SUPPLEMENT TO "A SIMPLE BOOTSTRAP METHOD FOR CONSTRUCTING NONPARAMETRIC CONFIDENCE BANDS FOR FUNCTIONS."

By Peter Hall[*,†,§], and Joel Horowitz[‡,¶]

University of Melbourne[*] and University of California, Davis[†] and Northwestern University[‡]

The supplementary material in Appendix B.1 outlines theoretical properties underpinning our methodology, while Appendix B.2 contains a proof of Theorem 4.1.

APPENDIX B: SUPPLEMENTARY MATERIAL

B.1. Outline of theoretical properties underpinning our methodology. Here we give a non-technical overview explaining why our methodology has the properties claimed (and proved theoretically) for it. We assume for simplicity that the problem is univariate (i.e. $r = 1$), and, except in the last paragraph of this appendix, we suppose that the set of points $x \in \mathcal{R}$ for which $b(x) = 0$ is of measure zero.

If the model at (2.1) obtains, and $\hat{g}$ is a local linear estimator computed using a bandwidth h and a nonnegative, symmetric, compactly supported kernel, then the bootstrap bias estimator admits the expansion

$$E\{\hat{g}^*(x) \mid \mathcal{Z}\} - \hat{g}(x) = c_1\, g''(x)\, h^2 + (nh)^{-1/2}\, f_X(x)^{-1/2}\, \sigma\, W(x/h) + \text{negligible terms}\,, \tag{B.1}$$

where, here and below, c_j denotes a positive constant, c_1 and c_2 depend only on the kernel, f_X is the common density of the design points X_i, σ^2 denotes the variance of the errors ϵ_i in the model at (2.1), and W is a c_2-dependent, stationary Gaussian process with zero mean and unit variance. (Although the covariance structure, and hence the distribution, of the process W are fixed, a different version of W is used for each sample size, to ensure that (B.1) holds.) The value of c_2, in the claim of "c_2-dependence," depends on the length of the support of the kernel used to construct the local linear estimator. The first term, $c_1\, g''(x)\, h^2$, on the right-hand side of (B.1), is identical to the asymptotic bias of $\hat{g}(x)$, but the second term on the right makes the bias estimator, on the left-hand side of (B.1), inconsistent.

Still in the context of (B.1), the bandwidth h, which would have been chosen by a standard empirical method to minimise a version of L_p error where $1 \le p < \infty$, is asymptotic to $c_3\, n^{-1/5}$, say. The asymptotic variance of $\hat{g}(x)$ equals $\sigma\,(nh)^{-1}\, f_X(x)^{-1}$, where f_X is the density of the design variables X_i. In this notation, $b(x)$ in (2.4) is given by

$$b(x) = -\lim_{h\to 0} \Big[c_1\, g''(x)\, h^2 / \{\sigma\,(nh)^{-1}\, f_X(x)^{-1}\}^{1/2} \Big] = -c_4\, g''(x)\, f_X(x)^{1/2}\,. \tag{B.2}$$

However, the limiting form of the bootstrap estimator of bias is strongly influenced by the second term on the right-hand side of (B.1), as well as by the first term, with the result that

$$\hat{\pi}(x, \alpha) = \Phi\{z + b(x) + \Delta(x)\} - \Phi\{-z + b(x) + \Delta(x)\} + \text{negligible terms}\,, \tag{B.3}$$

[§]Research supported by ARC and NSF grants.

[¶]Research supported by NSF grant SES-0817552.

uniformly in x, where, as in (2.4), $z = z_{1-(\alpha/2)} = \Phi^{-1}(1 - \frac{1}{2}\,\alpha)$, and

$$\Delta(x) = -W(x/h)\,. \tag{B.4}$$

Formula (B.3) gives us an approximation,

$$\hat\pi_{\rm app}(x,\alpha) = \Phi\{z_{1-(\alpha/2)} + b(x) + \Delta(x)\} - \Phi\{-z_{1-(\alpha/2)} + b(x) + \Delta(x)\}\,, \tag{B.5}$$

to the bootstrap estimator, $\hat\pi(x,\alpha)$, of the probability that the naive confidence interval $\mathcal{B}(\alpha)$ (see (2.2)) covers $g(x)$ at x. Thus it can be proved that $\hat\beta(x,\alpha_0)$, the solution in β of $\hat\pi(x,\beta) = 1-\alpha_0$, satisfies $\hat\beta(x,\alpha_0) = \hat\alpha_{\rm app}(x,\alpha_0) + o_p(1)$, where $\hat\alpha_{\rm app}(x,\alpha_0)$ is the solution in α of $\hat\pi_{\rm app}(x,\alpha) = 1-\alpha_0$. Note too that $\hat\alpha_{\rm app}(x,\alpha_0)$ is a smooth function, Ψ say, of both $b(x)+\Delta(x)$ and α_0:

$$\hat\alpha_{\rm app}(x,\alpha_0) = \Psi\{b(x) + \Delta(x), \alpha_0\}\,. \tag{B.6}$$

The next two paragraphs introduce a crucial property on which the performance of our methodology depends. They show that step 6 of the algorithm in Section 2.3 eliminates (in asymptotic terms) the stochastic process $\Delta(x)$ in quantities such as $\hat\pi_{\rm app}(x,\alpha)$, at (B.5), and $\hat\alpha_{\rm app}(x,\alpha_0)$, at (B.6). Now, it is the term $\Delta(x)$, which is of the same size as $b(x)$, that makes the conventional bootstrap bias estimator inconsistent. See the left-hand side of (B.1) for a definition of that estimator, and see the last paragraph of Section 2.2 for a discussion of its failings. As we note there, the failings are "caused by the stochastic error of the bias estimator" which "is of the same size as the bias itself." That is, they are caused by the size of $\Delta(x)$. By eliminating $\Delta(x)$ we remove most of the bias problem; in fact, in asymptotic terms we eliminate bias for all but a proportion ξ of points x.

To appreciate how Δ is removed, recall that W, appearing in the definition of Δ at (B.4), is a stationary, c_2-dependent Gaussian process with zero mean and unit variance, and note that the process $W(\,\cdot\,/h)$ oscillates with a frequency that diverges as $h \to 0$. This symmetry, and increasingly high frequency, ensure that if we take the median value of $\hat\alpha_{\rm app}(x_1,\alpha_0)$ over points x_1, say, in a δ-neighbourhood of x, and then let $h \to 0$, the limiting value of the median will equal $\Psi\{b(x)+\eta(x),\alpha_0\}$, where $\eta(x) \to 0$ as $\delta \to 0$. That is, the median value of $\hat\alpha_{\rm app}(x_1,\alpha_0)$, over x_1 in a small neighbourhood of the median, is very nearly $\Psi\{b(x),\alpha_0\}$. Thus, we have effectively eliminated $\Delta(x)$ from the formula at (B.6). Importantly, this elimination would not occur if we were to use the mean, rather than the median; then we would get $E[\Psi\{b(x)+Z,\alpha_0\}]$, instead of $\Psi\{b(x),\alpha_0\}$, where the random variable Z has the standard normal distribution.

We used the concept of the median here because we feel its properties may be easier for the reader to grasp than those of a quantile. However, the same property holds if (as in step 6 of the algorithm in Section 2.3) we construct the ξ-level quantile of values of $\hat\alpha_{\rm app}(x,\alpha_0)$, over values $x \in \mathcal{R}$. As $h \to 0$ this empirical quantile converges in probability to the ξ-level theoretical quantile of values of $\Psi\{b(x),\alpha_0\}$. Therefore we have again eliminated the process $\Delta(x)$ from the function $\Psi\{b(x)+\Delta(x),\alpha_0\}$ (see (B.6)), for a fraction $1-\xi$ of values of x, and so we have removed the problem of bias for those x, at least in an asymptotic sense.

Finally we summarise the remainder of the argument that shows the effectiveness of our method. Noting step 6 in Section 2.3, let x_j, for $1 \le j \le N$, say, denote the centre of the jth block in a regular rectangular grid in $\mathcal{R}$. Let $\langle u \rangle$ denote the integer part of a general positive number u, write $\hat\alpha_{{\rm app}\,\xi}(\alpha_0)$ for the $\langle \xi\, N\rangle$th largest among the values of $\hat\alpha_{\rm app}(x_j,\alpha_0)$ for $1 \le j \le N$, and let $\alpha_{{\rm app}\,\xi\,0}(\alpha_0)$ denote the version of $\hat\alpha_{\rm app}(x_j,\alpha_0)$ that is obtained if, in the formula for $\hat\alpha_{\rm app}(x,\alpha_0)$ at (B.6), we replace $\Delta(x)$ on the right-hand side by 0. (Since $\alpha_{{\rm app}\,\xi\,0}(\alpha_0)$ does not depend on data then we have not used a hat in this notation.) Then, as argued in the previous paragraph,

$$\hat\alpha_{{\rm app}\,\xi}(\alpha_0) - \alpha_{{\rm app}\,\xi\,0}(\alpha_0) \to 0 \tag{B.7}$$

in probability. Moreover, by (B.3) and (B.5), $\hat{\alpha}_\xi(\alpha_0) - \hat{\alpha}_{\mathrm{app}\,\xi}(\alpha_0) \to 0$ in probability, and so by (B.7),

$$\hat{\alpha}_\xi(\alpha_0) - \alpha_{\mathrm{app}\,\xi\,0}(\alpha_0) \to 0\,. \tag{B.8}$$

By definition, $\alpha_{\mathrm{app}\,\xi\,0}(\alpha_0)$ equals the largest value of α such that the interval $(-z_{1-(\alpha/2)}, z_{1-(\alpha/2)})$ asymptotically covers $-b(x)$ for a fraction $1-\xi$ of values of x. Therefore, in view of (B.8), the simultaneous confidence band with nominal level $\hat{\alpha}_\xi(\alpha_0)$ will also have asymptotically correct coverage for at least a fraction $1-\xi$ of values of x.

More generally, without assuming that the set of x for which $b(x) = 0$ is of measure zero, the points x that are undercovered are those for which $b(x) \neq 0$, and x will be covered at level at least $1-\alpha_0$ if and only if the nominal coverage is increased to at least the level $1-\beta$, where β solves $\Phi\{z_{1-(\beta/2)} + b(x)\} - \Phi\{-z_{1-(\beta/2)} + b(x)\} = 1-\alpha_0$. The properties discussed in Section 2.6 follow directly from this result.

B.2. Proof of Theorem 4.1. We shall prove only (4.5), since (4.6) can be derived by adapting standard results on the rate of convergence in the central limit theorems for sums of independent random variables, for example Theorem 6, page 115 of Petrov (1975). In the present context the independent random variables are the quantities ϵ_i^* multiplied by weights depending only on $\mathcal{Z}$, which is conditioned on when computing the probability on the left-hand side of (4.6).

Step 1. Preliminaries. For the sake of clarity we give the proof only in the case $r = k = 1$, where $\hat{g}$ is defined by (2.5) and (2.6). However, in step 6 below we shall mention changes that have to be made for multivariate design and polynomials of higher degree. Define $\kappa_2 = \int u^2\, K(u)\, du$ and $\kappa = \int K^2$.

Noting the model at (2.1), and defining

$$\begin{aligned}\tilde{g}(x) &= \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} A_{i_1}(x)\, A_{i_2}(X_{i_1})\, g(X_{i_2})\,,\\ e_1(x) &= \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} A_{i_1}(x)\, A_{i_2}(X_{i_1})\, \epsilon_{i_2}\,,\end{aligned} \tag{B.9}$$

where A_i as at (2.6), we have:

$$E\{\hat{g}^*(x) \mid \mathcal{Z}\} = \frac{1}{n} \sum_{i=1}^{n} A_i(x)\, \hat{g}(X_i) = \tilde{g}(x) + e_1(x)\,. \tag{B.10}$$

Writing $x_{i_1 i_2}$ for a quantity between 0 and $X_{i_2} - X_{i_1}$, and noting that $\sum_i A_i(x) \equiv n$ and $\sum_i A_i(x)\,(X_i - x) \equiv 0$, it can be shown that, for $x \in \mathcal{R}$,

$$\begin{aligned}\frac{1}{n} \sum_{i_2=1}^{n} A_{i_2}(X_{i_1})\, g(X_{i_2}) &= \frac{1}{n} \sum_{i_2=1}^{n} A_{i_2}(X_{i_1}) \Big\{ g(X_{i_1}) + (X_{i_2} - X_{i_1})\, g'(x)\\ &\qquad + \frac{1}{2}\, (X_{i_2} - X_{i_1})^2\, g''(X_{i_1} + x_{i_1 i_2}) \Big\}\\ &= g(X_{i_1}) + \frac{1}{2}\, h^2\, g''(X_{i_1}) + R(x, X_{i_1})\,,\end{aligned}$$

where

$$R(x, X_{i_1}) = \frac{1}{2n} \sum_{i_2=1}^{n} A_{i_2}(X_{i_1})\, (X_{i_2} - X_{i_1})^2 \left\{ g''(X_{i_1} + x_{i_1 i_2}) - g''(X_{i_1}) \right\}. \tag{B.11}$$

In this notation,

$$\tilde{g}(x) = \frac{1}{n} \sum_{i=1}^{n} A_i(x) \left\{ g(X_i) + \frac{1}{2} h^2 g''(X_i) + R(x, X_i) \right\}$$
$$= \hat{g}(x) + \frac{1}{2} h^2 \kappa_2 g''(x) - e_2(x) + \frac{1}{2} h^2 R(x) , \tag{B.12}$$

where

$$e_2(x) = \frac{1}{n} \sum_{i=1}^{n} A_i(x) \epsilon_i , \quad R(x) = \frac{1}{n} \sum_{i=1}^{n} A_i(x) \{ R(x, X_i) + g''(X_i) - g(x) \} . \tag{B.13}$$

Step 2. Bound for $|R(x)|$. The bound is given at (B.21) below. Let K be the kernel discussed in (4.2)(h). Since K is supported on a compact interval $[-B_1, B_1]$, for some $B_1 > 0$ (see (4.2)(h)), then $A_{i_2}(X_{i_1}) = 0$ unless $|X_{i_2} - X_{i_1}| \le 2B_1 h$, and therefore the contribution of the i_2th term to the right-hand side of (B.11) equals zero unless $|x_{i_1 i_2}| \le 2B_1 h$. However, g'' is Hölder continuous on an open set $\mathcal{O}$ containing $\mathcal{R}$ (see (4.2)(e)), and so $|g''(x_1) - g''(x_2)| \le B_2 |x_1 - x_2|^{B_3}$ for all $x_1, x_2 \in \mathcal{O}$, where $B_2, B_3 > 0$. Hence, by (B.11),

$$|R(x, X_{i_1})| \le \frac{B_4 h^{2+B_3}}{n} \sum_{i_2=1}^{n} |A_{i_2}(X_{i_1})| , \tag{B.14}$$

where $B_4 = \frac{1}{2} B_2 (2B_1)^{2+B_3}$. Now,

$$\frac{1}{n} \sum_{i=1}^{n} |A_i(x)| \le \frac{S_0(x) S_2(x) + B_1 S_0(x) |S_1(x)|}{S_0(x) S_2(x) - S_1(x)^2} . \tag{B.15}$$

We shall show in Lemma B.1, in step 9, that the open set $\mathcal{O}$ containing $\mathcal{R}$ can be chosen so that, for some $B_5 > 1$ and all $B_6 > 0$,

$$P\left\{ \max_{j=0,1,2} \sup_{x \in \mathcal{O}} |S_j(x)| > B_5 \right\} = O(n^{-B_6}) , \tag{B.16}$$

$$P\left[\min_{j=0,1,2} \inf_{x \in \mathcal{O}} \{ S_0(x) S_2(x) - S_1(x)^2 \} \le B_5^{-1} \right] = O(n^{-B_6}) . \tag{B.17}$$

Combining (B.15)–(B.17) we deduce that, for all $B_6 > 0$ and some $B_7 > 0$,

$$P\left\{ \frac{1}{n} \sum_{i=1}^{n} |A_i(x)| > B_5^2 \right\} = O(n^{-B_7}) . \tag{B.18}$$

Hence, by (B.14), for all $B_6 > 0$,

$$P\left\{ \sup_{x \in \mathcal{O}} \frac{1}{n} \sum_{i=1}^{n} |A_i(x)| \, |R(x, X_i)| > B_4 B_5^2 h^{B_3} \right\} = O(n^{-B_6}) . \tag{B.19}$$

More simply, since $A_i(x) = 0$ unless $|x - X_i| \le 2B_1 h$ then, for all $B_6 > 0$,

$$\sup_{x \in \mathcal{O}} \frac{1}{n} \sum_{i=1}^{n} |A_i(x)| \, |g''(X_i) - g''(x)| \le B_2 (2B_1 h)^{B_3} \sup_{x \in \mathcal{O}} \frac{1}{n} \sum_{i=1}^{n} |A_i(x)| ,$$

and so by (B.18), for all $B_6 > 0$,

$$P\Big\{\sup_{x\in\mathcal{O}} \frac{1}{n}\sum_{i=1}^{n} |A_i(x)|\,|g''(X_i) - g''(x)| > B_2\,(2B_1 h)^{B_3}\,B_5^2\Big\} = O(n^{-B_6})\,. \tag{B.20}$$

Combining (B.19) and (B.20), and noting the definition of $R(x)$ at (B.13), we deduce that, for all $B_6 > 0$ and some $B_7 > 0$,

$$P\Big\{\sup_{x\in\mathcal{O}} |R(x)| > B_7\,h^{2+B_3}\Big\} = O(n^{-B_6})\,. \tag{B.21}$$

Step 3. Expansion of $e_1(x)$. Recall that $e_1(x)$ was defined at (B.9); our expansion of $e_1(x)$ is given at (B.31) below, and the terms R_1 and R_2 in (B.31) satisfy (B.30) and (B.33), respectively. The expansion is designed to replace h, the bandwidth in (4.2)(g), which potentially depends on the errors ϵ_i as well as on the design variables X_i, by a deterministic bandwidth h_1. If h were a function of the design sequence alone then this step would not be necessary.

Define $h_1 = C_1\,n^{-1/(r+4k)} = C_1\,n^{-1/5}$ where C_1 is as in (4.2)(g), put $\delta_1 = (h_1 - h)/h_1$, and note that, if $|\delta_1| \le \frac{1}{2}$,

$$\frac{h_1}{h} = (1-\delta_1)^{-1} = 1 + \delta_1 + \frac{1}{2}\,\delta_1^2 + \frac{1}{3}\,\delta_1^3 + \dots\,.$$

Therefore, if $\ell \ge 1$ is an integer, and if K has $\ell + 1$ uniformly bounded derivatives, then there exist constants $B_8, B_9 > 0$ such that, when $|\delta_1| \le \frac{1}{2}$,

$$\bigg|K\Big(\frac{u}{h}\Big) - \bigg\{K\Big(\frac{u}{h_1}\Big) + \sum_{j_1}\sum_{j_2} c(j_1,j_2)\,\delta_1^{j_1+j_2}\Big(\frac{u}{h_1}\Big)^{j_2} K^{(j_2)}\Big(\frac{u}{h_1}\Big)\bigg\}\bigg| \le B_8\,|\delta_1|^{\ell+1}\,I(|u/h_1| \le B_9)\,, \tag{B.22}$$

where $c(j_1,j_2)$ denotes a constant and the double summation is over j_1 and j_2 such that $j_1 \ge 0$, $j_2 \ge 1$ and $j_1 + j_2 \le \ell$. (This range of summation is assumed also in the double summations in (B.24) and (B.28) below.) The constant B_9 is chosen so that $K(u)$ and its derivatives vanish for $|u| > B_9$. More simply,

$$\frac{u}{h} = \frac{u}{h_1}\Big(1 + \delta_1 + \frac{1}{2}\,\delta_1^2 + \frac{1}{3}\,\delta_1^3 + \dots\Big)\,. \tag{B.23}$$

Recall that $S_k(x) = n^{-1}\sum_i \{(x-X_i)/h\}^k\,K\{(x-X_i)/h\}$. Write this as $S_k(h,x)$, to indicate the dependence on h, and define

$$T_{kj}(x) = \frac{1}{nh_1}\sum_{i=1}^{n}\Big(\frac{x-X_i}{h_1}\Big)^{k+j} K^{(j)}\Big(\frac{x-X_i}{h_1}\Big)\,.$$

Results (B.22) and (B.23) imply that, for constants $c_k(j_1,j_2)$, and provided $|\delta_1| \le \frac{1}{2}$,

$$\bigg|S_k(h,x) - \bigg\{S_k(h_1,x) + \sum_{j_1}\sum_{j_2} c_k(j_1,j_2)\,\delta_1^{j_1+j_2}\,T_{kj_2}(x)\bigg\}\bigg| \le B_{10}\,\frac{|\delta_1|^{\ell+1}}{nh_1}\sum_{i=1}^{n} I\Big(\Big|\frac{x-X_i}{h_1}\Big| \le B_9\Big)\,. \tag{B.24}$$

The methods leading to (B.60), in the proof of Lemma B.1, can be used to show that there exists an open set $\mathcal{O}$, containing $\mathcal{R}$, such that for all $B_{11}, B_{12} > 0$, and each j and k,

$$P\Big\{\sup_{x\in\mathcal{O}} |(1-E)\,T_{kj}(x)| > (nh_1)^{-1/2}\,n^{B_{11}}\Big\} = O(n^{-B_{12}})\,. \tag{B.25}$$

Additionally, the argument leading to (B.61) can be used to prove that

$$\sup_{x\in\mathcal{O}} |E\{T_{kj}(x)\} - \ell_{kj}(x)| \to 0\,, \tag{B.26}$$

where $\ell_{kj}(x) = f_X(x) \int u^{k+j}\, K^{(j)}(u)\, du$.

The definition of $e_1(x)$, at (B.9), can be written equivalently as

$$\begin{aligned} e_1(x) = \frac{1}{(nh)^2} \sum_{i_1=1}^n \sum_{i_2=1}^n & \frac{S_2(h,x) - \{(x - X_{i_1})/h\}\, S_1(h,x)}{S_0(h,x)\, S_2(h,x) - S_1(h,x)^2} \\ & \times \frac{S_2(h,X_{i_1}) - \{(X_{i_1} - X_{i_2})/h\}\, S_1(h,X_{i_1})}{S_0(h,X_{i_1})\, S_2(h,X_{i_1}) - S_1(h,X_{i_1})^2} \\ & \times K\Big(\frac{x - X_{i_1}}{h}\Big)\, K\Big(\frac{X_{i_1} - X_{i_2}}{h}\Big)\, \epsilon_{i_2}\,. \end{aligned} \tag{B.27}$$

Write $A_i(h_1,x)$ for the version of $A_i(x)$, at (2.6), that would be obtained if h were replaced by h_1 in that formula, and in particular in the definitions of S_0, S_1, S_2 and K_i. Using (B.22) and (B.24) to substitute for $K(u/h)$ and $S_k(h,x)$, respectively, where $u = x - X_{i_1}$ or $X_{i_1} - X_{i_2}$ in the case of $K(u/h)$; and then Taylor expanding; it can be proved from (B.27), noting the properties at (B.24)–(B.26), that, provided $|\delta_1| \le \frac{1}{2}$,

$$\begin{aligned} e_1(x) = \frac{1}{n^2} & \sum_{i_1=1}^n \sum_{i_2=1}^n A_{i_1}(h_1,x)\, A_{i_2}(h_1,X_{i_1})\, \epsilon_{i_2} \\ & + \sum_{j_1} \sum_{j_2} c_1(j_1,j_2)\, \delta_1^{j_1+j_2}\, D_{j_1 j_2}(x) + |\delta_1|^{\ell+1}\, R_1(x)\,, \end{aligned} \tag{B.28}$$

where the constants $c_1(j_1,j_2)$ do not depend on h_1 or n and are uniformly bounded; each term $D_{j_1 j_2}(x)$ can be represented as

$$\frac{1}{nh} \sum_{i=1}^n J_i\Big(\frac{x - X_i}{h_1}\Big)\, \epsilon_i\,, \tag{B.29}$$

where the functions J_i depend on j_1, j_2 and x, on the design sequence $\mathcal{X}$ and the bandwidth h_1, but not on the errors or on h, and, for some B_{14} and B_{15}, and all B_{16}, satisfy

$$P\Big\{\sup_{x\in\mathcal{O}} \sup_{|u|\le B_{13}} \max_{1\le i\le n} |J_i(u)| > B_{14}\Big\} = O(n^{-B_{16}})\,,$$

$$P\Big\{\sup_{x\in\mathcal{O}} \sup_{|u|> B_{14}} \max_{1\le i\le n} |J_i(u)| = 0\Big\} = O(n^{-B_{16}})\,,$$

$$P\Big\{\sup_{x\in\mathcal{O}} \sup_{u_1,u_2\in\mathbb{R}} \max_{1\le i\le n} |J_i(u_1) - J_i(u_2)| \le B_{15}\, |u_1 - u_2|\Big\} = O(n^{-B_{16}})\,;$$

and, also for some B_{14} and all B_{16},

$$P\Big\{\sup_{x\in\mathcal{O}} |R_1(x)| > B_{14}\Big\} = O(n^{-B_{16}})\,. \tag{B.30}$$

Combining the results from (B.28) down, and using Markov's inequality and lattice arguments to bound the quantity at (B.29), we deduce that, if $|\delta_1| \le \frac{1}{2}$,

$$e_1(x) = T(x) + \delta_1\, R_2(x) + |\delta_1|^{\ell+1}\, R_1(x)\,, \tag{B.31}$$

where

$$T(x) = \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} A_{i_1}(h_1, x)\, A_{i_2}(h_1, X_{i_1})\, \epsilon_{i_2}\,, \tag{B.32}$$

R_1 satisfies (B.30) and R_2 satisfies

$$P\Big\{ \sup_{x\in\mathcal{O}} |R_2(x)| > (nh_1)^{-1/2}\, n^{B_{17}} \Big\} = O(n^{-B_{18}})\,. \tag{B.33}$$

In (B.33), for each fixed $B_{17} > 0$, B_{18} can be taken arbitrarily large, provided that $E|\epsilon|^{B_{19}} < \infty$ for sufficiently large B_{19}.

Step 4. Approximation to $T(x)$, defined at (B.32). The approximation is given by (B.41), with the remainder there controlled by (B.42).

Define

$$D_i(x) = \frac{1}{n} \sum_{i_1=1}^{n} A_{i_1}(h_1, x)\, A_i(h_1, X_{i_1})\,.$$

Result (B.61), derived during the proof of Lemma B.1, and assumption (4.2)(f) on f_X, imply that for a sequence of constants $\eta = \eta(n)$ decreasing to 0 at a polynomial rate as $n \to \infty$, and for all $B_{16} > 0$,

$$P\bigg[\sup_{x\in\mathcal{O}} \max_{1\le i\le n} \bigg\{ \bigg| h\, A_i(h_1, x)\, K\Big(\frac{x - X_i}{h_1}\Big)^{-1} - \frac{1}{f_X(x)} \bigg| \bigg\} > \eta \bigg] = O(n^{-B_{16}})\,. \tag{B.34}$$

Define

$$D(x_1, x_2) = \frac{1}{nh_1} \sum_{i=1}^{n} \frac{1}{f_X(X_i)}\, K\Big(\frac{x_1 - X_i}{h_1}\Big)\, K\Big(\frac{x_2 - X_i}{h_1}\Big)\,,$$

It follows from (B.34) and the compact support of K (see (4.2)(h)) that

$$D_i(x) = \frac{1 + \Delta_i(x)}{h_1\, f_X(x)}\, D(x, X_i)\,, \tag{B.35}$$

where the random functions Δ_i are measurable in the sigma-field generated by $\mathcal{X}$ (we refer to this below as "$\mathcal{X}$ measurable") and satisfy, for some $B_{20} > 0$ and all $B_{16} > 0$,

$$P\Big\{ \sup_{x\in\mathcal{O}} \max_{1\le i\le n} |\Delta_i(x)| > n^{-B_{20}} \Big\} = O(n^{-B_{16}})\,. \tag{B.36}$$

Recall that $L = K * K$, and note that $E\{D(x_1, x_2)\} = L\{(x_1 - x_2)h_1\}$, and that, since K is compactly supported, there exists $B_{21} > 0$ such that $D(x_1, x_2) = 0$ whenever $x_1 \in \mathcal{R}$ and $|x_1 - x_2| > B_{21}$. Furthermore, there exist B_{22} and $B_{23}(p)$, the latter for each choice of the integer $p \ge 1$, such that whenever $x_1 \in \mathcal{R}$,

$$\mathrm{var}\{D(x_1, x_2)\} \le \frac{B_{22}}{nh_1}\, I\Big(\Big| \frac{x_1 - x_2}{h_1} \Big| \le B_{21} \Big)\,,$$

$$E\Big\{ \frac{1}{f_X(X_i)}\, K\Big(\frac{x_1 - X_i}{h_1}\Big)\, K\Big(\frac{x_2 - X_i}{h_1}\Big) \Big\}^{2p} \le B_{23}(p)\, h_1\, I\Big(\Big| \frac{x_1 - x_2}{h_1} \Big| \le B_{21} \Big)\,.$$

Hence, by Rosenthal's inequality, whenever $x_1 \in \mathcal{R}$ and $x_2 \in \mathbb{R}$,

$$E\big|(1 - E)\, D(x_1, x_2)\big|^{2p} \le \frac{B_{24}(p)}{(nh_1)^p}\, I\Big(\Big| \frac{x_1 - x_2}{h_1} \Big| \le B_{21} \Big)\,.$$

Therefore, by Markov's inequality, for each $B_{25}, B_{26} > 0$,

$$\sup_{x_1 \in \mathcal{R},\, x_2 \in \mathbb{R}} P\Big\{\big|(1-E)\, D(x_1,x_2)\big| > (nh_1)^{-1/2}\, n^{B_{25}}\Big\} = O\big(n^{-B_{26}}\big)\,.$$

Approximating to $(1-E)\,D(x_1,x_2)$ on a polynomially fine lattice of pairs (x_1,x_2), with $x_1 \in \mathcal{R}$ and $|x_1 - x_2| \le B_{21}$, we deduce that the supremum here can be placed inside the probability statement: for each $B_{25}, B_{26} > 0$,

$$P\Big\{\sup_{x_1 \in \mathcal{R},\, x_2 \in \mathbb{R}} \big|(1-E)\, D(x_1,x_2)\big| > (nh_1)^{-1/2}\, n^{B_{25}}\Big\} = O\big(n^{-B_{26}}\big)\,. \tag{B.37}$$

Combining (B.35)–(B.37) we deduce that

$$D_i(x) = \frac{1+\Delta_i(x)}{h_1\, f_X(x)}\, L\Big(\frac{x-X_i}{h_1}\Big) + \frac{\Theta_i(x)}{h_1\,(nh_1)^{1/2}}\, I\Big(\Big|\frac{x-X_i}{h_1}\Big| \le B_{21}\Big)\,, \tag{B.38}$$

where the function Θ_i is $\mathcal{X}$-measurable and satisfies, for each B_{25}, $B_{26} > 0$,

$$P\Big\{\sup_{x\in\mathcal{R}} \max_{i\,:\,|x-X_i|\le B_{21}h_1} |\Theta_i(x)| > n^{B_{25}}\Big\} = O\big(n^{-B_{16}}\big)\,. \tag{B.39}$$

By (B.32) and (B.38),

$$T(x) = \frac{1}{n}\sum_{i=1}^n D_i(x)\,\epsilon_i = \frac{1}{nh_1\, f_X(x)} \sum_{i=1}^n L\Big(\frac{x-X_i}{h_1}\Big)\,\epsilon_i + T_1(x) + T_2(x)\,, \tag{B.40}$$

where

$$T_1(x) = \frac{1}{nh_1\, f_X(x)} \sum_{i=1}^n \Delta_i(x)\, L\Big(\frac{x-X_i}{h_1}\Big)\,\epsilon_i\,,$$
$$T_2(x) = \frac{1}{(nh_1)^{3/2}\, f_X(x)} \sum_{i=1}^n \Theta_i(x)\, I\Big(\Big|\frac{x-X_i}{h_1}\Big| \le B_{21}\Big)\,\epsilon_i\,.$$

Using the fact that the functions Δ_i and Θ_i are $\mathcal{X}$-measurable, as well as (B.36), (B.39), and approximations on polynomially fine lattices, it can be proved that if B_{27} (large) and B_{28} (small) are given then, provided $E|\epsilon|^{B_{29}} < \infty$ where B_{29} depends on B_{27} and B_{28}, we have for some $B_{30} > 0$,

$$P\Big\{\sup_{x\in\mathcal{R}} |T_1(x)| > (nh_1)^{-1/2}\, n^{-B_{30}}\Big\} = O\big(n^{-B_{27}}\big)\,,$$
$$P\Big\{\sup_{x\in\mathcal{R}} |T_2(x)| > (nh_1)^{-1}\, n^{B_{28}}\Big\} = O\big(n^{-B_{27}}\big)\,.$$

Therefore, by (B.40),

$$T(x) = \frac{1}{n}\sum_{i=1}^n D_i(x)\,\epsilon_i = \frac{1}{nh_1\, f_X(x)} \sum_{i=1}^n L\Big(\frac{x-X_i}{h_1}\Big)\,\epsilon_i + R_3(x)\,, \tag{B.41}$$

where the following property holds for $j=3$: for some $B_{31} > 0$, if $B_{32} > 0$ is given then, provided $E|\epsilon|^{B_{33}} < \infty$, with B_{33} depending on B_{32},

$$P\Big\{\sup_{x\in\mathcal{R}} |R_j(x)| > (nh_1)^{-1/2}\, n^{-B_{31}}\Big\} = O\big(n^{-B_{32}}\big)\,. \tag{B.42}$$

Step 5. Approximation to $e_2(x)$, defined at (B.13). Property (B.61), and arguments similar to those in step 5, permit us to show that

$$e_2(x) = \frac{1}{nh_1\, f_X(x)} \sum_{i=1}^{n} L\Big(\frac{x - X_i}{h_1}\Big)\, \epsilon_i + R_4(x)\,, \tag{B.43}$$

where (B.42) holds for $j = 4$.

Step 6. Gaussian approximation to $e_1(x) - e_2(x)$. The approximation is given at (B.50). Recall that $M = L - K = K * K - K$. By (B.31), (B.33) and (B.41)–(B.43),

$$e_1(x) - e_2(x) = \frac{1}{nh_1\, f_X(x)} \sum_{i=1}^{n} M\Big(\frac{x - X_i}{h_1}\Big)\, \epsilon_i + |\delta_1|^{\ell+1}\, R_1(x) + R_5(x)\,, \tag{B.44}$$

where R_1 and R_5 satisfy (B.30) and (B.42), respectively.

Next we use an approximation due to Komlós, Major and Tusnády (1976). Theorem 4 there implies that if B_{34} (small) and B_{35} (large) are given then there exists $B_{36} > 0$, depending on B_{34} and B_{35}, such that, if $E|\epsilon|^{B_{36}} < \infty$, then it is possible to construct a sequence of Normal random variables $Z_1, Z_2, \ldots$ with $E(Z_i) = E(\epsilon_i) = 0$ and $E(Z_i^2) = E(\epsilon_i)^2 = \sigma^2$, and for which

$$P\bigg\{ \max_{1\le i\le n} \bigg| \sum_{i_1=1}^{i} (\epsilon_{i_1} - Z_{i_1}) \bigg| > n^{B_{34}} \bigg\} = O\big(n^{-B_{35}}\big)\,. \tag{B.45}$$

Define $M_i(x) = M\{(x - X_i)/h_1\}$ for $1 \le i \le n$, $M_{n+1} = 0$, $V_i = \sum_{1\le i_1\le i} \epsilon_{i_1}$ and $N_i = \sum_{1\le i_1\le i} Z_{i_1}$, and note that, using Euler's method of summation,

$$\sum_{i=1}^{n} M\Big(\frac{x - X_i}{h_1}\Big)\, \epsilon_i = \sum_{i=1}^{n} M_i(x)\, \epsilon_i = \sum_{i=1}^{n} \{M_i(x) - M_{i+1}(x)\}\, V_i\,.$$

Therefore,

$$\sum_{i=1}^{n} M\Big(\frac{x - X_i}{h_1}\Big)\, (\epsilon_i - Z_i) = \sum_{i=1}^{n} \{M_i(x) - M_{i+1}(x)\}\, (V_i - N_i)\,. \tag{B.46}$$

Let $\mathcal{T} = \mathcal{T}(h_1)$ denote the set of all points $x_1 \in \mathbb{R}$ such that $(x - x_1)/h_1$ lies within the support of K for some $x \in \mathcal{R}$. Then $\mathcal{T}$ depends on n, and, for $n \ge B_{37}$ say, is a subset of the open set $\mathcal{O}$ introduced in step 9. Hence, if $x \in \mathcal{T}$ and $n \ge B_{37}$ then $f_X(x) > B_{38}$, where $B_{38} > 0$ is a lower bound for f_X on the open set referred to in (4.2)(f). Let ν denote the number of X_is, for $1 \le i \le n$, that lie in $\mathcal{T}$. Order the X_is so that these X_is are listed first in the sequence $X_1, \ldots, X_n$, and moreover, such that $X_1 \le \ldots \le X_\nu$. Let $X_{\nu+1}$ be the X_i that is nearest to X_ν and is not one of $X_1, \ldots, X_\nu$. Using properties of spacings of order statistics from a distribution the density of which is bounded away from zero, we deduce that if $B_{39} < 1$ then, for all $B_{40} > 0$,

$$P\Big(\max_{1\le i\le \nu} |X_i - X_{i+1}| > n^{-B_{39}} \Big) = O\big(n^{-B_{40}}\big)\,. \tag{B.47}$$

If $1 \le i \le \nu$ and $x \in \mathcal{R}$ then

$$\begin{aligned} |M_i(x) - M_{i+1}(x)| &= \Big| M\Big(\frac{x - X_i}{h_1}\Big) - M\Big(\frac{x - X_i}{h_1} + \frac{X_i - X_{i+1}}{h_1}\Big) \Big| \\ &\le h_1^{-1}\, (\sup |M'|)\, |X_i - X_{i+1}| \le h_1^{-1}\, (\sup |M'|)\, n^{-B_{39}}\,, \end{aligned}$$

where the identity and the first inequality hold with probability 1, and, by (B.47), the second inequality holds with probability $1 - O(n^{-B_{40}})$. If $n \geq B_{37}$ and $x \in \mathcal{R}$ then all the indices i for which $M_i(x) - M_{i+1}(x) \neq 0$ are in the range from 1 to ν, and therefore the second series on the right-hand side of (B.46) can be restricted to a sum from $i = 1, \ldots, \nu$. Combining the results in this paragraph we deduce that for all $B_{40} > 0$,

$$P\Big\{ \max_{1 \leq i \leq n} |M_i(x) - M_{i+1}(x)| \leq h_1^{-1} (\sup |M'|) \, n^{-B_{39}} \Big\} = 1 - O(n^{-B_{40}}) \,. \tag{B.48}$$

In multivariate cases, where the number of dimensions, r, satisfies $r \geq 2$, the spacings argument above should be modified by producing an ordering $X_1, \ldots, X_n$ of the X_is which is such that $\|X_i - X_{i+1}\|$ is small, for $1 \leq i \leq n-1$, where $\|\cdot\|$ is the Euclidean metric. We do this by taking $B < 1/r$ and, first of all, constructing a regular, rectangular lattice within $\mathcal{R}$ where the total number of cells, or lattice blocks, is bounded above and below by constant multiples of $n^{B+\delta}$ for a given $\delta \in (0, \frac{1}{3}(r^{-1} - B))$. (The sizes of the faces of the cells are in proportion to the sizes of the faces of $\mathcal{R}$.) We order the points X_i within each given cell so that $\|X_i - X_{i+1}\| \leq n^B$ is small. (With probability converging to 1 at a polynomial rate, this can be done simultaneously for each of the cells.) Then we choose one representative point X_i in each cell (it could be the point nearest to the cell's centre), and draw a path linking that point in one cell to its counterpart in an adjacent cell, such that those linked points are no further than $n^{B+2\delta}$ apart, and each cell is included in the chain after just $n-1$ links have been drawn. Again this can be achieved with probability converging to 1 at a polynomial rate. Once the linkage has been put in place, the n design points can be reordered so that $\|X_i - X_{i+1}\| \leq n^{B+3\delta}$ for $1 \leq i \leq n-1$. By taking $B > r$, but very close to r, and then choosing $\delta > 0$ but very close to 0, we see that, for any given $B' > r$, we can, with probability converging to 1 at a polynomial rate, construct an ordering $X_1, \ldots, X_n$ so that $\|X_i - X_{i+1}\| \leq n^{B'}$ for $1 \leq i \leq n-1$.

Result (B.46) implies that, if $x \in \mathcal{R}$,

$$\begin{aligned} &\bigg|\frac{1}{nh_1 \, f_X(x)} \sum_{i=1}^n M\Big(\frac{x - X_i}{h_1}\Big) (\epsilon_i - Z_i)\bigg| \\ &\qquad \leq \frac{1}{nh_1 B_{38}} \Big\{ \max_{1 \leq i \leq n} |M_i(x) - M_{i+1}(x)| \Big\} \Big\{ \max_{1 \leq i \leq n} \bigg| \sum_{i_1=1}^{i} (\epsilon_{i_1} - Z_{i_1}) \bigg| \Big\} \,. \end{aligned} \tag{B.49}$$

Combining (B.45), (B.48) and (B.49), recalling from step 3 that $h_1 = C_1 \, n^{-1/5}$, and taking $B_{35} > 1$ in (B.45), we conclude that for all $B_{40} > 0$,

$$\begin{aligned} P\bigg\{ &\sup_{x \in \mathcal{R}} \bigg|\frac{1}{nh_1 \, f_X(x)} \sum_{i=1}^n M\Big(\frac{x - X_i}{h_1}\Big) (\epsilon_i - Z_i)\bigg| \\ &\qquad\qquad \leq \frac{\sup |M'|}{B_{38} \, C_1^2 \, n^{(3/5) + B_{39} - B_{34}}} \bigg\} = 1 - O(n^{-B_{40}}) \,. \end{aligned}$$

Hence, by (B.44),

$$e_1(x) - e_2(x) = \zeta(x) + |\delta_1|^{\ell+1} \, R_1(x) + R_5(x) + R_6(x) \,, \tag{B.50}$$

where

$$\zeta(x) = \frac{1}{nh_1 \, f_X(x)} \sum_{i=1}^n M\Big(\frac{x - X_i}{h_1}\Big) \, Z_i \,, \tag{B.51}$$

R_1 and R_5 satisfy (B.30) and (B.42), respectively, and, for some $B_{41} > 0$ and all $B_{40} > 0$,

$$P\Big\{ \sup_{x\in\mathcal{R}} |R_6(x)| \le B_{41}\, n^{-(3/5)-B_{39}+B_{34}} \Big\} = 1 - O\big(n^{-B_{40}}\big)\,. \tag{B.52}$$

Step 7. Approximation to $\zeta(x)$ in terms of a Gaussian process. The approximation is given at (B.53). Conditional on the design sequence $\mathcal{X}$ the process ζ, at (B.51), is itself Gaussian, with zero mean and covariance

$$\mathrm{cov}\{\zeta(x_1), \zeta(x_2)\,|\,\mathcal{X}\} = \frac{\sigma^2}{\{nh_1\, f_X(x)\}^2} \sum_{i=1}^{n} M\Big(\frac{x_1 - X_i}{h_1}\Big)\, M\Big(\frac{x_2 - X_i}{h_1}\Big)\,,$$

and standard arguments show that for some $B_{42} > 0$ and all $B_{43} > 0$,

$$P\Big\{ \sup_{x_1,\, x_2\in\mathcal{R}} \Big| nh_1\, f_X(x)\, \mathrm{cov}\{\zeta(x_1), \zeta(x_2)\,|\,\mathcal{X}\} - (M * M)\Big(\frac{x_1 - x_2}{h_1}\Big)\Big| > n^{-B_{42}} \Big\} = O\big(n^{-B_{43}}\big)\,.$$

Hence, for each n there exists a Gaussian stationary process W, with zero mean and covariance given by (4.4), such that for some $B_{44} > 0$ and all $B_{43} > 0$,

$$P\Big\{ (nh_1)^{1/2} \sup_{x\in\mathcal{R}} |\zeta(x) - f_X(x)^{-1/2}\, W(x)| > n^{-B_{44}} \Big\} = O\big(n^{-B_{43}}\big)\,. \tag{B.53}$$

Step 8. Completion of proof of Theorem 4.1, except for Lemma B.1. Combining (B.10) and (B.12) we deduce that

$$E\{\hat g^*(x)\,|\,\mathcal{Z}\} = \hat g(x) + \frac{1}{2}\, h^2\, \kappa_2\, g''(x) + e_1(x) - e_2(x) + \frac{1}{2}\, h^2\, R(x)\,. \tag{B.54}$$

Combining (B.50) and (B.53), using the bounds at (B.30), (B.42) and (B.52) on the remainder terms R_1, R_5 and R_6 on the right-hand side of (B.50), and noting that, in view of (4.2)(g) and the definition $\delta_1 = (h_1 - h)/h_1$, $P(|\delta_1| > n^{-C_2/(r+4k)}) \to 0$, we see that if the exponent $\ell + 1$ in (B.50) can be taken sufficiently large (depending on $C_2 > 0$, and enabled by taking C_5 sufficiently large in (4.2)(h)), then for some $B_{45} > 0$,

$$P\Big\{ (nh_1)^{1/2} \sup_{x\in\mathcal{R}} |e_1(x) - e_2(x) - f_X(x)^{-1/2}\, W(x)| > n^{-B_{45}} \Big\} \to 0\,. \tag{B.55}$$

In view of the approximation to h by $h_1 = C_1\, n^{-1/5}$ asserted in (4.2)(g), (B.21) implies that

$$P\Big\{ \sup_{x\in\mathcal{R}} |R(x)| > B_7\, h_1^2\, n^{-B_{46}} \Big\} \to 0\,. \tag{B.56}$$

Result (4.5) follows on combining (B.54)–(B.56).

Step 9. Derivation of (B.16) and (B.17).

LEMMA B.1. *If (4.2) holds then there exists an open set $\mathcal{O}$, containing $\mathcal{R}$, such that (B.16) and (B.17) obtain.*

To derive the lemma, recall that

$$S_k(x) = \frac{1}{nh} \sum_{i=1}^{n} \Big(\frac{x - X_i}{h}\Big)^k K\Big(\frac{x - X_i}{h}\Big) .$$

Let $\mathcal{H} = [n^{-B_{49}}, n^{-B_{48}}]$, where $0 < B_{48} < B_{49} < 1$. As noted in (4.2), the bandwidth h is a function of the data in $\mathcal{Z}$, but initially we take h to be deterministic, denoting it by h_2 and denoting the corresponding value of $S_k(x)$ by $S_k(h_2, x)$. Then by standard calculations, for each integer $j \geq 1$, and for $k = 0, 1, 2$,

$$\sup_{h_2 \in \mathcal{H}} \sup_{x \in \mathbb{R}} (nh_2)^j \, E\big[\{(1 - E)\, S_k(h_2, x)\}^{2j}\big] \leq B(j) , \tag{B.57}$$

where $B(j)$ does not depend on n. Here we have used the uniform boundedness of f_X, asserted in (4.2)(f). Result (B.57), and Markov's inequality, imply that for all $B_{50}, B_{51} > 0$, and for $k = 0, 1, 2$,

$$\sup_{h_2 \in \mathcal{H}} \sup_{x \in \mathbb{R}} P\Big[\big|(1 - E)\, S_k(h_2, x)\big| > (nh_2)^{-1/2}\, n^{B_{50}}\Big] = O\big(n^{-B_{51}}\big) . \tag{B.58}$$

Let $B_{52} > 0$. It follows from (B.58) that if $\mathcal{S}$, contained in the open set referred to in (4.2)(f), is a compact subset of $\mathbb{R}$, if $\mathcal{S}(n)$ is any subset of $\mathcal{S}$ such that $\#\mathcal{S}(n) = O(n^{B_{52}})$, and if $\mathcal{H}(n)$ is any subset of $\mathcal{H}$ such that $\#\mathcal{H}(n) = O(n^{B_{52}})$, then for all $B_{50}, B_{51} > 0$, and for $k = 0, 1, 2$,

$$P\Big[\sup_{h_2 \in \mathcal{H}(n)} \sup_{x \in \mathcal{S}(n)} \big|(1 - E)\, S_k(h_2, x)\big| > (nh_2)^{-1/2}\, n^{B_{50}}\Big] = O\big(n^{-B_{51}}\big) . \tag{B.59}$$

Approximating to $S_k(h_2, x)$ on a polynomially fine lattice of values of h_2 and x, we deduce from (B.59) that, for all $B_{50}, B_{51} > 0$, and for $k = 0, 1, 2$,

$$P\Big[\sup_{h_2 \in \mathcal{H}} \sup_{x \in \mathcal{S}} \big|(1 - E)\, S_k(h_2, x)\big| > (nh_2)^{-1/2}\, n^{B_{50}}\Big] = O\big(n^{-B_{51}}\big) . \tag{B.60}$$

Choose $\mathcal{S}$ sufficiently large to contain an open set, $\mathcal{O}$, which contains $\mathcal{R}$ and has the property that, for some $\delta > 0$, the set of all closed balls of radius δ and centred at a point in $\mathcal{O}$ is contained in $\mathcal{S}$. Since K is compactly supported and $h_2 \in \mathcal{H}$ satisfies $h_2 \leq n^{-B_{48}}$, it can be proved from (4.2)(f) that from some $B_A > 0$,

$$\sup_{h_2 \in \mathcal{H}} \sup_{x \in \mathcal{O}} \Big[\big|E\{S_0(x)\} - f_X(x)\big| + \big|E\{S_1(x)\}\big| + \big|E\{S_2(x)\} - \kappa_2\, f_X(x)\big|\Big] = O\big(n^{-B_A}\big) .$$

Therefore, defining $\ell_k(x) = f_X(x), 0, \kappa_2\, f_X(x)$ according as $k = 0, 1, 2$, respectively, we deduce from (B.60) that for a sequence $\eta = \eta(n)$ decreasing to 0 at a polynomial rate in n as $n \to \infty$, for $k = 0, 1, 2$, and for all $B_{51} > 0$,

$$P\Big[\sup_{h_2 \in \mathcal{H}} \sup_{x \in \mathcal{O}} \big|S_k(h_2, x) - \ell_k(x)\big| > \eta\Big] = O\big(n^{-B_{51}}\big) . \tag{B.61}$$

Results (B.16) and (B.17) follow from (B.61) on noting the properties on h in (4.2)(g); we can take C_3 and C_4 there to be B_{48} and B_{49} above.

REFERENCES

PETROV, V.V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

Department of Mathematics and Statistics
The University of Melbourne
VIC 3010, Australia
and
Department of Statistics
University of California
Davis, CA 95616, USA.
E-mail: halpstat@ms.unimelb.edu.au

Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, Illinois 60208, USA.
E-mail: joel-horowitz@northwestern.edu