\nu\bar{\sigma}}(x) \right\} = O\left(\sum_{\sigma_j \in \mathcal{A}_\sigma^C} w_j \right).$$

The last equation holds since both $f_{\sigma_j}(x)$ and $f_{\nu\bar{\sigma}}(x)$ are bounded according to our assumption $\sigma_l^2 \le \sigma_j^2 \le \sigma_u^2$ for all j. Consider $\mathcal{A}_\mu \coloneqq \left\{ \mu : |\mu - x| \le \sqrt{\log(n)} \right\}$. We have

$$\frac{f_{\nu\bar{\sigma}}(x)}{f_{\sigma_j}(x)} = \frac{\int_{\mu \in \mathcal{A}_\mu} \phi_{\nu\bar{\sigma}}(x-\mu) dG_\mu(\mu)}{\int_{\mu \in \mathcal{A}_\mu} \phi_{\sigma_j}(x-\mu) dG_\mu(\mu)} \left\{ 1 + O(n^{-\kappa_1}) \right\}$$

for some $\kappa_1 > 0$, and in the range of $\mu \in \mathcal{A}_\mu$, we have

$$\phi_{\nu\bar{\sigma}}(\mu - x)/\phi_\sigma(\mu - x) = 1 + O(n^{-\kappa_2})$$

for some $\kappa_2 > 0$ and all j such that $\sigma_j \in \mathcal{A}_\sigma$. We conclude that the first term in (B.6) is $O(n^{-\kappa})$ for some $\kappa > 0$ since $f_{\sigma_j}(x)$ is bounded and $\sum_{j \in \mathcal{N}_\sigma} w_j \le 1$.

Now we focus on the asymptotic behavior of $\sum_{\sigma_j \in A_\sigma^C} \omega_{\sigma_j}(\sigma)$. Let K_1 be the event that

$$n^{-1} \sum_{j=1}^{n} \phi_{h_\sigma}(\sigma_j - \sigma) < \frac{1}{2} \{ g_\sigma * \phi_{h_\sigma} \}(\sigma)$$

and K_2 the event that

$$n^{-1} \sum_{j=1}^{n} \mathbb{1}_{\{\sigma_j \in A_\sigma^C\}} \phi_{h_\sigma}(\sigma_j - \sigma) > 2 \int_{A_\sigma^C} g_\sigma(y) \phi(y - \sigma) dy.$$

Let $Y_j = \phi_{h_\sigma}(\sigma_j - \sigma)$. Then for $a_j \le Y_j \le b_j$, we use Hoeffding's inequality

$$\mathbb{P}\left(|\bar{Y} - \mathbb{E}(\bar{Y})| \ge t \right) \le 2 \exp\left\{ - \frac{2n^2 t^2}{\sum_{j=1}^n (b_i - a_i)^2} \right\}.$$

Taking $t = \frac{1}{2}\mathbb{E}(Y_i)$, we have

$$\mathbb{P}(K_1) \le 2 \exp\left\{ - \frac{(1/2) n^2 \{\mathbb{E}(Y_i)\}^2}{n \cdot O(h_\sigma^{-1})} \right\} = O(n^{-\epsilon})$$

for some $\epsilon > 0$. Similarly we can show that $\mathbb{P}(K_2) = O(n^{-\epsilon})$ for some $\epsilon > 0$. Moreover, on the event $K = K_1^C \cap K_2^C$, we have

$$\sum_{\sigma_j \in A_\sigma^C} \omega_{\sigma_j}(\sigma) \le \frac{4\int_{A_\sigma^C} g_\sigma(y)\phi(y-\sigma)dy}{\{g_\sigma * \phi_{h_\sigma}\}(\sigma)} = O(n^{-\epsilon})$$

for some $\epsilon > 0$. We use the same ϵ in the previous arguments, which can be achieved easily by appropriate adjustments (taking the smallest). Previously we have shown that the first term in (B.6) is $O(n^{-\epsilon})$. Hence on event K, $R_2 = O(n^{-\kappa})$ for some $\kappa > 0$.

Now consider the domain $\mathbb{R}_x$. Define $\mathbb{S}_x := \{x : \bar{f}_\sigma(x) > n^{-\kappa'}\}$, where $0 < \kappa' < \kappa$. On $\mathbb{R}_x \cap \mathbb{S}_x^C$, we have

$$\int\int_{\mathbb{R}_x \cap \mathbb{S}_x^C} (\tilde{\delta} - \bar{\delta})^2 f_\sigma(x) dx dG_\sigma(\sigma) = O\{C_n'^2 \cdot \mathbb{P}(\mathbb{R}_x \cap \mathbb{S}_x^C)\} = O(n^{-\kappa}) \tag{B.7}$$

for some $\kappa > 0$. The previous claim holds true since the length of $\mathbb{R}_x$ is bounded by C_n', and both $\tilde{\delta}$ and $\bar{\delta}$ are truncated by C_n'.

Now we only need to prove the result for the region $\mathbb{R}_x \cap \mathbb{S}_x$. On event K, we have

$$\begin{aligned} & \mathbb{E}_{\boldsymbol{\sigma}^2}\left(\mathbb{1}_K \cdot \int\int_{\mathbb{R}_x \cap \mathbb{S}_x} \left[\frac{R_2 \bar{f}_\sigma^{(1)}(x)}{\bar{f}_\sigma(x)\{\bar{f}_\sigma(x) + R_2\}}\right]^2 f_\sigma(x) dx dG_\sigma(\sigma)\right) \\ = \;& O(C_n'^2) O\left(n^{-(\kappa - \kappa')}\right), \end{aligned}$$

which is $O(n^{-\eta})$ for some $\eta > 0$. On event K^C,

$$\mathbb{E}_{\boldsymbol{\sigma}^2}\left(\mathbb{1}_{K^C} \cdot \int\int_{\mathbb{R}_x \cap \mathbb{S}_x} (\tilde{\delta} - \bar{\delta})^2 f_\sigma(x) dx dG_\sigma(\sigma)\right) = O(C_n'^2) O(n^{-\epsilon}),$$

which is also $O(n^{-\eta})$. Hence the risk regarding the second term of (B.5) is vanishingly small. Similarly, we can show that the first term satisfies

$$\mathbb{E}_{\boldsymbol{\sigma}^2}\left(\int\int_{\mathbb{R}_x \cap \mathbb{S}_x} \left\{\frac{R_1}{\bar{f}_\sigma(x) + R_2}\right\}^2 f_\sigma(x) dx dG_\sigma(\sigma)\right) = O(n^{-\eta}).$$

Together with (B.7), we establish the desired result.

B.3. Proof of Lemma 4. Let $S_1 = \hat{f}_\sigma^{(1)}(x) - \tilde{f}_\sigma^{(1)}(x)$ and $S_2 = \hat{f}_\sigma(x) - \tilde{f}_\sigma(x)$. Then

$$(\tilde{\delta} - \hat{\delta})^2 \leq 2\sigma^4 \left[\left\{ \frac{\tilde{f}_\sigma^{(1)}(x)}{\tilde{f}_\sigma(x)} \right\}^2 \left\{ \frac{S_2}{S_2 + \tilde{f}_\sigma(x)} \right\}^2 + \left\{ \frac{S_1}{S_2 + \tilde{f}_\sigma(x)} \right\}^2 \right]. \tag{B.8}$$

According to the definition of $\tilde{f}_\sigma(x)$ [cf. equation (6.2)], we have $\mathbb{E}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}(S_2) = 0$. By doing differentiation on both sides we further have $\mathbb{E}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}(S_1) = 0$.

A key step in our analysis is to study the variance of S_2. We aim to show that

$$\mathbb{V}_{\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma}^2}(S_2) = O(n^{-1}h_\sigma^{-1}h_x^{-1}). \tag{B.9}$$

To see this, first note that

$$\mathbb{V}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}(S_2) = \sum_{j=1}^n w_j^2 \mathbb{V}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}\{\phi_{h_{xj}}(x - X_j)\}, \text{ where}$$

$$\begin{aligned}
& \mathbb{V}\{\phi_{h_{xj}}(x - X_j)\} \\
= \;& \int \{\phi_{h_{xj}}(x-y)\}^2 \{g_\mu * \phi_{\sigma_j}\}(y)dy - \left\{ \int \phi_{h_{xj}}(x-y)\{g_\mu * \phi_{\sigma_j}\}(y)dy \right\}^2 \\
= \;& \frac{1}{h_x\sigma_j^2} \int \phi^2(z) g_\mu * \phi(x + h_x\sigma_j z)dz - \left\{ \frac{1}{\sigma_j} \int \phi(z) g_\mu * \phi_{\sigma_j}(x + h_x\sigma_j z)dz \right\}^2 \\
= \;& \frac{1}{h_x\sigma_j^2} \left\{ \int \phi^2(z)dz \right\} f_{\sigma_j}(x)\{1 + o(1)\} - \left\{ \frac{1}{\sigma_j} f_{\sigma_j}(x) \right\}^2 \{1 + o(1)\} \\
= \;& O(h_x^{-1}).
\end{aligned}$$

Next we shall show that

$$\mathbb{E}_{\boldsymbol{\sigma}^2} \left\{ \textstyle\sum_{j=1}^n w_j^2 \right\} = O\left(n^{-1}h_\sigma^{-1}\right). \tag{B.10}$$

Observe that $\phi_{h_\sigma}(\sigma_j - \sigma) = O(h_\sigma^{-1})$ for all j. Therefore we have

$$\sum_{j=1}^n \phi_{h_\sigma}^2(\sigma_j - \sigma) = O(h_\sigma^{-1}) \sum_{j=1}^n \phi_{h_\sigma}(\sigma_j - \sigma),$$

which further implies that

$$\sum_{j=1}^n w_j^2 = \frac{\sum_{j=1}^n \phi_{h_\sigma}^2(\sigma_j - \sigma)}{\left\{ \sum_{j=1}^n \phi_{h_\sigma}(\sigma_j - \sigma) \right\}^2} = \frac{O(n^{-1}h_\sigma^{-1})}{n^{-1}\sum_{j=1}^n \phi_{h_\sigma}(\sigma_j - \sigma)}.$$

Let $Y_j = \phi_{h_\sigma}(\phi_i - \phi)$ and $\bar{Y} = n^{-1}\sum_{j=1}^n Y_i$. Then $0 \leq Y_j \leq (\sqrt{2\pi}h_\sigma)^{-1}$ and

$$\mathbb{E}(Y_j) = \{g_\sigma * \phi_{h_\sigma}\}(\sigma) = g_\sigma(\sigma) + O(h_\sigma^2).$$

Let E_1 be the event such that $\bar{Y} < \frac{1}{2}E(\bar{Y})$. We apply Hoeffding's inequality to obtain

$$\begin{aligned}\mathbb{P}\left\{\bar{Y} < \frac{1}{2}E(\bar{Y})\right\} &\leq \mathbb{P}\left\{|\bar{Y} - E(\bar{Y})| \geq \frac{1}{2}E(\bar{Y})\right\} \\ &\leq 2\exp\left\{-\frac{2n^2 g_\sigma * \phi_{h_\sigma}(\sigma)}{n(2\pi)^{-1}h_\sigma^{-2}}\right\} \\ &\leq 2\exp(Cnh_\sigma^2) = O(n^{-1}).\end{aligned}$$

Note that $\sum_{j=1}^n w_j^2 \leq \sum_{j=1}^n w_j = 1$. We have

$$\begin{aligned}\mathbb{E}(\sum_{j=1}^n w_j^2) &= \mathbb{E}\left(\sum_{j=1}^n w_j^2 \mathbb{1}_E\right) + \mathbb{E}\left(\sum_{j=1}^n w_j^2 \mathbb{1}_E^C\right) \\ &= O(n^{-1}h_\sigma^{-1}) + O(n^{-1}) \\ &= O(n^{-1}h_\sigma^{-1}),\end{aligned}$$

proving (B.10). Next, consider the variance decomposition

$$\mathbb{V}_{\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma}^2}(S_2) = \mathbb{V}_{\boldsymbol{\sigma}^2}\{\mathbb{E}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}(S_2)\} + \mathbb{E}_{\boldsymbol{\sigma}^2}\{\mathbb{V}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}(S_2)\}.$$

The first term is zero, and the second term is given by

$$\mathbb{E}_{\boldsymbol{\sigma}^2}\{\mathbb{V}_{\boldsymbol{X},\boldsymbol{\mu}|\boldsymbol{\sigma}^2}(S_2)\} = O(h_x^{-1})\mathbb{E}\left(\sum_{j=1}^n w_j^2\right) = O(n^{-1}h_\sigma^{-1}h_x^{-1}).$$

We simplify the notation and denote the variance of S_2 by $\mathbb{V}(S_2)$ directly. Therefore $\mathbb{V}(S_2) = O(n^{-\epsilon})$ for some $\epsilon > 0$. Consider the following space $\mathbb{Q}_x = \{x : \tilde{f}_\sigma(x) > n^{-\epsilon'}\}$, where $2\epsilon' < \epsilon$. In the proof of the previous lemmas, we showed that on $\mathbb{R}_x$,

$$\tilde{f}_\sigma(x) = f_\sigma(x)\{1 + O(n^{-\epsilon})\} + K_n,$$

where K_n is a bounded random variable due to the variability of σ_j^2, and $E_{\boldsymbol{\sigma}^2}(K_n) = O(n^{-\varepsilon})$ for some $\varepsilon > 0$. Next we show it is sufficient to only consider $\mathbb{Q}_x$. To see this, note that

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\sigma}^2}\left(\int\int_{\mathbb{R}_x\cap\mathbb{Q}_x^C}(\tilde{\delta} - \bar{\delta})^2 f_\sigma(x)dxdG_\sigma(\sigma)\right) \\ =\ &\mathbb{E}_{\boldsymbol{\sigma}^2}\left(\int\int_{\mathbb{R}_x\cap\mathbb{Q}_x^C}(\tilde{\delta} - \bar{\delta})^2\left[\tilde{f}_\sigma(x)\{1 + O(n^{-\epsilon})\} + K_n\right]dxdG_\sigma(\sigma)\right) \\ =\ &O(C_n'^3)\left\{O(n^{-\epsilon'}) + O(n^{-\epsilon'-\epsilon}) + O(n^{-1/2})\right\},\end{aligned}$$

which is also $O(n^{-\eta})$ for some $\eta > 0$. Let

$$Y_j = w_j \phi_{h_{xj}}(x - X_j) - w_j \left\{g_\mu * \phi_{\nu\sigma_j}\right\}(x)$$

and $\bar{Y} = n^{-1}\sum_{j=1}^n Y_j$. Then $\mathbb{E}(Y_j) = 0$, $S_2 = \sum_{j=1}^n Y_j$, and $0 \le Y_j \le D_n$, where $D_n \sim h_x^{-1}$. Let E_2 be the event such that $S_2 < -\frac{1}{2}\tilde{f}_\sigma(x)$. Then by applying Hoeffding's inequality,

$$\begin{aligned}\mathbb{P}(E_2) &\le \mathbb{P}\left\{|\bar{Y} - E(\bar{Y})| \ge \frac{1}{2}\tilde{f}_\sigma(x)\right\}\\ &\le 2\exp\left\{-\frac{2n^2\{\frac{1}{2}\tilde{f}_\sigma(x)\}^2}{nD_n^2}\right\} = O(n^{-\epsilon})\end{aligned}$$

for some $\epsilon > 0$. Note that on event E_2, we have

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma}^2}\left\{(\hat{\delta} - \tilde{\delta})^2 \mathbb{1}_{E_2}\right\} = O(C_n^2)O(n^{-\epsilon}) = o(1).$$

Therefore, we only need to focus on the event E_2^C, on which we have

$$\tilde{f}_\sigma(x) + S_2 \ge \frac{1}{2}\tilde{f}_\sigma(x).$$

It follows that on E_2^C, we have

$$\{S_2/(\tilde{f}_\sigma(x) + S_2)\}^2 \le 4S_2^2/\{\tilde{f}_\sigma(x)\}^2.$$

Therefore the first term on the right of (B.8) can be controlled as

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma}^2}\left(\mathbb{1}_{E_2^C}\cdot\int\int_{\mathbb{R}_x\cap\mathbb{Q}_x}\left\{\frac{\tilde{f}_\sigma^{(1)}(x)}{\tilde{f}_\sigma(x)}\right\}^2\left\{\frac{S_2}{S_2 + \tilde{f}_\sigma(x)}\right\}^2 f_\sigma(x)dxdG_\sigma(\sigma)\right)\\ = \;& O(C_n'^2)O(n^{-(\epsilon-2\epsilon')}) = O(n^{-\eta})\end{aligned}$$

for some $\eta > 0$. Hence we show that the first term of (B.8) is vanishingly small.

For the second term in (B.8), we need to evaluate the variance term of S_1, which can be similarly shown to be of order $O(n^{-\eta})$ for some $\eta > 0$. Following similar arguments, we can prove that the expectation of the second term in (B.8) is also vanishingly small, establishing the desired result. $\square$

Table 5
Proportion of large schools. Bolded terms represent best performances.

Method	2002–2004	2003–2005	2004–2006	2005–2007	2006–2008	2007–2009	2008–2010
Naive	0.10	0.11	0.12	0.12	0.12	0.14	0.16
NEST	**0.20**	**0.19**	**0.26**	**0.26**	.29	**0.25**	**.24**
TF	0.13	0.16	0.15	0.18	0.19	0.24	0.29
Scaled	0.11	0.12	0.12	0.12	0.12	0.17	0.18
2 Group	0.13	0.15	0.15	0.18	0.18	**0.25**	0.29
3 Group	0.13	0.15	0.15	0.16	0.18	0.22	**0.26**
4 Group	0.13	0.15	0.15	0.17	0.18	**0.25**	0.27
5 Group	0.13	0.15	0.15	0.17	0.17	0.26	0.29
Group L	0.13	0.16	0.17	0.18	**0.24**	0.29	0.28
SURE-M	0.16	0.17	0.19	0.21	**0.24**	0.34	0.28
SURE-SG	0.16	0.17	0.18	0.21	0.22	0.32	**0.26**

APPENDIX C: TABLE OF LARGE SCHOOL PROPORTIONS

Table 5 shows three–year windows for the proportion of large schools selected into the 100 schools by each method. The three–year windows range from 2002 - 2004 to 2008 - 2010. NEST is among the closest to giving large schools 25% representation for 6 of 7 years.

401W Bridge Hall
Department of Data Sciences and Operations
Marshall School of Business
University of Southern California
Los Angeles, CA 90089
E-mail: luella.fu.2017@marshall.usc.edu
gareth@marshall.usc.edu
wenguans@marshall.usc.edu