SUPPLEMENTARY MATERIAL TO 'EFFICIENT MULTIVARIATE ENTROPY ESTIMATION VIA *K*-NEAREST NEIGHBOUR DISTANCES'

By Thomas B. Berrett[*,§], Richard J. Samworth[†,§] and Ming Yuan[‡,¶]

University of Cambridge[§]
University of Wisconsin–Madison[¶]

Appendix. This is the supplementary material to Berrett, Samworth and Yuan (2017), hereafter referred to as the main text.

A.1. *Proofs of auxiliary results.*

Proof of Proposition 9 in the main text. Fix $\tau \in \bigl(\frac{d}{\alpha+d}, 1\bigr]$. We first claim that given any $\epsilon > 0$, there exists $A_\epsilon > 0$ such that $a(\delta) \leq A_\epsilon \delta^{-\epsilon}$ for all $\delta \in (0,\gamma]$. To see this, observe that there exists $\delta_0 \in (0,\gamma]$ such that $a(\delta) \leq \delta^{-\epsilon}$ for $\delta \leq \delta_0$. But then

$$\sup_{\delta \in (0,\gamma]} \delta^\epsilon a(\delta) \leq \max\bigl\{1, \gamma^\epsilon a(\delta_0)\bigr\} \leq \gamma^\epsilon \delta_0^{-\epsilon},$$

which establishes the claim, with $A_\epsilon := \gamma^\epsilon \delta_0^{-\epsilon}$. Now choose $\epsilon = \frac{1}{3}\bigl(\tau - \frac{d}{\alpha+d}\bigr)$ and let $\tau' := \frac{\tau}{3} + \frac{2d}{3(\alpha+d)} \in \bigl(\frac{d}{\alpha+d}, 1\bigr)$. Then, by Hölder's inequality, and since $\alpha\tau'/(1-\tau') > d$,

$$\begin{aligned}\sup_{f \in \mathcal{F}_{d,\theta}} \int_{\{x : f(x) < \delta\}} a\bigl(f(x)\bigr) f(x)^\tau \, dx &\leq A_\epsilon \delta^\epsilon \sup_{f \in \mathcal{F}_{d,\theta}} \int_{\{x : f(x) < \delta\}} f(x)^{\tau'} \, dx \\ &\leq A_\epsilon \delta^\epsilon (1+\nu)^{\tau'} \biggl\{ \int_{\mathbb{R}^d} (1 + \|x\|^\alpha)^{-\frac{\tau'}{1-\tau'}} \, dx \biggr\}^{1-\tau'} \to 0\end{aligned}$$

as $\delta \searrow 0$, as required.

---

[*]Research supported by a Ph.D. scholarship from the SIMS fund.

[†]Research supported by an EPSRC Early Career Fellowship and a grant from the Leverhulme Trust.

[‡]Research supported by NSF FRG Grant DMS-1265202 and NIH Grant 1-U54AI117924-01.

AMS 2000 subject classifications: 62G05, 62G20

Keywords and phrases: efficiency, entropy estimation, Kozachenko–Leonenko estimator, weighted nearest neighbours

For the second part, fix $\rho > 0$, set $\epsilon := \frac{1}{2}\bigl(\tau - \frac{d}{\alpha+d}\bigr)$ and $\tau' := \frac{\tau}{2} + \frac{d}{2(\alpha+d)} \in \bigl(\frac{d}{\alpha+d}, 1\bigr)$. Then, by Hölder's inequality again,

$$\sup_{f\in\mathcal{F}_{d,\theta}} \int_{\mathcal{X}} a\bigl(f(x)\bigr)^{\rho} f(x)^{\tau}\,dx \leq A_{\epsilon/\rho} \sup_{f\in\mathcal{F}_{d,\theta}} \int_{\mathcal{X}} f(x)^{\tau'}\,dx$$
$$\leq A_{\epsilon/\rho}(1+\nu)^{\tau'}\biggl\{\int_{\mathbb{R}^d} (1+\|x\|^{\alpha})^{-\frac{\tau'}{1-\tau'}}\,dx\biggr\}^{1-\tau'} < \infty,$$

as required. □

Proof of Lemma 10. (i) The lower bound is immediate from the fact that $h_x(r) \leq V_d\|f\|_\infty r^d$ for any $r > 0$. For the upper bound, observe that by Markov's inequality, for any $r > 0$,

$$h_x(\|x\|+r) = \int_{B_x(\|x\|+r)} f(y)\,dy \geq \int_{B_0(r)} f(y)\,dy \geq 1 - \frac{\mu_\alpha(f)}{r^\alpha}.$$

The result follows on substituting $r = \bigl(\frac{\mu_\alpha(f)}{1-s}\bigr)^{1/\alpha}$ for $s \in (0,1)$.

(ii) We first prove this result in the case $\beta \in (2,4]$, giving the stated form of $b_1(\cdot)$. Let $C := 4dV_d^{-\beta/d}/(d+\beta)$, and let $y := Ca(f(x))^{\beta/2}s\{s/f(x)\}^{\beta/d}$. Now, by the mean value theorem, we have for $r \leq r_a(x)$ that

$$\biggl|h_x(r) - V_d r^d f(x) - \frac{V_d}{2(d+2)} r^{d+2}\Delta f(x)\biggr| \leq a(f(x))f(x)\frac{dV_d}{2(d+\beta)}r^{d+\beta}.$$

It is convenient to write

$$s_{x,y} := s - \frac{s^{1+2/d}\Delta f(x)}{2(d+2)V_d^{2/d} f(x)^{1+2/d}} + y.$$

Then, provided $s_{x,y} \in (0, V_d r_a^d(x) f(x)]$, we have

$$h_x\biggl(\frac{s_{x,y}^{1/d}}{\{V_d f(x)\}^{1/d}}\biggr)$$
$$\geq s_{x,y} + \frac{V_d^{-2/d}\Delta f(x)}{2(d+2)f(x)^{1+2/d}} s_{x,y}^{1+2/d} - \frac{a(f(x))dV_d^{-\beta/d}}{2(d+\beta)f(x)^{\beta/d}} s_{x,y}^{1+\beta/d}.$$

Now, by our hypothesis, we know that

$$\sup_{f\in\mathcal{F}_{d,\theta}} \sup_{s\in\mathcal{S}_n} \sup_{x\in\mathcal{X}_n} \max\biggl\{\frac{V_d^{-2/d}s^{2/d}|\Delta f(x)|}{2(d+2)f(x)^{1+2/d}}, \frac{y}{s}\biggr\}$$
$$\leq \max\biggl\{\frac{d^{1/2}V_d^{-2/d}C_n^{2/d}}{2(d+2)}, CC_n^{\beta/d}\biggr\} \to 0$$

as $n \to \infty$. Hence there exists $n_1 = n_1(d,\theta) \in \mathbb{N}$ such that for all $n \geq n_1$, all $f \in \mathcal{F}_{d,\theta}$, $s \in \mathcal{S}_n$ and $x \in \mathcal{X}_n$, we have

$$\frac{1}{2(d+2)}(s_{x,y}^{1+2/d} - s^{1+2/d}) \geq -\frac{s^{1+2/d}}{2d}\biggl\{\frac{d^{1/2}V_d^{-2/d}a(f(x))s^{2/d}}{2(d+2)f(x)^{2/d}} + \frac{y}{s}\biggr\}.$$

Moreover, there exists $n_2 = n_2(d,\theta) \in \mathbb{N}$ such that for all $n \geq n_2$, all $s \in \mathcal{S}_n$, $x \in \mathcal{X}_n$ and $f \in \mathcal{F}_{d,\theta}$ we have

$$|s_{x,y}|^{1+\beta/d} \leq 2s^{1+\beta/d}.$$

Finally, we can choose $n_3 = n_3(d,\theta) \in \mathbb{N}$ such that

$$\max\biggl\{\frac{C_n^{(4-\beta)/d}}{4(d+2)V_d^{(4-\beta)/d}}, \frac{2d^{1/2}C_n^{2/d}}{(d+\beta)V_d^{2/d}}, \frac{d^{3/2}C_n^{2/d}}{2(d+2)(d+\beta)V_d^{2/d}}\biggr\} \leq \frac{d}{d+\beta}$$

and such that $C_n \leq (8d^{1/2})^{-d}V_d/2$ for $n \geq n_3$. It follows that for $n \geq \max(n_1,n_2,n_3) =: n_*$, for $f \in \mathcal{F}_{d,\theta}$, $s \in \mathcal{S}_n$ and for $x \in \mathcal{X}_n$, we have that $s_{x,y} \in (0, V_d r_a^d(x)f(x)]$ and

$$\begin{aligned}
&h_x\biggl(\frac{s_{x,y}^{1/d}}{\{V_d f(x)\}^{1/d}}\biggr) - s \\
&\geq y - \frac{a(f(x))s^{1+2/d}}{2d^{1/2}V_d^{2/d}f(x)^{2/d}}\biggl\{\frac{d^{1/2}V_d^{-2/d}a(f(x))s^{2/d}}{2(d+2)f(x)^{2/d}} + \frac{y}{s}\biggr\} - \frac{da(f(x))s^{1+\beta/d}}{(d+\beta)V_d^{\frac{\beta}{d}}f(x)^{\frac{\beta}{d}}} \\
&\geq \frac{a(f(x))^{\beta/2}s^{1+\beta/d}}{f(x)^{\beta/d}}\biggl[C - \frac{a(f(x))^{2-\beta/2}}{4(d+2)V_d^{4/d}}\biggl\{\frac{s}{f(x)}\biggr\}^{(4-\beta)/d} \\
&\qquad - \frac{Ca(f(x))}{2d^{1/2}V_d^{2/d}}\biggl\{\frac{s}{f(x)}\biggr\}^{2/d} - \frac{dV_d^{-\beta/d}}{d+\beta}\biggr] \geq 0.
\end{aligned}$$

The lower bound is proved by very similar calculations, and the result for the case $\beta \in (2,4]$ follows. The general case can be proved using very similar arguments, and is omitted for brevity. □

A.2. *Auxiliary results for the proof of Theorem 2 in the main text.* Recall the definition of $V(f)$ given in the statement of Theorem 1.

Lemma 1. *For each $d \in \mathbb{N}$ and $\theta \in \Theta$ and $m \in \mathbb{N}$, we have*

(i) $\sup_{f \in \mathcal{F}_{d,\theta}} \int_{\mathcal{X}} f(x)|\log^m f(x)|\,dx < \infty$;
(ii) $\inf_{f \in \mathcal{F}_{d,\theta}} V(f) > 0$;

Proof of Lemma 1. Fix $d \in \mathbb{N}$ and $\theta = (\alpha, \beta, \gamma, \nu, a) \in \Theta$.

(i) For $\epsilon \in (0,1)$ and $t \in (0,1]$, we have

$$\log \frac{1}{t} \leq \frac{1}{\epsilon} t^{-\epsilon}.$$

Let $\epsilon = \frac{\alpha}{m(\alpha+2d)}$, so that $\frac{\alpha(1-m\epsilon)}{m\epsilon} = 2d$. Then, by Hölder's inequality, for any $f \in \mathcal{F}_{d,\theta}$,

$$\begin{aligned}\int_{\mathcal{X}} f(x) |\log^m f(x)| \, dx &\leq 2^{m-1} \int_{\mathcal{X}} f(x) \log^m \Bigl(\frac{\|f\|_\infty}{f(x)} \Bigr) \, dx + 2^{m-1} |\log^m \|f\|_\infty| \\ &\leq \frac{2^{m-1} \|f\|_\infty^{m\epsilon}}{\epsilon^m} \int_{\mathcal{X}} f(x)^{1-m\epsilon} \, dx + 2^{m-1} |\log^m \|f\|_\infty| \\ &\leq \frac{2^{m-1} \gamma^{m\epsilon}}{\epsilon^m} (1+\nu)^{1-m\epsilon} \Bigl\{ \int_{\mathcal{X}} (1 + \|x\|^\alpha)^{-\frac{1-m\epsilon}{m\epsilon}} \, dx \Bigr\}^{m\epsilon} \\ &+ 2^{m-1} \max \Bigl\{ \log^m \gamma \, , \, \frac{1}{\alpha^m} \log^m \Bigl(\frac{V_d^\alpha \nu^d (\alpha+d)^{\alpha+d}}{\alpha^\alpha d^d} \Bigr) \Bigr\},\end{aligned}$$

where the bound on $\bigl|\log^m \|f\|_\infty\bigr|$ comes from (13) in the main text.

(ii) Now define

$$A_{d,\theta} := \max \Bigl\{ \sup_{f \in \mathcal{F}_{d,\theta}} |H(f)| \, , \, -\frac{1}{2} \log \inf_{f \in \mathcal{F}_{d,\theta}} \|f\|_\infty \, , \, 1 \Bigr\}$$

and the set $S_{d,\theta} := \{x \in \mathcal{X} : e^{-4A_{d,\theta}} \leq f(x) \leq e^{-2A_{d,\theta}}\}$. For $f \in \mathcal{F}_{d,\theta}, x \in S_{d,\theta}$ and $y \in B_x(\{8d^{1/2} a(e^{-4A_{d,\theta}})\}^{-1/(\beta \wedge 1)})$ we have by Lemma 2 below that

$$|f(y) - f(x)| \leq \frac{15 d^{1/2}}{7} a(e^{-4A_{d,\theta}}) e^{-2A_{d,\theta}} \|y - x\|^{\beta \wedge 1}. \tag{1}$$

By the continuity of f, there exists $x_0 \in S_{d,\theta}$ such that $f(x_0) = \frac{1}{2} e^{-2A_{d,\theta}}(1 + e^{-2A_{d,\theta}})$. Thus, by (1), we have that $B_{x_0}(r_{d,\theta}) \subseteq S_{d,\theta}$, where

$$r_{d,\theta} := \Bigl\{ \frac{7(1 - e^{-2A_{d,\theta}})}{30 d^{1/2} a(e^{-4A_{d,\theta}})} \Bigr\}^{1/(\beta \wedge 1)} \wedge \frac{1}{\{8 d^{1/2} a(e^{-4A_{d,\theta}})\}^{1/(\beta \wedge 1)}}.$$

Hence

$$V(f) = \mathbb{E}_f[\{\log f(X_1) + H(f)\}^2] \geq A_{d,\theta}^2 \mathbb{P}_f(X_1 \in S_{d,\theta}) \geq A_{d,\theta}^2 e^{-4A_{d,\theta}} V_d r_{d,\theta}^d,$$

as required. □

The following auxiliary result provides control on deviations of the density arising from the smoothness condition of our $\mathcal{F}_{d,\theta}$ classes.

Lemma 2. *For $\theta = (\alpha, \beta, \gamma, \nu, a) \in \Theta$, $m := \lceil \beta \rceil - 1$, $f \in \mathcal{F}_{d,\theta}$ and $y \in B_x\bigl(r_a(x)\bigr)$, we have, for multi-indices t with $|t| \leq m$, that*

$$\Bigl|\frac{\partial f^t}{\partial x^t}(y) - \frac{\partial f^t}{\partial x^t}(x)\Bigr| \leq \frac{15d^{1/2}}{7} a\bigl(f(x)\bigr) f(x) \|y-x\|^{\min(\beta - |t|, 1)}.$$

Proof. If $|t| = m$ then the result follows immediately from the definition of $\mathcal{F}_{d,\theta}$. Henceforth, therefore, assume that $m \geq 1$ and $|t| \leq m-1$. Writing $|\!|\!|\cdot|\!|\!|$ here for the largest absolute entry of an array, we have for $y \in B_x\bigl(r_a(x)\bigr)$ that

$$\begin{aligned}
&\Bigl|\frac{\partial f^t}{\partial x^t}(y) - \frac{\partial f^t}{\partial x^t}(x)\Bigr| \leq \|y-x\| \sup_{z \in B_x(\|y-x\|)} \Bigl\|\nabla \frac{\partial f^t}{\partial x^t}(z)\Bigr\| \\
&\leq \|y-x\| \|f^{(|t|+1)}(x)\| + d^{1/2}\|y-x\| \sup_{z \in B_x(\|y-x\|)} \Bigl|\!\Bigl|\!\Bigl| f^{(|t|+1)}(z) - f^{(|t|+1)}(x) \Bigr|\!\Bigr|\!\Bigr| \\
&\leq \sum_{\ell=1}^{m-|t|} d^{(\ell-1)/2} \|y-x\|^{\ell} \|f^{(|t|+\ell)}(x)\| \\
&\qquad + d^{(m-|t|)/2}\|y-x\|^{m-|t|} \sup_{z \in B_x(\|y-x\|)} \Bigl|\!\Bigl|\!\Bigl| f^{(m)}(z) - f^{(m)}(x) \Bigr|\!\Bigr|\!\Bigr| \\
&\leq a(f(x)) f(x) \|y-x\| \biggl\{ \frac{1}{1 - d^{1/2}\|y-x\|} + d^{(\beta-|t|)/2}\|y-x\|^{\beta-|t|-1} \biggr\} \\
&\leq \frac{15d^{1/2}}{7} a(f(x)) f(x) \|y-x\|,
\end{aligned}$$

as required. ◻

Lemma 3. *Under the conditions of Theorem 1 in the main text, we have that*

$$\sup_{k \in \{k_0^*, \ldots, k_1^*\}} \sup_{f \in \mathcal{F}_{d,\theta}} \mathbb{E}_f[\{\tilde{V}_n^w - V(f)\}^2] \to 0.$$

Proof. For $w = (w_1, \ldots, w_k)^T \in \mathcal{W}^{(k)}$, write $\mathrm{supp}(w) := \{j : w_j \neq 0\}$. Then

$$\begin{aligned}
&|\mathbb{E}_f \tilde{V}_n^w - V(f)| \\
&\leq \biggl|\sum_{j=1}^k w_j \mathbb{E}_f \log^2 \xi_{(j),1} - \int_{\mathcal{X}} f \log^2 f\biggr| + \bigl|\mathbb{E}_f\{(\hat{H}_n^w)^2\} - H(f)^2\bigr| \\
&\leq \|w\|_1 \max_{j \in \mathrm{supp}(w)} \biggl|\mathbb{E}_f \log^2 \xi_{(j),1} - \int_{\mathcal{X}} f \log^2 f\biggr| + \mathrm{Var}_f \hat{H}_n^w + |(\mathbb{E}_f \hat{H}_n^w)^2 - H(f)^2|.
\end{aligned}$$

Thus, by Theorem 1 in the main text, (18) in the proof of that result and Lemma 1(i), we have that $\sup_{k\in\{k_0^*,\ldots,k_1^*\}}\sup_{f\in\mathcal{F}_{d,\theta}}|\mathbb{E}_f\tilde{V}_n^w - V(f)| \to 0$. Now,

$$\begin{aligned}\mathrm{Var}_f\tilde{V}_n^w \leq \frac{\|w\|_1^2}{n}\max_{j\in\mathrm{supp}(w)}\mathrm{Var}_f\log^2\xi_{(j),1}& \\ + \|w\|_1^2\max_{j,\ell\in\mathrm{supp}(w)}\bigl|\mathrm{Cov}_f(\log^2\xi_{(j),1},\log^2\xi_{(\ell),2})\bigr|.&\end{aligned} \tag{2}$$

Let $a_{n,j}^- := (j-3j^{1/2}\log^{1/2}n)\vee 0$ and $a_{n,j}^+ := (j+3j^{1/2}\log^{1/2}n)\wedge(n-1)$. Mimicking arguments in the proof of Theorem 1, for any $m\in\mathbb{N}$, $j\in\mathrm{supp}(w)$ and $\epsilon>0$,

$$\begin{aligned}\mathbb{E}_f\bigl\{\log^m(\xi_{(j),1}f(X_1))\bigr\}&\\ &= \int_{\mathcal{X}} f(x)\int_0^\infty \log^m\biggl(\frac{V_d(n-1)f(x)h_x^{-1}(s)^d}{e^{\Psi(j)}}\biggr)\mathrm{B}_{j,n-j}(s)\,ds\,dx\\ &= \int_{\frac{a_{n,j}^-}{n-1}}^{\frac{a_{n,j}^+}{n-1}}\log^m\biggl(\frac{(n-1)s}{e^{\Psi(j)}}\biggr)\mathrm{B}_{j,n-j}(s)\,ds\\ &\qquad + O\biggl(\max\biggl\{\frac{k^{\beta/d}}{n^{\beta/d}}\log^{m-1}n\,,\,\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}\biggr\}\biggr)\to 0,\end{aligned}$$

uniformly for $j\in\mathrm{supp}(w)$, $k\in\{k_0^*,\ldots,k_1^*\}$ and $f\in\mathcal{F}_{d,\theta}$. Moreover, by Cauchy–Schwarz, we can now show, for example, that

$$\mathbb{E}_f\log^4\xi_{(j),1} = \mathbb{E}_f[\{\log(\xi_{(j),1}f(X_1)) - \log f(X_1)\}^4] \to \mathbb{E}_f\log^4 f(X_1)$$

uniformly for $j\in\mathrm{supp}(w)$, $k\in\{k_0^*,\ldots,k_1^*\}$ and $f\in\mathcal{F}_{d,\theta}$. The result follows upon noting that we may use a similar approach for the covariance term in (2) to see that $\sup_{k\in\{k_0^*,\ldots,k_1^*\}}\sup_{f\in\mathcal{F}_{d,\theta}}\mathrm{Var}\,\tilde{V}_n^w\to 0$. □

A.3. *Proof of Proposition 5 in the main text.*

Proof of Proposition 5 in the main text. In each of the three examples, we provide $\theta=(\alpha,\beta,\gamma,\nu,a)\in\Theta$ such that $f\in\mathcal{F}_{d,\theta}$. In fact, $\beta>0$ may be chosen arbitrarily in each case.

(i) We may choose any $\alpha>0$, and then set $\nu=2^{\alpha/2}\Gamma\bigl(\frac{\alpha}{2}+\frac{d}{2}\bigr)/\Gamma(d/2)$. We may also set $\gamma=(2\pi)^{-d/2}$. It remains to find $a\in\mathcal{A}$ such that (6) in the main text holds. Write $H_r(y) := (-1)^r e^{y^2/2}\frac{d^r}{dy^r}e^{-y^2/2}$ for the rth Hermite polynomial, and note that $|H_r(y)|\leq p_r(|y|)$, where p_r is a polynomial of degree r with non-negative coefficients. Using multi-index notation for partial derivatives, if $t = (t_1, \ldots, t_d) \in \{0, 1, \ldots, \}^d$ with $|t| := t_1 + \ldots + t_d$, we have

$$\left|\frac{\partial f^t(x)}{\partial x^t}\right| = f(x) \prod_{j=1}^d |H_{t_j}(x_j)| \leq f(x) \prod_{j=1}^d p_{t_j}(\|x\|) \leq f(x) q_{|t|}(\|x\|),$$

for some polynomial q_r of degree r, with non-negative coefficients. It follows that if $y \in B_x^\circ(1)$, then for any $\beta > 0$ with $m = \lceil \beta \rceil - 1$,

$$\begin{aligned}
\frac{\|f^{(m)}(x) - f^{(m)}(y)\|}{f(x)\|y-x\|^{\beta-m}} &\leq \frac{d^{m/2}}{f(x)\|y-x\|^{\beta-m}} \max_{t:|t|=m} \left| \frac{\partial f^t(x)}{\partial x^t} - \frac{\partial f^t(y)}{\partial x^t} \right| \\
&\leq \frac{d^{(m+1)/2}}{f(x)} \max_{t:|t|=m+1} \sup_{w \in B_0(1)} \left| \frac{\partial f^t(x+w)}{\partial x^t} \right| \\
&\leq d^{(m+1)/2} \sup_{w \in B_0(1)} \frac{f(x+w) q_{m+1}(\|x+w\|)}{f(x)} \\
&\leq d^{(m+1)/2} e^{\|x\|} q_{m+1}(\|x\|+1).
\end{aligned}$$

Similarly,

$$\max_{r=1,\ldots,m} \frac{\|f^{(r)}(x)\|}{f(x)} \leq d^{m/2} \max_{r=1,\ldots,m} q_r(\|x\|).$$

Write $g(\delta) := \bigl\{-2\log\bigl(\delta(2\pi)^{d/2}\bigr)\bigr\}^{1/2}$ and define $a \in \mathcal{A}$ by setting $a(\delta) := \max\{1, \tilde{a}(\delta)\}$, where

$$\begin{aligned}
\tilde{a}(\delta) &:= d^{m/2} \sup_{x:\|x\| \leq g(\delta)} \max\biggl\{ \max_{r=1,\ldots,m} q_r(\|x\|) \, , \, d^{1/2} e^{\|x\|} q_{m+1}(\|x\|+1) \biggr\} \\
&= d^{m/2} \max\biggl\{ \max_{r=1,\ldots,m} q_r\bigl(g(\delta)\bigr) \, , \, d^{1/2} e^{g(\delta)} q_{m+1}\bigl(g(\delta)+1\bigr) \biggr\}.
\end{aligned}$$

Then $\sup_{x:f(x)\geq\delta} M_{f,a,\beta}(x) \leq a(\delta)$ and $a(\delta) = o(\delta^{-\epsilon})$ for every $\epsilon > 0$, so (6) in the main text holds.

(ii) We may choose any $\alpha < \rho$, and set

$$\nu = 2^{\alpha/2} \frac{\Gamma\bigl(\frac{\alpha}{2} + \frac{d}{2}\bigr)}{\Gamma\bigl(\frac{d}{2}\bigr)} \frac{(\rho/2)^{\alpha/2} \Gamma\bigl(\frac{\rho-\alpha}{2}\bigr)}{\Gamma\bigl(\frac{\rho}{2}\bigr)}.$$

We may also set $\gamma = \frac{\Gamma\left(\frac{\rho}{2}+\frac{d}{2}\right)}{\Gamma(\rho/2)\rho^{d/2}\pi^{d/2}}$. To verify (6) in the main text for suitable $a \in \mathcal{A}$, we note by induction, that if $t = (t_1, \ldots, t_d) \in \{0, 1, \ldots, \}^d$ with $|t| := t_1 + \ldots + t_d$, then

$$\left|\frac{\partial f^t(x)}{\partial x^t}\right| \leq \frac{f(x) q_{|t|}(\|x\|)}{(1 + \|x\|^2/\rho)^{|t|}},$$

where q_r is a polynomial of degree r with non-negative coefficients. Thus, similarly to the Gaussian example, for any $\beta > 0$ with $m = \lceil \beta \rceil - 1$,

$$\sup_{x \in \mathbb{R}^d} \sup_{y \in B_x^\circ(1)} \frac{\|f^{(m)}(x) - f^{(m)}(y)\|}{f(x)\|y-x\|^{\beta-m}} \leq d^{(m+1)/2} \sup_{x \in \mathbb{R}^d} \sup_{w \in B_0(1)} \frac{f(x+w)q_{m+1}(\|x+w\|)}{f(x)(1+\|x\|^2/\rho)^{m+1}} =: A^{(1)}_{d,m,\rho},$$

say, where $A^{(1)}_{d,m,\rho} \in [0,\infty)$. Similarly,

$$\sup_{x \in \mathbb{R}^d} \max_{r=1,\ldots,m} \frac{\|f^{(r)}(x)\|}{f(x)} \leq d^{m/2} \sup_{x \in \mathbb{R}^d} \max_{r=1,\ldots,m} \frac{q_r(\|x\|)}{(1+\|x\|^2/\rho)^r} =: A^{(2)}_{d,m,\rho},$$

say, where $A^{(2)}_{d,m,\rho} \in [0,\infty)$. Now defining $a \in \mathcal{A}$ to be the constant function

$$a(\delta) := \max\{1, A^{(1)}_{d,m,\rho}, A^{(2)}_{d,m,\rho}\},$$

we again have that $\sup_{x:f(x)\geq\delta} M_{f,a,\beta}(x) \leq a(\delta)$, so (6) in the main text holds.

(iii) We may take any $\alpha > 0$ and $\nu = 1$, $\gamma = 3$. To verify (6) in the main text, fix $\beta > 0$, set $m := \lceil \beta \rceil - 1$, and define $a \in \mathcal{A}$ by

$$a(\delta) := A_m \max\biggl\{1\,,\, \log^{2(m+1)}\Bigl(\frac{1}{\delta}\Bigr)\biggr\},$$

for some $A_m \geq 1$ depending only on m. Then, by induction, we find that for some constants $A'_m, B'_m > 0$ depending only on m, and $x \in (-1,1)$

$$M_{f,a,\beta}(x) \leq \max\biggl\{\max_{r=1,\ldots,m} \frac{A'_r}{(1-x^2)^{2r}}\,,\, \sup_{y:0<|y-x|\leq r_a(x)} \frac{A'_{m+1}f(y)}{(1-y^2)^{2(m+1)}f(x)}\biggr\} \leq \frac{B'_{m+1}}{(1-x^2)^{2(m+1)}} \leq a\bigl(f(x)\bigr),$$

provided A_m in the definition of a is chosen sufficiently large. Hence (6) in the main text again holds. □

A.4. *Proof of Proposition 6 in the main text.*

PROOF OF PROPOSITION 6 IN THE MAIN TEXT. To deal with the integrals over $\mathcal{X}_n^c$, we first observe that by (13) in the main text there exists a constant $C_{d,f} > 0$, depending only on d and f, such that

$$\int_{\mathcal{X}_n^c} f(x) \int_0^1 \mathrm{B}_{k,n-k}(s) \log u_{x,s} \, ds \, dx$$

(3)

$$\leq C_{d,f} \int_{\mathcal{X}_n^c} f(x) \biggl\{ \log n + \log\biggl(1 + \frac{\|x\|}{\mu_\alpha^{1/\alpha}(f)}\biggr)\biggr\} \, dx = O\bigl(\max\{q_n \log n, q_n^{1-\epsilon}\}\bigr),$$

for every $\epsilon > 0$. Moreover,

$$\biggl| \int_{\mathcal{X}_n^c} f(x) \log f(x) \, dx \biggr| = O(q_n^{1-\epsilon}), \tag{4}$$

for every $\epsilon > 0$. Now, a slightly simpler argument than that used in the proof of Lemma 10(ii) in the main text gives that for $r \in (0, r_x]$, we have

$$|h_x(r) - V_d f(x) r^d| \leq \frac{dV_d}{d + \tilde{\beta}} C_{n,\tilde{\beta}}(x) r^{d+\tilde{\beta}}.$$

We deduce, again using a slightly simplified version of the argument in Lemma 10(ii) in the main text, that there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, $s \in [0, \frac{a_n}{n-1}]$ and $x \in \mathcal{X}_n$, we have

$$\bigl| V_d f(x) h_x^{-1}(s)^d - s \bigr| \leq \frac{2dV_d^{-\tilde{\beta}/d}}{d + \tilde{\beta}} s^{1+\tilde{\beta}/d} \frac{C_{n,\tilde{\beta}}(x)}{f(x)^{1+\tilde{\beta}/d}} \leq \frac{s}{2}. \tag{5}$$

It follows from (3), (4), (5) and an almost identical argument to that leading to (15) in the main text that for every $n \geq n_0$ and $\epsilon > 0$,

$$\begin{aligned}
|\mathbb{E}_f(\hat{H}_n) - H| &\leq \biggl| \int_{\mathcal{X}_n} f(x) \int_0^{\frac{a_n}{n-1}} \mathrm{B}_{k,n-k}(s) \log\biggl(\frac{V_d f(x) h_x^{-1}(s)^d}{s}\biggr) \, ds \, dx \biggr| \\
&\qquad + O\bigl(\max\{q_n^{1-\epsilon}, q_n \log n, n^{-1}\}\bigr) \\
&\leq 2 \int_{\mathcal{X}_n} f(x) \int_0^{\frac{a_n}{n-1}} \mathrm{B}_{k,n-k}(s) \biggl| \frac{V_d f(x) h_x^{-1}(s)^d - s}{s} \biggr| \, ds \, dx \\
&\qquad + O\bigl(\max\{q_n^{1-\epsilon}, q_n \log n, n^{-1}\}\bigr) \\
&\leq \frac{4dV_d^{-\tilde{\beta}/d}}{d + \tilde{\beta}} \frac{\mathrm{B}_{k+\tilde{\beta}/d,n-k}}{\mathrm{B}_{k,n-k}} \int_{\mathcal{X}_n} \frac{C_{n,\tilde{\beta}}(x)}{f(x)^{\tilde{\beta}/d}} \, dx + O\bigl(\max\{q_n^{1-\epsilon}, q_n \log n, n^{-1}\}\bigr),
\end{aligned}$$

as required. □

A.5. *Completion of the proof of Lemma 7 in the main text.* To prove Lemma 7 in the main text, it remains to bound several error terms arising from arguments that approximate the variance of the unweighted Kozachenko–Leonenko estimator $\hat{H}_n$, and then to show how these arguments may be adapted to yield the desired asymptotic expansion for $\mathrm{Var}(\hat{H}_n^w)$.

A.5.1. *Bounds on $S_1, \ldots, S_5$.* *To bound S_1:* By similar methods to those used to bound R_1 in the proof of Lemma 3 in the main text, it is straightforward to show that for every $\epsilon > 0$, we have

$$S_1 = \int_{\mathcal{X}_n^c} f(x) \int_0^1 \mathrm{B}_{k,n-k}(s) \log^2 u_{x,s} \, ds \, dx = O\biggl(\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}} \biggr).$$

To bound S_2: For every $\epsilon > 0$, we have that

$$S_2 = \int_{\mathcal{X}_n} f(x) \int_{\frac{a_n}{n-1}}^1 \mathrm{B}_{k,n-k}(s) \log^2 u_{x,s} \, ds \, dx = o(n^{-(3-\epsilon)}),$$

by very similar arguments to those used to bound R_2 in the proof of Lemma 3 in the main text.

To bound S_3: We have

$$\begin{aligned}
&\log^2 u_{x,s} - \log^2\biggl(\frac{(n-1)s}{e^{\Psi(k)} f(x)} \biggr) \\
&= \biggl\{ 2\log\biggl(\frac{(n-1)s}{e^{\Psi(k)} f(x)} \biggr) + \log\biggl(\frac{V_d f(x) h_x^{-1}(s)^d}{s} \biggr) \biggr\} \log\biggl(\frac{V_d f(x) h_x^{-1}(s)^d}{s} \biggr).
\end{aligned}$$

It therefore follows from Lemma 10(ii) in the main text that for every $\epsilon > 0$,

$$\begin{aligned}
S_3 &= \int_{\mathcal{X}_n} f(x) \int_0^{\frac{a_n}{n-1}} \mathrm{B}_{k,n-k}(s) \biggl\{ \log^2 u_{x,s} - \log^2\biggl(\frac{(n-1)s}{e^{\Psi(k)} f(x)} \biggr) \biggr\} \, ds \, dx \\
&= O\biggl\{ \max\biggl(\frac{k^{\beta/d}}{n^{\beta/d}} \log n \, , \, \frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}} \biggr) \biggr\}.
\end{aligned}$$

To bound S_4: A simplified version of the argument used to bound R_4 in Lemma 3 of the main text shows that for every $\epsilon > 0$,

$$S_4 = \int_{\mathcal{X}_n} f(x) \int_{\frac{a_n}{n-1}}^1 \mathrm{B}_{k,n-k}(s) \log^2\biggl(\frac{(n-1)s}{e^{\Psi(k)} f(x)} \biggr) \, ds \, dx = o(n^{-(3-\epsilon)}).$$

To bound S_5: Very similar arguments to those used to bound R_1 in Lemma 3 in the main text show that for every $\epsilon > 0$,

$$S_5 = \int_{\mathcal{X}_n^c} f(x) \log^2 f(x)\, dx = O\biggl(\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}\biggr).$$

A.5.2. *Bounds on T_1, T_2 and T_3.* *To bound T_1*: Let $\mathrm{B} \sim \mathrm{Beta}(k-1, n-k-1)$. By (13) in the main text, for every $\epsilon > 0$,

$$\begin{aligned}
T_{11} &:= \biggl|\int_{\mathcal{X}_n^c \times \mathcal{X}_n^c} f(x) f(y) \log f(y) \int_{\tilde{u}_{n,x,y}}^\infty \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\, dy\, dx\biggr| \\
&\leq \frac{n-2}{n-k-1} \int_{\mathcal{X}_n^c \times \mathcal{X}_n^c} f(x) f(y) |\log f(y)| \\
&\qquad\qquad \int_0^1 \bigl|\log(u_{x,s} f(x))\bigr| \mathrm{B}_{k-1,n-k-1}(s) \biggl|1 - \frac{(n-2)s}{k-1}\biggr|\, ds\, dy\, dx \\
&\lesssim \int_{\mathcal{X}_n^c \times \mathcal{X}_n^c} f(x) f(y) |\log f(y)| \biggl[\mathbb{E}\biggl\{\biggl(\log\frac{1}{\mathrm{B}} + \log\frac{1}{1-B}\biggr)\biggl|1 - \frac{(n-2)\mathrm{B}}{k-1}\biggr|\biggr\} \\
&\quad + \biggl\{\log n + |\log f(x)| + \log\biggl(1 + \frac{\|x\|}{\mu_\alpha^{1/\alpha}(f)}\biggr)\biggr\}\mathbb{E}\biggl|1 - \frac{(n-2)\mathrm{B}}{k-1}\biggr|\biggr]\, dy\, dx \\
&= o\biggl(\frac{k^{-\frac{1}{2}+\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{\frac{2\alpha}{\alpha+d}-\epsilon}}\biggr),
\end{aligned}$$

where we used the Cauchy–Schwarz inequality and elementary properties of beta random variables to obtain the final bound.

Now let

$$u_n^*(x) := u_{x,a_n/(n-1)} = \frac{V_d (n-1) h_x^{-1}(\frac{a_n}{n-1})^d}{e^{\Psi(k)}},$$

and consider

$$T_{12} := \biggl|\int_{\mathcal{X}_n^c} \int_{\mathcal{X}_n} f(x) f(y) \log f(y) \int_{\tilde{u}_{n,x,y}}^\infty \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\, dy\, dx\biggr|.$$

If $\tilde{u}_{n,x,y} \geq u_n^*(x)$, then by very similar arguments to those used to bound R_1 and R_2 (cf. (13) and (14) in the main text), together with Cauchy–Schwarz,

$$\begin{aligned}
&\int_{\tilde{u}_{n,x,y}}^\infty \bigl|\log(uf(x))\bigr|\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u) \\
&\qquad\qquad \leq \int_{\frac{a_n}{n-1}}^1 |\log(u_{x,s} f(x))| \{\mathrm{B}_{k-1,n-k}(s) + \mathrm{B}_{k,n-k-1}(s)\}\, ds \\
&\qquad\qquad \lesssim \frac{\log n + |\log f(x)| + \log\Bigl(1 + \frac{\|x\|}{\mu_\alpha^{1/\alpha}(f)}\Bigr)}{n^{3-\epsilon}}, \qquad (6)
\end{aligned}$$

for every $\epsilon > 0$. On the other hand, if $\tilde{u}_{n,x,y} < u_n^*(x)$, then $\|x - y\| < r_{n,u_n^*(x)} + r_{n,u_n^*(y)}$, where we have added the $r_{n,u_n^*(y)}$ term to aid a calculation later in the proof. Define the sequence

$$\rho_n := \bigl[c_n \log^{1/d}(n-1)\bigr]^{-1}.$$

From Lemma 10(ii) in the main text,

$$\sup_{y\in\mathcal{X}_n} r_{n,u_n^*(y)} = \sup_{y\in\mathcal{X}_n} h_y^{-1}\Bigl(\frac{a_n}{n-1}\Bigr) \lesssim \sup_{y\in\mathcal{X}_n}\Bigl\{\frac{k\log n}{nf(y)}\Bigr\}^{1/d} \leq \Bigl(\frac{k\log n}{n\delta_n}\Bigr)^{1/d} = o(\rho_n).$$

Now suppose that $x \in \mathcal{X}_n^c$ and $y \in \mathcal{X}_n$ satisfy $\|y - x\| \leq \rho_n$. Choose $n_0 \in \mathbb{N}$ large enough that $r_{n,u_n^*(y)} \leq \rho_n/2$ for all $y \in \mathcal{X}_n$, and that $\log(n-1) \geq \max\{(3/2)^d(8d^{1/2})^{d/\beta}, 12V_d^{-1}2^d\}$ for all $n \geq n_0$ and $k \in \{k_0^*, \dots, k_1^*\}$. Then when $\beta \in (0,1]$ and $n \geq n_0$, using the fact that $B_x(\rho_n/2) \subseteq B_y(3\rho_n/2)$, we have

$$\begin{aligned}\int_{B_x(\rho_n/2)} f(w)\,dw &\geq V_d f(y)(\rho_n/2)^d - V_d a(f(y))f(y)(\rho_n/2)^d(3\rho_n/2)^\beta \\ &\geq V_d f(y)(\rho_n/2)^d\{1-(3c_n\rho_n/2)^\beta\} \geq \frac{1}{2}V_d(\rho_n/2)^d\delta_n \geq \frac{a_n}{n-1}. \end{aligned} \tag{7}$$

Hence, for all $n \geq n_0$, $x \in \mathcal{X}_n^c$, $y \in \mathcal{X}_n$ with $\|y-x\| \leq \rho_n$ and $k \in \{k_0^*, \dots, k_1^*\}$,

$$r_{n,u_n^*(x)} + r_{n,u_n^*(y)} \leq \rho_n. \tag{8}$$

On other hand, suppose instead that $x \in \mathcal{X}_n^c$ and $\rho_x^* := \inf_{y\in\mathcal{X}_n}\|y-x\| \geq \rho_n$. Since $\mathcal{X}_n$ is a closed subset of $\mathbb{R}^d$, we can find $y^* \in \mathcal{X}_n$ such that $\|y^*-x\| = \rho_x^*$, and set $\tilde{x} := \frac{\rho_n}{\rho_x^*}x + \bigl(1 - \frac{\rho_n}{\rho_x^*}\bigr)y^*$. Then $\|\tilde{x} - y^*\| = \rho_n$, so from (7), we have $r_{n,u_n^*(\tilde{x})} \leq \rho_n/2$ for $n \geq n_0$ and $k \in \{k_0^*, \dots, k_1^*\}$. Since $B_{\tilde{x}}(\rho_n/2) \subseteq B_x(\rho_x^* - \rho_n/2)$, we deduce that $r_{n,u_n^*(x)} \leq \rho_x^* - \rho_n/2$ and

$$\{y \in \mathcal{X}_n : \|x-y\| < r_{n,u_n^*(x)} + r_{n,u_n^*(y)}\} = \emptyset \tag{9}$$

for $n \geq n_0$ and $k \in \{k_0^*, \dots, k_1^*\}$. But for $n \geq n_0$,

$$\sup_{x\in\mathcal{X}_n^c}\ \sup_{y\in\mathcal{X}_n:\|y-x\|\leq\rho_n} \frac{1}{f(y)}|f(x)-f(y)| \leq \frac{15d^{1/2}}{7}(c_n\rho_n)^\beta < \frac{1}{2}, \tag{10}$$

so that if $x \in \mathcal{X}_n^c$, $y \in \mathcal{X}_n$ and $\|x-y\| \leq \rho_n$, then $f(y) < 2\delta_n$ for $n \geq n_0$ and $k \in \{k_0^*, \dots, k_1^*\}$.

It therefore follows from (6), (8), (9), (10) and the argument used to bound T_{11} that for each $\epsilon > 0$ and $n \geq n_0$,

$$\begin{aligned}
T_{12} &\leq \int_{\mathcal{X}_n^c} \int_{\mathcal{X}_n} f(x)f(y)|\log f(y)| \mathbb{1}_{\{\|x-y\| < r_{n,u_n^*(x)} + r_{n,u_n^*(y)}\}} \\
&\qquad\qquad \int_0^\infty |\log(uf(x))| \, d(\tilde{F}_{n,x} - F_{n,x}^-)(u) \, dy \, dx + o(n^{-2}) \\
&\leq \int_{\mathcal{X}_n^c} \int_{y: f(y) < 2\delta_n} f(x)f(y)|\log f(y)| \\
&\qquad\qquad \int_0^\infty |\log(uf(x))| \, d(\tilde{F}_{n,x} - F_{n,x}^-)(u) \, dy \, dx + o(n^{-2}) \\
&= o\biggl(\frac{k^{-\frac{1}{2} + \frac{2\alpha}{\alpha+d} - \epsilon}}{n^{\frac{2\alpha}{\alpha+d} - \epsilon}}\biggr).
\end{aligned}$$

Finally for T_1, we define

$$\begin{aligned}
T_{13} := \biggl| \int_{\mathcal{X}_n} \int_{B_x^c\left(\frac{r_{n,1} d_n}{f(x)^{1/d}}\right)} & f(x)f(y) \log f(y) \\
& \int_{\tilde{u}_{n,x,y}}^\infty \log\bigl(uf(x)\bigr) \, d(\tilde{F}_{n,x} - F_{n,x}^-)(u) \, dy \, dx \biggr|.
\end{aligned}$$

By Lemma 10(ii) in the main text, we can find $n_1 \in \mathbb{N}$ such that for $n \geq n_1$, $k \in \{k_0^*, \ldots, k_1^*\}$, $x \in \mathcal{X}_n$ and $s \leq a_n/(n-1)$, we have $V_d f(x) h_x^{-1}(s)^d \leq 2s$. Thus, for $n \geq n_1$, $k \in \{k_0^*, \ldots, k_1^*\}$, $x \in \mathcal{X}_n$ and $y \in B_x^c(\frac{r_{n,1} d_n}{f(x)^{1/d}})$,

$$\tilde{u}_{n,x,y} \geq \frac{24 \log n}{f(x)} \geq \frac{2a_n}{f(x) e^{\Psi(k)}} \geq u_n^*(x).$$

Thus, from (6), $T_{13} = O(n^{-2} \log n)$. We conclude that for every $\epsilon > 0$,

$$|T_1| \leq T_{11} + T_{12} + T_{13} = o\biggl(\frac{k^{-\frac{1}{2} + \frac{2\alpha}{\alpha+d} - \epsilon}}{n^{\frac{2\alpha}{\alpha+d} - \epsilon}}\biggr).$$

To bound T_2: Fix $x \in \mathcal{X}_n$ and $z \in B_0(d_n)$. Choosing $n_2 \in \mathbb{N}$ large enough that $\frac{r_{n,1} d_n}{\delta_n^{1/d}} \leq (8d^{1/2})^{-1/\beta} c_n^{-1}$ for $n \geq n_2$, we have by Lemma 2 that

$$\sup_{y \in B_x\left(\frac{r_{n,1} d_n}{\delta_n^{1/d}}\right)} \left| \frac{f(y)}{f(x)} - 1 \right| \leq \frac{1}{2}$$

for $n \geq n_2$, $k \in \{k_0^*, \ldots, k_1^*\}$. Also, for all $n \geq n_2, k \in \{k_0^*, \ldots, k_1^*\}$, we have

$$\begin{aligned}\bigl|f(y_{x,z}) \log f(y_{x,z}) - f(x) \log f(x)\bigr| &\leq f(y_{x,z}) |\log(f(y_{x,z})/f(x))| + |\log f(x)||f(y_{x,z}) - f(x)| \\ &\leq a(f(x)) f(x) \|y_{x,z} - x\|^{\beta} \{|\log f(x)| + 4\}.\end{aligned}$$

Moreover, by arguments used to bound T_{11},

$$\begin{aligned}\biggl|\int_{\|z\|^d/f(x)}^{\infty} \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\biggr| &\lesssim \mathbb{E}\biggl|\log(\mathrm{B})\biggl(1 - \frac{(n-2)\mathrm{B}}{k-1}\biggr)\biggr| \\ &+ \biggl\{\log n + |\log f(x)| + \log\biggl(1 + \frac{\|x\|}{\mu_\alpha^{1/\alpha}(f)}\biggr)\biggr\} \mathbb{E}\biggl|1 - \frac{(n-2)\mathrm{B}}{k-1}\biggr|,\end{aligned}$$

where $\mathrm{B} \sim \mathrm{Beta}(k-1, n-k-1)$. It follows that for every $\epsilon > 0$,

$$\begin{aligned}T_2 &= \frac{e^{\Psi(k)}}{V_d(n-1)} \int_{\mathcal{X}_n} \int_{B_0(d_n)} \{f(y_{x,z}) \log f(y_{x,z}) - f(x) \log f(x)\} \\ &\qquad \int_{\|z\|^d/f(x)}^{\infty} \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\, dz\, dx \\ &= O\biggl(\frac{k^{1/2}}{n} \max\biggl\{\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}, \frac{k^{\beta/d}}{n^{\beta/d}} \log^{2+\beta/d} n\biggr\}\biggr).\end{aligned}$$

To bound T_3: Note that by Fubini's theorem,

$$\begin{aligned}&\int_{\mathcal{X}_n} f(x) \log f(x) \int_{B_0(d_n)} \int_{\frac{\|z\|^d}{f(x)}}^{\infty} \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\, dz\, dx \\ &= V_d \int_{\mathcal{X}_n} f(x) \log f(x) \int_0^{\infty} \min\{uf(x), d_n^d\} \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\, dx \\ &= V_d \int_{\mathcal{X}_n} f(x) \log f(x) \int_0^{u_n^*(x)} uf(x) \log(uf(x))\, d(\tilde{F}_{n,x} - F_{n,x}^-)(u)\, dx \\ &\qquad + O(n^{-(3-\epsilon)}),\end{aligned}$$

for all $\epsilon > 0$, where the order of the error term follows from a similar argument to that used to obtain (6) and Lemma 10(i). Thus, for every $\epsilon > 0$,

$$\begin{aligned}T_3 &= \frac{k-1}{n-k-1} \int_{\mathcal{X}_n} f(x) \log f(x) \int_0^{\frac{a_n}{n-1}} \biggl\{\frac{V_d f(x) h_x^{-1}(s)^d}{s} \log(u_{x,s} f(x)) \\ &\qquad - \log\biggl(\frac{(n-1)s}{e^{\Psi(k)}}\biggr)\biggr\} \mathrm{B}_{k,n-k-1}(s) \biggl\{1 - \frac{(n-2)s}{k-1}\biggr\}\, ds\, dx + O(n^{-(3-\epsilon)}) \\ &= O\biggl(\frac{k^{1/2}}{n} \max\biggl\{\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}, \frac{k^{\beta/d}}{n^{\beta/d}} \log n\biggr\}\biggr).\end{aligned}$$

A.5.3. *Bounds on U_1 and U_2.* *To bound U_1*: Using Lemma 10(i) and (13) in the main text as in our bounds on T_{11} we have that for every $\epsilon > 0$,

$$
\begin{aligned}
U_{11} &:= \biggl| \int_{\mathcal{X}_n^c} f(x) \int_0^{u_n^*(x)} \log\bigl(uf(x)\bigr)\, d(F_{n,x}^- - F_{n,x})(u)\, dx \biggr| \\
&\leq \int_{\mathcal{X}_n^c} f(x) \int_0^{\frac{a_n}{n-1}} |\log(u_{x,s} f(x))| \mathrm{B}_{k,n-k-1}(s) \biggl| \frac{(n-1)s-k}{n-k-1} \biggr|\, ds\, dx \\
&= o\biggl(\frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}} \biggr). \qquad (11)
\end{aligned}
$$

Moreover, using arguments similar to those used to bound R_2 in the proof of Lemma 3 in the main text, for every $\epsilon > 0$,

$$
U_{12} := \biggl| \int_{\mathcal{X}} f(x) \int_{u_n^*(x)}^{\infty} \log\bigl(uf(x)\bigr)\, d(F_{n,x}^- - F_{n,x})(u)\, dx \biggr| = o(n^{-(3-\epsilon)}). \qquad (12)
$$

From (11), and (12), we have for every $\epsilon > 0$ that

$$
|U_1| \leq U_{11} + U_{12} = o\biggl(\frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}} \biggr).
$$

To bound U_2: By Lemma 10(ii) and letting $\mathrm{B} \sim \mathrm{Beta}(k+\beta/d, n-k-1)$, we have that for every $\epsilon > 0$,

$$
\begin{aligned}
U_{21} &:= \biggl| \int_{\mathcal{X}_n} f(x) \int_0^{\frac{a_n}{n-1}} \log\Bigl(\frac{V_d f(x) h_x^{-1}(s)^d}{s} \Bigr) \mathrm{B}_{k,n-k-1}(s) \Bigl\{ \frac{(n-1)s-k}{n-k-1} \Bigr\} ds\, dx \biggr| \\
&\lesssim \frac{k^{\beta/d}}{n^{\beta/d}} \mathbb{E}\biggl(\biggl| \frac{(n-1)\mathrm{B}-k}{n-k-1} \biggr| \biggr) \int_{\mathcal{X}_n} a(f(x)) f(x)^{1-\beta/d}\, dx \\
&= O\biggl(\frac{k^{1/2}}{n} \max\biggl\{ \frac{k^{\beta/d}}{n^{\beta/d}}, \frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}} \biggr\} \biggr).
\end{aligned}
$$

Moreover, we can use similar arguments to those used to bound R_4 in the proof of Lemma 3 in the main text to show that for every $\epsilon > 0$,

$$
\begin{aligned}
U_{22} &:= \biggl| \int_{\mathcal{X}_n} f(x) \int_{\frac{a_n}{n-1}}^1 \log\biggl(\frac{(n-1)s}{e^{\Psi(k)}} \biggr) \mathrm{B}_{k,n-k-1}(s) \biggl\{ \frac{(n-1)s-k}{n-k-1} \biggr\}\, ds\, dx \biggr| \\
&= o(n^{-(3-\epsilon)}).
\end{aligned}
$$

We deduce that for every $\epsilon > 0$,

$$
|U_2| \leq U_{21} + U_{22} = O\biggl(\frac{k^{1/2}}{n} \max\biggl\{ \frac{k^{\beta/d}}{n^{\beta/d}}, \frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}} \biggr\} \biggr).
$$

A.5.4. *Bounds on* $W_1, \ldots, W_4$. *To bound* W_1: We partition the region $([l_x, v_x] \times [l_y, v_y])^c$ into eight rectangles as follows:

$$\begin{aligned}\big([l_x, v_x] \times [l_y, v_y]\big)^c = &\big([0, l_x) \times [0, l_y)\big) \cup \big([0, l_x) \times [l_y, v_y]\big) \cup \big([0, l_x) \times (v_y, \infty)\big) \\ &\cup \big([l_x, v_x] \times [0, l_y)\big) \cup \big([l_x, v_x] \times (v_y, \infty)\big) \cup \big((v_x, \infty) \times [0, l_y)\big) \\ &\cup \big((v_x, \infty) \times [l_y, v_y]\big) \cup \big((v_x, \infty) \times (v_y, \infty)\big).\end{aligned}$$

Recall our shorthand $h(u, v) = \log(uf(x)) \log(vf(y))$. By Lemma 10(i) in the main text and the Cauchy–Schwarz inequality, as well as very similar arguments to those used to bound R_2 in the proof of Lemma 3 in the main text, we can bound the contributions from each rectangle individually, to obtain that for every $\epsilon > 0$,

$$\begin{aligned}W_1 &= \int_{\mathcal{X} \times \mathcal{X}} f(x) f(y) \int_{([l_x, v_x] \times [l_y, v_y])^c} h(u, v)\, d(F_{n,x,y} - F_{n,x} F_{n,y})(u, v)\, dx\, dy \\ &= o(n^{-(9/2-\epsilon)}).\end{aligned}$$

To bound W_2: We have

$$W_2 = \int_{\mathcal{X} \times \mathcal{X}} f(x) f(y) \int_{l_x}^{v_x} \int_{l_y}^{v_y} h(u, v)\, d(G_{n,x,y} - F_{n,x} F_{n,y})(u, v)\, dx\, dy + \frac{1}{n}.$$

We write $\mathrm{B}_{a,b,c} := \frac{\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(a+b+c)}$, and, for $s, t > 0$ with $s + t < 1$, let

$$\mathrm{B}_{a,b,c}(s, t) := \frac{s^{a-1} t^{b-1} (1 - s - t)^{c-1}}{\mathrm{B}_{a,b,c}} \tag{13}$$

denote the density of a Dirichlet(a, b, c) random vector at (s, t). For $a, b > -1$, writing $I_n := [a_n^-/(n-1), a_n^+/(n-1)]$, let

$$\begin{aligned}\mathrm{B}^{(n)}_{k+a,n-k} &:= \int_{I_n} s^{k+a-1} (1 - s)^{n-k-1}\, ds, \\ \mathrm{B}^{(n)}_{k+a,n-k}(s) &:= s^{k+a-1} (1 - s)^{n-k-1} / \mathrm{B}^{(n)}_{k+a,n-k} \\ \mathrm{B}^{(n)}_{k+a,k+b,n-2k-1} &:= \int_{I_n \times I_n} s^{k+a-1} t^{k+b-1} (1 - s - t)^{n-2k-2}\, ds\, dt \\ \mathrm{B}^{(n)}_{k+a,k+b,n-2k-1}(s, t) &:= s^{k+a-1} t^{k+b-1} (1 - s - t)^{n-2k-2} / \mathrm{B}^{(n)}_{k+a,k+b,n-2k-1}.\end{aligned}$$

Then by the triangle and Pinsker's inequalities, and Beta tail bounds similar to those used previously, we have that

$$\int_{I_n\times I_n} \bigl|\mathrm{B}_{k+a,k+b,n-2k-1}(s,t)-\mathrm{B}_{k+a,n-k}(s)\mathrm{B}_{k+b,n-k}(t)\bigr|\,ds\,dt$$
$$\le \left|\frac{\mathrm{B}^{(n)}_{k+a,k+b,n-2k-1}}{\mathrm{B}_{k+a,k+b,n-2k-1}}-1\right| + \left|\frac{\mathrm{B}^{(n)}_{k+a,n-k}\mathrm{B}^{(n)}_{k+b,n-k}}{\mathrm{B}_{k+a,n-k}\mathrm{B}_{k+b,n-k}}-1\right|$$
$$+\left\{2\int_{I_n\times I_n}\mathrm{B}^{(n)}_{k+a,k+b,n-2k-1}(s,t)\log\Bigl(\frac{\mathrm{B}^{(n)}_{k+a,k+b,n-2k-1}(s,t)}{\mathrm{B}^{(n)}_{k+a,n-k}(s)\mathrm{B}^{(n)}_{k+b,n-k}(t)}\Bigr)\,ds\,dt\right\}^{1/2}$$
$$=\left\{2\int_0^1\int_0^{1-t}\mathrm{B}_{k+a,k+b,n-2k-1}(s,t)\log\biggl(\frac{\mathrm{B}_{k+a,k+b,n-2k-1}(s,t)}{\mathrm{B}_{k+a,n-k}(s)\mathrm{B}_{k+b,n-k}(t)}\biggr)\,ds\,dt\right\}^{1/2} + o(n^{-2})$$
$$=2^{1/2}\biggl[\log\Bigl(\frac{\Gamma(n+a+b-1)\Gamma(n-k)^2}{\Gamma(n-2k-1)\Gamma(n+a)\Gamma(n+b)}\Bigr)+(n-2k-2)\psi(n-2k-1)$$
$$-(n-k-1)\{\psi(n+b-k-1)+\psi(n+a-k-1)\}$$
$$+n\psi(n+a+b-1)\biggr]^{1/2}+o(n^{-2})$$

(14)
$$=\frac{k}{n}\{1+o(1)\}.$$

As a first step towards bounding W_2 note that

$$W_{21}:=\int_{\mathcal{X}_n\times\mathcal{X}_n} f(x)f(y)\int_{l_x}^{v_x}\int_{l_y}^{v_y} h(u,v)\,d(G_{n,x,y}-F_{n,x}F_{n,y})(u,v)\,dx\,dy$$
$$=\int_{\mathcal{X}_n\times\mathcal{X}_n} f(x)f(y)\int_{I_n\times I_n}\log(u_{x,s}f(x))\log(u_{y,t}f(y))$$
$$\bigl\{\mathrm{B}_{k,k,n-2k-1}(s,t)-\mathrm{B}_{k,n-k}(s)\mathrm{B}_{k,n-k}(t)\bigr\}\,ds\,dt\,dx\,dy$$
$$=\int_{\mathcal{X}_n\times\mathcal{X}_n} f(x)f(y)\int_{I_n\times I_n}\log\biggl(\frac{(n-1)s}{e^{\Psi(k)}}\biggr)\log\biggl(\frac{(n-1)t}{e^{\Psi(k)}}\biggr)$$
$$\bigl\{\mathrm{B}_{k,k,n-2k-1}(s,t)-\mathrm{B}_{k,n-k}(s)\mathrm{B}_{k,n-k}(t)\bigr\}\,ds\,dt\,dx\,dy+W_{211}$$

(15)
$$=-\frac{1}{n}+O\biggl(\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}}\biggr)+O(n^{-2})+W_{211},$$

for every $\epsilon>0$. But, by Lemma 10(ii) in the main text and (14), for every

$\epsilon > 0$,

$$\begin{aligned} |W_{211}| &= \biggl| \int_{\mathcal{X}_n \times \mathcal{X}_n} f(x)f(y) \int_{I_n \times I_n} \biggl\{ 2\log\biggl(\frac{V_d h_x^{-1}(s)^d f(x)}{s}\biggr) \log\biggl(\frac{(n-1)t}{e^{\Psi(k)}}\biggr) \\ &\qquad + \log\biggl(\frac{V_d h_x^{-1}(s)^d f(x)}{s}\biggr) \log\biggl(\frac{V_d h_y^{-1}(t)^d f(y)}{t}\biggr) \biggr\} \\ &\qquad \bigl\{ \mathrm{B}_{k,k,n-2k-1}(s,t) - \mathrm{B}_{k,n-k}(s)\mathrm{B}_{k,n-k}(t) \bigr\} \, ds \, dt \, dx \, dy \biggr| \\ &\leq 2 \biggl| \int_{\mathcal{X}_n \times \mathcal{X}_n} f(x)f(y) \int_{I_n} \log\biggl(\frac{V_d h_x^{-1}(s)^d f(x)}{s}\biggr) \\ &\qquad \Bigl[\bigl\{ \log(n-1) - \Psi(n-k-1) + \log(1-s) \bigr\} \mathrm{B}_{k,n-k-1}(s) \\ &\qquad - \bigl\{ \log(n-1) - \Psi(n) \bigr\} \mathrm{B}_{k,n-k}(s) \Bigr] \, ds \, dx \, dy \biggr| \\ &\qquad + O\biggl(\max\biggl\{ \frac{k^{1+\frac{2\beta}{d}}}{n^{1+\frac{2\beta}{d}}}, \frac{k^{1+\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{2\alpha}{\alpha+d}-\epsilon}} \biggr\} \biggr) \\ &= O\biggl(\frac{k^{1/2}}{n} \max\biggl\{ \frac{k^{\beta/d}}{n^{\beta/d}}, \frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}} \biggr\} \biggr). \end{aligned} \tag{16}$$

Moreover, by Lemma 10(i) and (ii) in the main text and very similar arguments, for every $\epsilon > 0$,

$$\begin{aligned} W_{22} &:= \int_{\mathcal{X}_n \times \mathcal{X}_n^c} f(x)f(y) \int_{l_x}^{v_x} \int_{l_y}^{v_y} h(u,v) \, d(G_{n,x,y} - F_{n,x}F_{n,y})(u,v) \, dx \, dy \\ &= O\biggl(\frac{k^{1+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}} \max\biggl\{ \frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}, \frac{k^{\beta/d}}{n^{\beta/d}}, \frac{1}{k^{1/2}} \biggr\} \biggr) \\ W_{23} &:= \int_{\mathcal{X}_n^c \times \mathcal{X}_n^c} f(x)f(y) \int_{l_x}^{v_x} \int_{l_y}^{v_y} h(u,v) \, d(G_{n,x,y} - F_{n,x}F_{n,y})(u,v) \, dx \, dy \\ &= O\biggl(\frac{k^{1+\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{2\alpha}{\alpha+d}-\epsilon}} \biggr). \end{aligned} \tag{17}$$

Incorporating our restrictions on k, we conclude from (15), (16) and (17) that for every $\epsilon > 0$,

$$|W_2| \leq \biggl| W_{21} + \frac{1}{n} \biggr| + 2|W_{22}| + |W_{23}| = O\biggl(\frac{k^{1/2}}{n} \max\biggl\{ \frac{k^{\beta/d}}{n^{\beta/d}}, \frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}} \biggr\} \biggr).$$

To bound W_3: We write h_u, h_v and h_{uv} for the partial derivatives of $h(u,v)$ and write, for example, $(h_u F)(u,v) = h_u(u,v)F(u,v)$. We find on integrating by parts that, writing $F = F_{n,x,y} - G_{n,x,y}$,

$$\int_{[l_x,v_x]\times[l_y,v_y]} (h\, dF)(u,v) - \int_{l_x}^{v_x}\int_{l_y}^{v_y} (h_{uv}F(u,v))\, du\, dv$$
$$= \int_{l_x}^{v_x}\big[(h_uF)(u,l_y) - (h_uF)(u,v_y)\big]du + \int_{l_y}^{v_y}\big[(h_vF)(l_x,v) - (h_vF)(v_x,v)\big]dv$$
$$+ (hF)(v_x,v_y) + (hF)(l_x,l_y) - (hF)(v_x,l_y) - (hF)(l_x,v_y). \tag{18}$$

Using standard binomial tail bounds as used to bound W_1 together with (13) in the main text we therefore see that for every $\epsilon > 0$,

$$W_{31} := \int_{\mathcal{X}\times\mathcal{X}} f(x)f(y)\Big\{\int_{l_x}^{v_x}\int_{l_y}^{v_y}(h\, dF)(u,v) - \int_{l_x}^{v_x}\int_{l_y}^{v_y}(h_{uv}F)(u,v)\, du\, dv\Big\} dx\, dy$$
$$= - \int_{\mathcal{X}\times\mathcal{X}} f(x)f(y)\Big\{\int_{l_x}^{v_x}(h_uF)(u,v_y)\, du + \int_{l_y}^{v_y}(h_vF)(v_x,v)\, dv\Big\}\, dx\, dy$$
$$+ o(n^{-(9/2-\epsilon)}). \tag{19}$$

Now, uniformly for $u \in [l_x, v_x]$ and $(x,y) \in \mathcal{X}\times\mathcal{X}$ and for every $\epsilon > 0$,

$$F(u,v_y) = \mathbb{1}_{\{\|x-y\|\le r_{n,u}\}}\binom{n-2}{k-1}p_{n,x,u}^{k-1}(1-p_{n,x,u})^{n-k-1} + o(n^{-(9/2-\epsilon)})$$
$$= \mathbb{1}_{\{\|x-y\|\le r_{n,u}\}}\frac{\mathrm{B}_{k,n-k}(p_{n,x,u})}{n-1} + o(n^{-(9/2-\epsilon)})$$
$$\le \mathbb{1}_{\{\|x-y\|\le r_{n,v_x}\}}\frac{1}{(2\pi k)^{1/2}}\{1+o(1)\} + o(n^{-(9/2-\epsilon)}). \tag{20}$$

By (10) and the arguments leading up to it, we have

$$\sup_{x\in\mathcal{X}_n^c}\ \sup_{y\in\mathcal{X}_n\cap B_x(r_{n,v_x}+r_{n,v_y})}\Big|\frac{f(x)}{f(y)} - 1\Big| \to 0. \tag{21}$$

We therefore have by (13) in the main text that, for every $\epsilon > 0$,

$$\int_{\mathcal{X}_n^c\times\mathcal{X}} f(x)f(y)\int_{l_x}^{v_x}(h_uF)(u,v_y)\, du\, dy\, dx = O\bigg(\frac{k^{-\frac{1}{2}+\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{\frac{2\alpha}{\alpha+d}-\epsilon}}\bigg). \tag{22}$$

Now, using Lemma 10(ii) in the main text, for $x \in \mathcal{X}_n$,

$$\max\{|l_xf(x)-1|, |v_xf(x)-1|\} \lesssim a(f(x))\Big(\frac{k}{nf(x)}\Big)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}. \tag{23}$$

We also need some control over $vf(y)$. By (10) and the work leading up to it, for $n \geq \max(n_0, 5), x \in \mathcal{X}_n$ and $\|y - x\| \leq r_{n,v_x} + r_{n,v_y}$,

$$f(y) \geq \Big\{1 - \frac{15d^{1/2}}{7}(c_n \rho_n)^\beta\Big\}\delta_n \geq \delta_n/2 \geq k/(n-1).$$

Thus $a(f(y)) \leq c_n^\beta$ and using (21) we may apply Lemma 10(ii) in the main text to the set

$$\mathcal{X}_n' = \mathcal{X}_n \cup \{y : \|y - x\| \leq r_{n,v_x} + r_{n,v_y} \text{ for some } x \in \mathcal{X}_n\}.$$

From this and (21), for any $x \in \mathcal{X}_n$ and $y \in B_x(r_{n,v_x} + r_{n,v_y})$,

$$\max(|l_y f(y) - 1|, |v_y f(y) - 1|) \lesssim a(f(y))\Big(\frac{k}{nf(x)}\Big)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}. \tag{24}$$

Using (21) again, we have that $a(f(y_{x,z})) \lesssim f(x)^{-\epsilon}$ for each $\epsilon > 0$, uniformly for $x \in \mathcal{X}_n$ and $\|z\| \leq \{v_x f(x)\}^{1/d} + \{v_y f(x)\}^{1/d}$. From (20), (23) and (24) we therefore have that

$$\begin{aligned}
&\Big|\int_{\mathcal{X}_n \times \mathcal{X}} f(x)f(y)\int_{l_x}^{v_x} (h_u F)(u, v_y)\,du\,dy\,dx\Big| \\
&\lesssim k^{-1/2}\int_{\mathcal{X}_n \times \mathcal{X}} f(x)f(y)\mathbb{1}_{\{\|x-y\| < r_{n,v_x}\}}|\log(v_y f(y))|\log(v_x/l_x)\,dy\,dx \\
&= O\bigg(\max\bigg\{\frac{k^{1/2+2\beta/d}}{n^{1+2\beta/d}}, \frac{\log n}{nk^{1/2}}, \frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}}\bigg\}\bigg) \qquad (25)
\end{aligned}$$

for every $\epsilon > 0$. By (19), (22) and (25), it follows that

$$W_{31} = O\bigg(\max\bigg\{\frac{k^{1/2+2\beta/d}}{n^{1+2\beta/d}}, \frac{\log n}{nk^{1/2}}, \frac{k^{-1/2+\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{\frac{2\alpha}{\alpha+d}-\epsilon}}\bigg\}\bigg). \tag{26}$$

Finally, by (13) in the main text and (21), we have since $F = 0$ when $\|x - y\| > r_{n,u} + r_{n,v}$ that

(27)

$$W_{32} := \int_{\mathcal{X}_n^c \times \mathcal{X}} f(x)f(y)\int_{l_x}^{v_x}\int_{l_y}^{v_y} (h_{uv}F)(u, v)\,du\,dv\,dx\,dy = O\bigg(\frac{k^{\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{\frac{2\alpha}{\alpha+d}-\epsilon}}\bigg).$$

Combining (26) and (27) we have that

$$W_3 = W_{31} + W_{32} = O\bigg(\max\bigg\{\frac{k^{1/2+2\beta/d}}{n^{1+2\beta/d}}, \frac{\log n}{nk^{1/2}}, \frac{k^{\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{\frac{2\alpha}{\alpha+d}-\epsilon}}\bigg\}\bigg).$$

To bound W_4: Let $p_\cap := \int_{B_x(r_{n,u})\cap B_y(r_{n,v})} f(w)\,dw$ and let $(N_1,N_2,N_3,N_4) \sim \mathrm{Multi}(n-2, p_{n,x,u}-p_\cap, p_{n,y,v}-p_\cap, p_\cap, 1-p_{n,x,u}-p_{n,y,v}+p_\cap)$. Further, let

$$F^{(1)}_{n,x,y}(u,v) := \mathbb{P}(N_1+N_3 \geq k, N_2+N_3 \geq k),$$

so that

$$\begin{aligned}(F_{n,x,y}-F^{(1)}_{n,x,y})(u,v) &= \mathbb{P}(N_1+N_3 = k-1, N_2+N_3 \geq k)\mathbb{1}_{\{\|x-y\|\leq r_{n,u}\}} \\ &+ \mathbb{P}(N_2+N_3 = k-1, N_1+N_3 \geq k)\mathbb{1}_{\{\|x-y\|\leq r_{n,v}\}} \\ &+ \mathbb{P}(N_1+N_3 = k-1, N_2+N_3 = k-1)\mathbb{1}_{\{\|x-y\|\leq r_{n,u}\wedge r_{n,v}\}}.\end{aligned}$$

Now $\mathbb{P}(N_1+N_3=k-1) = \binom{n-2}{k-1}p_{n,x,u}^{k-1}(1-p_{n,x,u})^{n-k-1} \leq (2\pi k)^{-1/2}\{1+o(1)\}$ and $F_{n,x,y}(u,v) = G_{n,x,y}(u,v)$ if $\|x-y\| > r_{n,u}+r_{n,v}$, and so, by (23) and (24), we have that

$$\begin{aligned}&\int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_x}^{v_x}\int_{l_y}^{v_y}\frac{(F_{n,x,y}-G_{n,x,y})(u,v)}{uv}\,du\,dv\,dx\,dy \\ &= \int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_x}^{v_x}\int_{l_y}^{v_y}\frac{(F^{(1)}_{n,x,y}-G_{n,x,y})(u,v)}{uv}\,du\,dv\,dx\,dy \\ &\quad + O\biggl(\max\biggl\{\frac{\log n}{nk^{1/2}}, \frac{k^{\frac{1}{2}+\frac{2\beta}{d}}}{n^{1+\frac{2\beta}{d}}}, \frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}}\biggr\}\biggr). \qquad (28)\end{aligned}$$

We can now approximate $F^{(1)}_{n,x,y}(u,v)$ by $\Phi_\Sigma(k^{1/2}\{uf(x)-1\}, k^{1/2}\{vf(x)-1\})$ and $G_{n,x,y}(u,v)$ by $\Phi(k^{1/2}\{uf(x)-1\})\Phi(k^{1/2}\{vf(x)-1\})$. To avoid repetition, we focus on the former of these terms. To this end, for $i = 3,\ldots,n$, let

$$Y_i := \begin{pmatrix}\mathbb{1}_{\{X_i\in B_x(r_{n,u})\}} \\ \mathbb{1}_{\{X_i\in B_y(r_{n,v})\}}\end{pmatrix},$$

so that $\sum_{i=3}^n Y_i = \begin{pmatrix}N_1+N_3 \\ N_2+N_3\end{pmatrix}$. We also define

$$\begin{aligned}\mu &:= \mathbb{E}(Y_i) = \begin{pmatrix}p_{n,x,u} \\ p_{n,y,v}\end{pmatrix} \\ V &:= \mathrm{Cov}(Y_i) = \begin{pmatrix}p_{n,x,u}(1-p_{n,x,u}) & p_\cap - p_{n,x,u}p_{n,y,v} \\ p_\cap - p_{n,x,u}p_{n,y,v} & p_{n,y,v}(1-p_{n,y,v})\end{pmatrix},\end{aligned}$$

When $x \in \mathcal{X}_n$ and $y \in B_x^\circ(r_{n,v_x}+r_{n,v_y})$ we have that, writing Δ for the symmetric difference and using (21), $\mathbb{P}(X_1 \in B_x(r_{n,u})\Delta B_y(r_{n,v})) > 0$ and so V is invertible for such x and y. We may therefore set $Z_i := V^{-1/2}(Y_i-\mu)$.

Then by the Berry–Esseen bound of Götze (1991), writing $\mathcal{C}$ for the set of closed, convex subsets of $\mathbb{R}^2$ and letting $Z \sim N_2(0, I)$, there exists a universal constant $C_2 > 0$ such that

$$\sup_{C \in \mathcal{C}} \biggl| \mathbb{P}\biggl(\frac{1}{(n-2)^{1/2}} \sum_{i=3}^{n} Z_i \in C \biggr) - \mathbb{P}(Z \in C) \biggr| \leq \frac{C_2 \mathbb{E}(\|Z_3\|^3)}{(n-2)^{1/2}}. \tag{29}$$

The distribution of Z_3 depends on x, y, u and v, but, recalling the substitution $y = y_{x,z}$ as defined in (22) in the main text, we claim that for $x \in \mathcal{X}_n$, $y = y_{x,z} \in B_x(r_{n,u} + r_{n,v})$, $u \in [l_x, v_x]$ and $v \in [l_y, v_y]$,

$$\mathbb{E}(\|Z_3\|^3) \lesssim \biggl(\frac{n}{k\|z\|} \biggr)^{1/2}. \tag{30}$$

To establish this, note that for $x \in \mathcal{X}_n$ and $\|y - x\| \leq r_{n,v_x} + r_{n,v_y}$, we have by (21), (23) and (24) that $\|y - x\| \lesssim (\frac{k}{nf(x)})^{1/d}$. Thus, for $v \in [l_y, v_y]$, and using Lemma 2, we also have that

$$\begin{aligned} |vf(x) - 1| &\leq \max(|v_y f(y) - 1|, |l_y f(y) - 1|) + v_y |f(y) - f(x)| \\ &\lesssim a(f(x) \wedge f(y)) \biggl(\frac{k}{nf(x)} \biggr)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}. \end{aligned} \tag{31}$$

Now, by the definition of l_x and v_x,

$$\max\bigl\{ |p_{n,x,u} - k/(n-1)| \, , \, |p_{n,y,v} - k/(n-1)| \bigr\} \leq \frac{3k^{1/2} \log^{1/2} n}{n-1} \tag{32}$$

for all $x, y \in \mathcal{X}$ and $u \in [l_x, v_x], v \in [l_y, v_y]$. Next, we bound $|\frac{n-2}{k} p_\cap - \alpha_z|$ for $x \in \mathcal{X}_n$ and $y = y_{x,z}$ with $\|z\| \leq \{v_x f(x)\}^{1/d} + \{v_y f(x)\}^{1/d}$. First suppose that $u \geq v$. We may write

$$B_x(r_{n,u}) \cap B_y(r_{n,v}) = \{B_x(r_{n,v}) \cap B_y(r_{n,v})\} \cup [\{B_x(r_{n,u}) \setminus B_x(r_{n,v})\} \cap B_y(r_{n,v})],$$

where this is a disjoint union. Writing $I_{a,b}(x) := \int_0^x \mathrm{B}_{a,b}(s)\, ds$ for the regularised incomplete beta function and recalling that μ_d denotes Lebesgue measure on $\mathbb{R}^d$, we have

$$\begin{aligned} \mu_d\bigl(B_x(r_{n,v}) \cap B_y(r_{n,v})\bigr) &= V_d r_{n,v}^d I_{\frac{d+1}{2}, \frac{1}{2}} \biggl(1 - \frac{\|x - y\|^2}{4 r_{n,v}^2} \biggr) \\ &= \frac{v e^{\Psi(k)}}{n-1} I_{\frac{d+1}{2}, \frac{1}{2}} \biggl(1 - \frac{\|z\|^2}{4\{vf(x)\}^{2/d}} \biggr) \end{aligned}$$

and

$$\alpha_z = I_{\frac{d+1}{2},\frac{1}{2}}\biggl(1 - \frac{\|z\|^2}{4}\biggr).$$

Now,

$$\biggl|\frac{d}{dr} I_{\frac{d+1}{2},\frac{1}{2}}\biggl(1 - \frac{r^2}{4}\biggr)\biggr| = \frac{(1-r^2/4)^{\frac{d-1}{2}}}{\mathrm{B}_{(d+1)/2,1/2}} \leq \frac{1}{\mathrm{B}_{(d+1)/2,1/2}}.$$

Hence by the mean value inequality,

$$\biggl|\mu_d\bigl(B_x(r_{n,v}) \cap B_y(r_{n,v})\bigr) - \frac{e^{\Psi(k)}\alpha_z}{(n-1)f(x)}\biggr| \\ \leq \frac{e^{\Psi(k)}}{n-1}\biggl[\frac{v\|z\| |1 - \{vf(x)\}^{-1/d}|}{\mathrm{B}_{(d+1)/2,1/2}} + \frac{\alpha_z}{f(x)}|1 - vf(x)|\biggr].$$

It follows that for all $x \in \mathcal{X}_n$, $y \in B_x(r_{n,v_x} + r_{n,v_y})$ and $v \in [l_y, v_y]$,

$$\biggl|\int_{B_x(r_{n,v})\cap B_y(r_{n,v})} f(w)\,dw - \frac{e^{\Psi(k)}\alpha_z}{n-1}\biggr| \\ \lesssim \frac{k}{n} a(f(x)\wedge f(y))\biggl(\frac{k}{nf(x)}\biggr)^{\beta/d} + \frac{k^{1/2}\log^{1/2} n}{n}$$

using (31) and Lemma 2. We also have by (32) that

$$\int_{\{B_x(r_{n,u})\setminus B_x(r_{n,v})\}\cap B_y(r_{n,v})} f(w)\,dw \leq p_{n,x,u} - p_{n,x,v} \\ \lesssim \frac{k}{n} a(f(x)\wedge f(y))\biggl(\frac{k}{nf(x)}\biggr)^{\beta/d} + \frac{k^{1/2}\log^{1/2} n}{n}.$$

Thus, when $x \in \mathcal{X}_n$, $y = y_{x,z} \in B_x(r_{n,v_x} + r_{n,v_y})$, $u \in [l_x, v_x]$, $v \in [l_y, v_y]$ and $u \geq v$,

$$\biggl|\frac{n-2}{k} p_\cap - \alpha_z\biggr| \lesssim a(f(x)\wedge f(y))\biggl(\frac{k}{nf(x)}\biggr)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}. \tag{33}$$

We can prove the same bound when $v > u$ similarly, using (23), (31) and Lemma 2. We will also require a lower bound on $p_{n,x,u} + p_{n,y,v} - 2p_\cap$ in the region where $B_x(r_{n,u}) \cap B_y(r_{n,v}) \neq \emptyset$, i.e., $\|z\| \leq \{uf(x)\}^{1/d} + \{vf(x)\}^{1/d}$. By the mean value theorem,

$$1 - I_{\frac{d+1}{2},\frac{1}{2}}(1-\delta^2) \geq 2^{1/2}\delta \max\biggl\{\frac{2^{-d/2}}{\mathrm{B}_{(d+1)/2,1/2}}\,,\, 1 - I_{\frac{d+1}{2},\frac{1}{2}}(1/2)\biggr\}$$

for all $\delta \in [0,1]$. Thus, for $u \geq v$, with $v \in [l_y, v_y]$, $x \in \mathcal{X}_n$, and $y = y_{x,z}$ with $\|z\| \leq 2\{vf(x)\}^{1/d}$, by (31) we have,

$$\mu_d\bigl(B_x(r_{n,u}) \cap B_y(r_{n,v})^c\bigr) \geq \mu_d\bigl(B_x(r_{n,v}) \cap B_y(r_{n,v})^c\bigr) = V_d r_{n,v}^d \biggl\{1 - I_{\frac{d+1}{2},\frac{1}{2}}\biggl(1 - \frac{\|x-y\|^2}{4r_{n,v}^2}\biggr)\biggr\} \gtrsim \frac{k\|z\|}{nf(x)}.$$

When $\|z\| > 2\{vf(x)\}^{1/d}$ we simply have $\mu_d\bigl(B_x(r_{n,v}) \cap B_y(r_{n,v})^c\bigr) = V_d r_{n,v}^d$ and the same overall bound applies. Moreover, the same lower bound for $\mu_d\bigl(B_y(r_{n,v}) \cap B_x(r_{n,u})^c\bigr)$ holds when $u < v$, $u \in [l_x, v_x]$, $x \in \mathcal{X}_n$, and $y = y_{x,z} \in B_x(r_{n,v_x} + r_{n,v_y})$. We deduce that for all $x \in \mathcal{X}_n$, $y = y_{x,z} \in B_x(r_{n,v_x} + r_{n,v_y})$, $u \in [l_x, v_x]$ and $v \in [l_y, v_y]$,

$$p_{n,x,u} + p_{n,y,v} - 2p_\cap \geq \max\{p_{n,x,u} - p_\cap\,,\, p_{n,y,v} - p_\cap\} \gtrsim \frac{k}{n}\|z\|. \tag{34}$$

We are now in a position to bound $\mathbb{E}(\|Z_3\|^3)$ above for $x \in \mathcal{X}_n$, $y = y_{x,z} \in B_x(r_{n,v_x} + r_{n,v_y})$, $u \in [l_x, v_x]$, $v \in [l_y, v_y]$. We write

$$\begin{aligned}\mathbb{E}(\|Z_3\|^3) = p_\cap \biggl\|V^{-1/2}\begin{pmatrix}1 - p_{n,x,u}\\ 1 - p_{n,y,v}\end{pmatrix}\biggr\|^3 + (p_{n,x,u} - p_\cap)\biggl\|V^{-1/2}\begin{pmatrix}1 - p_{n,x,u}\\ -p_{n,y,v}\end{pmatrix}\biggr\|^3 \\ + (p_{n,y,v} - p_\cap)\biggl\|V^{-1/2}\begin{pmatrix}-p_{n,x,u}\\ 1 - p_{n,y,v}\end{pmatrix}\biggr\|^3 \\ + (1 - p_{n,x,u} - p_{n,y,v} + p_\cap)\biggl\|V^{-1/2}\begin{pmatrix}p_{n,x,u}\\ p_{n,y,v}\end{pmatrix}\biggr\|^3,\end{aligned} \tag{35}$$

and bound each of these terms in turn. First,

$$\begin{aligned}&p_\cap \biggl\|V^{-1/2}\begin{pmatrix}1 - p_{n,x,u}\\ 1 - p_{n,y,v}\end{pmatrix}\biggr\|^3 \\ &= p_\cap |V|^{-3/2}\{(1-p_{n,x,u})(1-p_{n,y,v})(p_{n,x,u} + p_{n,y,v} - 2p_\cap)\}^{3/2} \\ &= p_\cap \Biggl\{\frac{(1-p_{n,x,u})(1-p_{n,y,v})}{p_\cap - p_{n,x,u}p_{n,y,v} + \frac{(p_{n,x,u}-p_\cap)(p_{n,y,v}-p_\cap)}{p_{n,x,u}+p_{n,y,v}-2p_\cap}}\Biggr\}^{3/2} \\ &\leq p_\cap \min\biggl\{\frac{p_{n,x,u}+p_{n,y,v}}{|V|}, \frac{1}{p_\cap - p_{n,x,u}p_{n,y,v}}\biggr\}^{3/2} \lesssim n^{1/2}/k^{1/2},\end{aligned} \tag{36}$$

using (32) and (33), and where we derive the final bound from the left hand side of the minimum if $\|z\| \geq 1$ and the right hand side if $\|z\| < 1$. Similarly,

$$(p_{n,x,u} - p_\cap)\biggl\|V^{-1/2}\begin{pmatrix}1 - p_{n,x,u}\\ -p_{n,y,v}\end{pmatrix}\biggr\|^3 \leq (p_{n,x,u} - p_\cap)p_{n,y,v}^{3/2}|V|^{-3/2} \lesssim \biggl(\frac{n}{k\|z\|}\biggr)^{1/2}, \tag{37}$$

where we have used (34) for the final bound. By symmetry, the same bound holds for the third term on the right-hand side of (35). Finally, very similar arguments yield

$$(1 - p_{n,x,u} - p_{n,y,v} + p_\cap)\left\| V^{-1/2}\begin{pmatrix} p_{n,x,u} \\ p_{n,y,v}\end{pmatrix}\right\|^3 \lesssim (k/n)^{3/2}. \tag{38}$$

Combining (36), (37) and (38) gives (30).

Writing $\mathbf{\Phi}_A(\cdot)$ for the measure associated with the $N_2(0,A)$ distribution for invertible A, and ϕ_A for the corresponding density, we have by Pinsker's inequality and a Taylor expansion of the log-determinant function that

$$\begin{aligned} 2\sup_{C\in\mathcal{C}} |\mathbf{\Phi}_A(C) - \mathbf{\Phi}_B(C)|^2 &\le \int_{\mathbb{R}^2} \phi_A \log\frac{\phi_A}{\phi_B} \\ &= \frac{1}{2}\{\log|B| - \log|A| + \mathrm{tr}(B^{-1}(A-B))\} \le \|B^{-1/2}(A-B)B^{-1/2}\|^2, \end{aligned}$$

provided $\|B^{-1/2}(A-B)B^{-1/2}\| \le 1/2$. Hence

$$\sup_{C\in\mathcal{C}} |\mathbf{\Phi}_A(C) - \mathbf{\Phi}_B(C)| \le \min\{1, 2\|B^{-1/2}(A-B)B^{-1/2}\|\}.$$

We now take $A = (n-2)V/k$, $B = \Sigma$ and use the submultiplicativity of the Frobenius norm along with (32) and (33) and the fact that $\|\Sigma^{-1/2}\| = \{(1+\alpha_z)^{-1} + (1-\alpha_z)^{-1}\}^{1/2}$ to deduce that

$$\sup_{C\in\mathcal{C}} |\mathbf{\Phi}_A(C) - \mathbf{\Phi}_B(C)| \lesssim \frac{1}{\|z\|}\left\{a(f(x)\wedge f(y))\left(\frac{k}{nf(x)}\right)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}\right\} \tag{39}$$

for $x \in \mathcal{X}_n$, $y \in B_x^\circ(r_{n,v_x} + r_{n,v_y})$, $u \in [l_x, v_x]$, $v \in [l_y, v_y]$. Now let $u = f(x)^{-1}(1 + k^{-1/2}s)$ and $v = f(x)^{-1}(1+k^{-1/2}t)$. By the mean value theorem, (23) and (31),

$$\begin{aligned} &\left|\Phi_\Sigma\left(k^{-1/2}\left\{(n-2)\mu - \begin{pmatrix} k \\ k\end{pmatrix}\right\}\right) - \Phi_\Sigma(s,t)\right| \\ &\qquad \le \frac{1}{(2\pi)^{1/2}}\left\{\left|\frac{(n-2)p_{n,x,u} - k}{k^{1/2}} - s\right| + \left|\frac{(n-2)p_{n,y,v} - k}{k^{1/2}} - t\right|\right\} \\ &\qquad \lesssim k^{1/2} a(f(x)\wedge f(y))\left(\frac{k}{nf(x)}\right)^{\beta/d} + k^{-1/2}. \end{aligned} \tag{40}$$

It follows by (29), (30), (39) and (40) that for $x \in \mathcal{X}_n$ and $y \in B_x^\circ(r_{n,v_x} +$

$r_{n,v_y})$,

$$\sup_{u\in[l_x,v_x],v\in[l_y,v_y]} |F^{(1)}_{n,x,y}(u,v) - \Phi_\Sigma(s,t)| \lesssim \min\biggl\{1, \frac{\log^{1/2} n}{k^{1/2}\|z\|} + a(f(x)\wedge(f(y))\biggl(\frac{k}{nf(x)}\biggr)^{\beta/d}\biggl(k^{1/2}+\frac{1}{\|z\|}\biggr)\biggr\}.$$

Therefore, by (23) and (24), and since $f(y) \geq f(x)/2$ for $x \in \mathcal{X}_n$, $y \in B_x(r_{n,v_x} + r_{n,v_y})$ and $n \geq n_0$, we conclude that for each $\epsilon > 0$ and $n \geq n_0$

$$\biggl|\int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_x}^{v_x}\int_{l_y}^{v_y} \frac{F^{(1)}_{n,x}(u,v)-\Phi_\Sigma(s,t)}{uv}\mathbb{1}_{\{\|x-y\|\leq r_{n,u}+r_{n,v}\}}\,du\,dv\,dy\,dx\biggr|$$
$$\lesssim \frac{k}{n}\int_{\mathcal{X}_n} f(x)\biggl\{\frac{\log^{1/2} n}{k^{1/2}} + a(f(x)/2)\biggl(\frac{k}{nf(x)}\biggr)^{\beta/d}\biggr\}^2 \int_{B_0(3)} \sup_{u\in[l_x,v_x],v\in[l_{y_{x,z}},v_{y_{x,z}}]} |F^{(1)}_{n,x,y_{x,z}}(u,v)-\Phi_\Sigma(s,t)|\,dz\,dx$$
$$= O\biggl(\frac{k}{n}\max\biggl\{\frac{\log^{5/2} n}{k^{3/2}}, \frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}, \frac{k^{-1/2+\beta/d}\log n}{n^{\beta/d}}, \frac{k^{1/2+2\beta/d}}{n^{2\beta/d}}\biggr\}\biggr). \tag{41}$$

By similar (in fact, rather simpler) means we can establish the same bound for the approximation of $G_{n,x,y}$ by $\Phi(k^{1/2}\{uf(x)-1\})\Phi(k^{1/2}\{vf(x)-1\})$.

To conclude the proof for the unweighted case, we write $\mathcal{X}_n = \mathcal{X}_n^{(1)} \cup \mathcal{X}_n^{(2)}$, where

$$\mathcal{X}_n^{(1)} := \{x : f(x) \geq k^{\frac{d}{2\beta}}\delta_n\}, \quad \mathcal{X}_n^{(2)} := \{x : \delta_n \leq f(x) < k^{\frac{d}{2\beta}}\delta_n\},$$

and deal with these two regions separately. We have by Slepian's inequality that $\Phi_\Sigma(s,t) \geq \Phi(s)\Phi(t)$ for all s and t. Hence, recalling that $s = s_{x,u} = k^{1/2}\{uf(x)-1\}$ and $t = t_{x,v} = k^{1/2}\{vf(x)-1\}$, by (21), (23) and (31), for every $\epsilon > 0$,

$$\int_{\mathcal{X}_n^{(2)}\times\mathcal{X}} f(x)f(y)\int_{l_x}^{v_x}\int_{l_y}^{v_y}\frac{\Phi_\Sigma(s,t)-\Phi(s)\Phi(t)}{uv}\mathbb{1}_{\{\|x-y\|\leq r_{n,u}+r_{n,v}\}}du\,dv\,dy\,dx$$
$$\leq \frac{e^{\Psi(k)}}{V_d(n-1)k}\int_{\mathcal{X}_n^{(2)}}\int_{\mathbb{R}^d} f(y_{x,z})\frac{\mathbb{1}_{\{\|x-y_{x,z}\|\leq r_{n,v_x}+r_{n,v_{y_{x,z}}}\}}}{f(x)^2 l_x l_{y_{x,z}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{\Phi_\Sigma(s,t)-\Phi(s)\Phi(t)\}\,ds\,dt\,dz\,dx$$
$$\lesssim \frac{1}{n}\int_{\mathcal{X}_n^{(2)}} f(x)\int_{B_0(2)}\alpha_z\,dz\,dx = o\biggl(\frac{k^{(1+\frac{d}{2\beta})\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}}\biggr), \tag{42}$$

where to obtain the final error term, we have used the fact that $\int_{B_0(2)} \alpha_z \, dz = V_d$. By (23) and (24) we have, for each $\epsilon > 0$,

$$\int_{\mathcal{X}_n^{(1)} \times \mathcal{X}} f(x) f(y) \int_{l_x}^{v_x} \int_{l_y}^{v_y} \frac{\Phi_\Sigma(s,t) - \Phi(s)\Phi(t)}{uv} \mathbb{1}_{\{\|x-y\| \leq r_{n,u} + r_{n,v}\}} \, du \, dv \, dy \, dx$$
$$\leq \frac{e^{\Psi(k)}}{V_d (n-1)k} \int_{\mathcal{X}_n^{(1)}} \int_{\mathbb{R}^d} f(y_{x,z}) \frac{\mathbb{1}_{\{\|x - y_{x,z}\| \leq r_{n,v_x} + r_{n,v_{y_{x,z}}}\}}}{f(x)^2 l_x l_{y_{x,z}}} \alpha_z \, dz \, dx$$
$$= \frac{e^{\Psi(k)}}{(n-1)k} \int_{\mathcal{X}_n^{(1)}} f(x) \, dx + O\biggl(\max\biggl\{ \frac{\log^{1/2} n}{nk^{1/2}}, \frac{k^{\beta/d}}{n^{1+\beta/d}}, \frac{k^{\frac{\alpha}{\alpha+d} - \epsilon}}{n^{1 + \frac{\alpha}{\alpha+d} - \epsilon}} \biggr\} \biggr)$$

(43)

$$= \frac{e^{\Psi(k)}}{(n-1)k} + O\biggl(\max\biggl\{ \frac{\log^{1/2} n}{nk^{1/2}}, \frac{k^{\beta/d}}{n^{1+\beta/d}}, \frac{k^{(1+\frac{d}{2\beta})\frac{\alpha}{\alpha+d} - \epsilon}}{n^{1 + \frac{\alpha}{\alpha+d} - \epsilon}} \biggr\} \biggr).$$

By applying Lemma 10(ii) in the main text as for (31) we have, for $x \in \mathcal{X}_n^{(1)}$ and $y \in B_x(r_{n,v_x} + r_{n,v_y})$, that

(44)

$$\max_{v \in \{v_x, v_y\}} |vf(x) - 1 - 3k^{-1/2} \log^{1/2} n| \lesssim a(f(x) \wedge f(y)) \Bigl(\frac{k}{nf(x)} \Bigr)^{\beta/d} = o(k^{-1/2}),$$

with similar bounds holding for l_x and l_y. A corresponding lower bound of the same order for the left-hand side of (43) follows from (44) and the fact that

$$\int_{-2\sqrt{\log n}}^{2\sqrt{\log n}} \int_{-2\sqrt{\log n}}^{2\sqrt{\log n}} \{\Phi_\Sigma(s,t) - \Phi(s)\Phi(t)\} \, ds \, dt = \alpha_z + O(n^{-2})$$

uniformly for $z \in \mathbb{R}^d$. It now follows from (28), (41), (42) and (43) that for each $\epsilon > 0$,

$$W_4 = O\biggl(\max\biggl\{ \frac{\log^{5/2} n}{nk^{1/2}}, \frac{k^{\frac{3}{2} + \frac{\alpha-\epsilon}{\alpha+d}}}{n^{1+\frac{\alpha-\epsilon}{\alpha+d}}}, \frac{k^{3/2 + 2\beta/d}}{n^{1+2\beta/d}}, \frac{k^{(1+\frac{d}{2\beta})\frac{\alpha-\epsilon}{\alpha+d}}}{n^{1+\frac{\alpha-\epsilon}{\alpha+d}}}, \frac{k^{\frac{1}{2}+\frac{\beta}{d}} \log n}{n^{1+\frac{\beta}{d}}} \biggr\} \biggr),$$

as required.

We now turn our attention to the variance of the weighted Kozachenko–Leonenko estimator $\hat{H}_n^w$. We first claim that

(45)

$$\mathrm{Var}\biggl(\sum_{j=1}^k w_j \log \xi_{(j),1} \biggr) = \sum_{j,l=1}^k w_j w_l \, \mathrm{Cov}(\log \xi_{(j),1}, \log \xi_{(l),1}) = V(f) + o(1).$$

By (18), (19) and Lemma 3 in the main text, for j such that $w_j \neq 0$,

$$\mathrm{Var} \log \xi_{(j),1} = V(f) + o(1)$$

as $n \to \infty$. For $l > j$, using similar arguments to those used in the proof of Lemma 3 in the main text, and writing $u_{x,s}^{(k)} := u_{x,s} = V_d(n-1)h_x^{-1}(s)^d e^{-\Psi(k)}$ for clarity, we have

$$\begin{aligned}
&\mathbb{E}(\log \xi_{(j),1} \log \xi_{(l),1}) \\
&= \int_{\mathcal{X}} f(x) \int_0^1 \int_0^{1-s} \log(u_{x,s}^{(j)}) \log(u_{x,s+t}^{(l)}) \mathrm{B}_{j,l-j,n-l}(s,t)\, dt\, ds\, dx \\
&= \int_{\mathcal{X}} f(x) \int_0^1 \int_0^{1-s} \log\Bigl(\frac{(n-1)s}{f(x)e^{\Psi(j)}}\Bigr) \log\Bigl(\frac{(n-1)(s+t)}{f(x)e^{\Psi(l)}}\Bigr) \mathrm{B}_{j,l-j,n-l}(s,t) dt\, ds\, dx + o(1) \\
&= \int_{\mathcal{X}} f(x) \log^2 f(x)\, dx + o(1)
\end{aligned}$$

as $n \to \infty$, uniformly for $1 \leq j < l \leq k_1^*$. Now (45) follows on noting that $\sup_{k \geq k_d} \|w\| < \infty$.

Next we claim that

$$\mathrm{Cov}\Bigl(\sum_{j=1}^k w_j \log \xi_{(j),1}, \sum_{l=1}^k w_l \log \xi_{(l),2}\Bigr) = o(n^{-1}) \tag{46}$$

as $n \to \infty$. In view of (20) in the main text and the fact that $\sup_{k \geq k_d} \|w\| < \infty$, it is sufficient to show that

$$\mathrm{Cov}\bigl(\log(f(X_1)\xi_{(j),1}), \log(f(X_2)\xi_{(l),2})\bigr) = o(n^{-1})$$

as $n \to \infty$, whenever $w_j, w_l \neq 0$. We suppose without loss of generality here that $j < l$, since the $j = l$ case is dealt with in (27). We broadly follow the same approach used to bound $W_1, \ldots, W_4$, though we require some new (similar) notation. Let $F'_{n,x,y}$ denote the conditional distribution function of $(\xi_{(j),1}, \xi_{(l),2})$ given $X_1 = x, X_2 = y$ and let $F_{n,x}^{(j)}$ denote the conditional distribution function of $\xi_{(j),1}$ given $X_1 = x$. Let

$$r_{n,u}^{(j)} := \Bigl\{\frac{ue^{\Psi(j)}}{V_d(n-1)}\Bigr\}^{1/d}, \quad p_{n,x,u}^{(j)} := h_x(r_{n,u}^{(j)}).$$

Recall the definitions of $a_{n,j}^{\pm}$ given in the proof of Lemma 3, and let $v_{x,j} := \inf\{u \geq 0 : (n-1)p_{n,x,u}^{(j)} = a_{n,j}^{+}\}$ and $l_{x,j} := \inf\{u \geq 0 : (n-1)p_{n,x,u}^{(j)} = a_{n,j}^{-}\}$.

For pairs (u,v) with $u \leq v_{x,j}$ and $v \leq v_{y,l}$, let $(M_1, M_2, M_3) \sim \text{Multi}(n-2; p^{(j)}_{n,x,u}, p^{(l)}_{n,y,v}, 1-p^{(j)}_{n,x,u}-p^{(l)}_{n,y,v})$ and write

$$G'_{n,x,y}(u,v) := \mathbb{P}(M_1 \geq j, M_2 \geq l).$$

Also write

$$\Sigma' := \begin{pmatrix} 1 & (j/l)^{1/2}\alpha'_z \\ (j/l)^{1/2}\alpha'_z & 1 \end{pmatrix},$$

where $\alpha'_z := V_d^{-1}\mu_d\big(B_0(1) \cap B_z(\exp(\Psi(l)-\Psi(j))^{1/d})\big)$. Writing W'_i for remainder terms to be bounded later, we have

$$\begin{aligned}
&\operatorname{Cov}\big(\log(f(X_1)\xi_{(j),1}), \log(f(X_2)\xi_{(l),2})\big) \\
&= \int_{\mathcal{X}\times\mathcal{X}} f(x)f(y)\int_{[l_{y,l},v_{y,l}]\times[l_{x,j},v_{x,j}]} h(u,v)\, d(F'_{n,x,y} - F^{(j)}_{n,x}F^{(l)}_{n,y})(u,v)\,dx\,dy + W'_1 \\
&= \int_{\mathcal{X}\times\mathcal{X}} f(x)f(y)\int_{[l_{y,l},v_{y,l}]\times[l_{x,j},v_{x,j}]} h(u,v)\, d(F'_{n,x,y} - G'_{n,x,y})(u,v)\,dx\,dy - \frac{1}{n} + \sum_{i=1}^{2} W'_i \\
&= \int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_{y,l}}^{v_{y,l}}\int_{l_{x,j}}^{v_{x,j}} \frac{(F'_{n,x,y} - G'_{n,x,y})(u,v)}{uv}\,du\,dv\,dx\,dy - \frac{1}{n} + \sum_{i=1}^{3} W'_i \\
&= \frac{V_d^{-1}e^{\Psi(j)}}{(n-1)(jl)^{1/2}}\int_{\mathbb{R}^d}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{\Phi_{\Sigma'}(s,t) - \Phi(s)\Phi(t)\}\,ds\,dt\,dz - \frac{1}{n} + \sum_{i=1}^{4} W'_i
\end{aligned}$$

(47)

$$= \frac{V_d^{-1}e^{\Psi(j)}}{(n-1)l}\int_{\mathbb{R}^d}\alpha'_z\,dz - \frac{1}{n} + \sum_{i=1}^{4} W_i = O\Big(\frac{1}{nk}\Big) + \sum_{i=1}^{4} W'_i$$

as $n \to \infty$. The final equality here follows from the fact that, for Borel measurable sets $K, L \subseteq \mathbb{R}^d$,

$$\int_{\mathbb{R}^d}\mu_d\big((K+z)\cap L\big)\,dz = \mu_d(K)\mu_d(L), \tag{48}$$

so that $\int_{\mathbb{R}^d}\alpha'_z\,dz = V_d e^{\Psi(l)-\Psi(j)}$.

To bound W'_1: Very similar arguments to those used to bound W_1 show that $W'_1 = o(n^{-(9/2-\epsilon)})$ as $n \to \infty$, for every $\epsilon > 0$.

To bound W_2': Similar to our work used to bound W_2, we may show that

$$\int_{\frac{a_{n,j}^-}{n-1}}^{\frac{a_{n,j}^+}{n-1}} \int_{\frac{a_{n,l}^-}{n-1}}^{\frac{a_{n,l}^+}{n-1}} |\mathrm{B}_{j+a,l+b,n-j-l-1}(s,t) - \mathrm{B}_{j+a,n-j}(s)\mathrm{B}_{l+b,n-l}(t)|\,dt\,ds \leq \frac{(jl)^{1/2}}{n}\{1+o(1)\}$$

as $n \to \infty$, for fixed $a, b > -1$. Also,

$$\int_0^1\int_0^{1-s} \log\Bigl(\frac{(n-1)s}{e^{\Psi(j)}}\Bigr)\log\Bigl(\frac{(n-1)t}{e^{\Psi(l)}}\Bigr)\{\mathrm{B}_{j,l,n-j-l-1}(s,t) - \mathrm{B}_{j,n-j}(s)\mathrm{B}_{l,n-l}(t)\}dtds = -\frac{1}{n} + O(n^{-2})$$

as $n \to \infty$. Using these facts and very similar arguments to those used to bound W_2 we have for every $\epsilon > 0$ that

$$W_2' = O\biggl(\frac{k^{1/2}}{n}\max\biggl\{\frac{k^{\beta/d}}{n^{\beta/d}}\,,\,\frac{k^{\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}\biggr\}\biggr).$$

To bound W_3': Similarly to (18) and the surrounding work, we can show that for every $\epsilon > 0$,

$$W_3' = O\biggl(\max\biggl\{\frac{\log n}{nk^{1/2}}\,,\,\frac{k^{\frac{1}{2}+\frac{2\beta}{d}}}{n^{1+\frac{2\beta}{d}}}\,,\,\frac{k^{\frac{2\alpha}{\alpha+d}-\epsilon}}{n^{\frac{2\alpha}{\alpha+d}-\epsilon}}\biggr\}\biggr).$$

To bound W_4': Let $(N_1, N_2, N_3, N_4) \sim \mathrm{Multi}(n-2; p_{n,x,u}^{(j)} - p_\cap, p_{n,y,v}^{(l)} - p_\cap, p_\cap, 1 - p_{n,x,u}^{(j)} - p_{n,y,v}^{(l)} + p_\cap)$, where $p_\cap := \int_{B_x(r_{n,u}^{(j)})\cap B_y(r_{n,v}^{(l)})} f(w)\,dw$. Further, let

$$F_{n,x,y}^{\prime,(1)} := \mathbb{P}(N_1 + N_3 \geq j, N_2 + N_3 \geq l).$$

Then, as in (28), we have

$$\begin{aligned}\int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_{x,j}}^{v_{x,j}}\int_{l_{y,l}}^{v_{y,l}} \frac{(F_{n,x,y}' - G_{n,x,y}')(u,v)}{uv}\,du\,dv\,dx\,dy \\ = \int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_{x,j}}^{v_{x,j}}\int_{l_{y,l}}^{v_{y,l}} \frac{(F_{n,x,y}^{\prime,(1)} - G_{n,x,y}')(u,v)}{uv}\,du\,dv\,dx\,dy \\ + O\biggl(\max\biggl\{\frac{\log n}{nk^{1/2}}\,,\,\frac{k^{\frac{1}{2}+\frac{2\beta}{d}}}{n^{1+\frac{2\beta}{d}}}\,,\,\frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{1+\frac{\alpha}{\alpha+d}-\epsilon}}\biggr\}\biggr).\end{aligned}$$

We can now approximate $F_{n,x,y}^{\prime,(1)}(u,v)$ by $\Phi_{\Sigma'}(j^{1/2}\{uf(x)-1\}, l^{1/2}\{vf(x)-1\})$ and $G'_{n,x,y}(u,v)$ by $\Phi(j^{1/2}\{uf(x)-1\})\Phi(l^{1/2}\{vf(x)-1\})$. This is rather similar to the corresponding approximation in the bounds on W_4, so we only present the main differences. First, let

$$Y'_i := \begin{pmatrix} \mathbb{1}_{\{X_i \in B_x(r_{n,u}^{(j)})\}} \\ \mathbb{1}_{\{X_i \in B_y(r_{n,v}^{(l)})\}} \end{pmatrix}.$$

We also define

$$\mu' := \mathbb{E}(Y'_i) = \begin{pmatrix} p_{n,x,u}^{(j)} \\ p_{n,y,v}^{(l)} \end{pmatrix}$$

and

$$V' := \mathrm{Cov}(Y'_i) = \begin{pmatrix} p_{n,x,u}^{(j)}(1-p_{n,x,u}^{(j)}) & p_\cap - p_{n,x,u}^{(j)}p_{n,y,v}^{(l)} \\ p_\cap - p_{n,x,u}^{(j)}p_{n,y,v}^{(l)} & p_{n,y,v}^{(l)}(1-p_{n,y,v}^{(l)}) \end{pmatrix},$$

and set $Z'_i := V'^{-1/2}(Y'_i - \mu)$. Our aim is to provide a bound on $p_\cap$. Since the function

$$(r,s) \mapsto \mu_d\big(B_0(r^{1/d}) \cap B_z(s^{1/d})\big),$$

is Lipschitz we have for $x \in \mathcal{X}_n, y = x + f(x)^{-1/d} r_{n,1}^{(j)} z \in B_x(r_{n,v_{x,j}}^{(j)} + r_{n,v_{y,l}}^{(l)}), u \in [l_{x,j}, v_{x,j}]$ and $v \in [l_{y,l}, v_{y,l}]$ that

$$\left| \frac{n-2}{e^{\Psi(j)}} p_\cap - \alpha'_z \right| \lesssim a(f(x) \wedge (f(y)) \left(\frac{k}{nf(x)} \right)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}, \tag{49}$$

using similar equations to (23), (24) and (31). From this and similar bounds to (32), we find that $|V'| \gtrsim k^2/n^2$ and $\|(V')^{-1/2}\| \lesssim (n/k)^{1/2}$. We therefore have

$$\mathbb{E}\|Z'_3\|^3 \leq \|(V')^{-1/2}\|^3 \mathbb{E}\|Y'_3 - \mu'\|^3 \lesssim n^{1/2}/k^{1/2},$$

which is as in the $l = j$ case except with the factor of $\|z\|^{-1/2}$ missing. Note now that

$$\limsup_{n\to\infty} \sup_{\substack{(j,l):j<l \\ w_j,w_l \neq 0}} \sup_{z \in B_0(1+e^{(\Psi(l)-\Psi(j))/d})} \|(\Sigma')^{-1/2}\| < \infty.$$

Hence, using (49), similar bounds to (32) and the same arguments as leading up to (39),

$$\sup_{C \in \mathcal{C}} |\mathbf{\Phi}_A(C) - \mathbf{\Phi}_B(C)| \lesssim a(f(x) \wedge f(y)) \left(\frac{k}{nf(x)} \right)^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}}, \tag{50}$$

where $B := \Sigma'$ and

$$A := (n-2)\begin{pmatrix} j^{-1}p_{n,x,u}^{(j)}(1-p_{n,x,u}^{(j)}) & j^{-1/2}l^{-1/2}(p_\cap - p_{n,x,u}^{(j)}p_{n,y,v}^{(l)}) \\ j^{-1/2}l^{-1/2}(p_\cap - p_{n,x,u}^{(j)}p_{n,y,v}^{(l)}) & l^{-1}p_{n,y,v}^{(l)}(1-p_{n,y,v}^{(l)}) \end{pmatrix}.$$

Now let $u := f(x)^{-1}(1+j^{-1/2}s)$ and $v := f(x)^{-1}(1+l^{-1/2}t)$. Similarly to (40), we have

$$\begin{aligned}\Bigl|\Phi_{\Sigma'}\Bigl(\frac{(n-2)p_{n,x,u}^{(j)}-j}{j^{1/2}},\frac{(n-2)p_{n,y,v}^{(l)}-l}{l^{1/2}}\Bigr) - \Phi_{\Sigma'}(s,t)\Bigr| \\ \lesssim k^{1/2}a(f(x)\wedge f(y))\Bigl(\frac{k}{nf(x)}\Bigr)^{\beta/d} + k^{-1/2}.\end{aligned}$$

Similarly to the arguments leading up to (41), it follows that

$$\begin{aligned}&\Bigl|\int_{\mathcal{X}_n\times\mathcal{X}} f(x)f(y)\int_{l_{x,j}}^{v_{x,j}}\int_{l_{y,l}}^{v_{y,l}} \frac{F_{n,x}^{',(1)}(u,v)-\Phi_{\Sigma'}(s,t)}{uv}\mathbb{1}_{\{\|x-y\|\leq r_{n,u}^{(j)}+r_{n,v}^{(l)}\}}\,du\,dv\,dy\,dx\Bigr| \\ &= O\biggl(\frac{k}{n}\max\biggl\{\frac{\log^{3/2} n}{k^{3/2}}, \frac{k^{\frac{1}{2}+\frac{\alpha}{\alpha+d}-\epsilon}}{n^{\frac{\alpha}{\alpha+d}-\epsilon}}, \frac{k^{-1/2+\beta/d}\log n}{n^{\beta/d}}, \frac{k^{1/2+2\beta/d}}{n^{2\beta/d}}\biggr\}\biggr),\end{aligned}$$

where the power on the first logarithmic factor is smaller because of the absence of the factor of the $\|z\|^{-1}$ term in (50). The remainder of the work required to bound W_4' is very similar to the work done from (42) to (43), using also (48), so is omitted. We conclude that

$$W_4' = O\biggl(\max\biggl\{\frac{\log^{\frac{3}{2}} n}{nk^{\frac{1}{2}}}, \frac{k^{\frac{3}{2}+\frac{\alpha-\epsilon}{\alpha+d}}}{n^{1+\frac{\alpha-\epsilon}{\alpha+d}}}, \frac{k^{\frac{3}{2}+\frac{2\beta}{d}}}{n^{1+\frac{2\beta}{d}}}, \frac{k^{(1+\frac{d}{2\beta})\frac{\alpha-\epsilon}{\alpha+d}}}{n^{1+\frac{\alpha-\epsilon}{\alpha+d}}}, \frac{k^{\frac{1}{2}+\frac{\beta}{d}}\log n}{n^{1+\frac{\beta}{d}}}\biggr\}\biggr).$$

The equation (47), together the bounds on $W_1', \ldots, W_4'$ just proved, establish the claim (46). We finally conclude from (45) and (46) that

$$\begin{aligned}\mathrm{Var}(\hat{H}_n^w) &= \frac{1}{n}\mathrm{Var}\biggl(\sum_{j=1}^k w_j \log \xi_{(j),1}\biggr) \\ &\qquad + \Bigl(1-\frac{1}{n}\Bigr)\mathrm{Cov}\biggl(\sum_{j=1}^k w_j\log\xi_{(j),1}, \sum_{l=1}^k w_l \log\xi_{(l),2}\biggr) \\ &= V(f) + o(n^{-1}),\end{aligned}$$

as required.

A. Proof of Theorem 8.

PROOF OF THEOREM 8. For the first part of the theorem we aim to apply Theorem 25.21 of van der Vaart (1998), and follow the notation used there. With $\dot{\mathcal{P}} := \{\lambda(\log f + H(f)) : \lambda \in \mathbb{R}\}$ we will first show that the entropy functional H is differentiable at f relative to the tangent set $\dot{\mathcal{P}}$, with efficient influence function $\tilde{\psi}_f = -\log f - H(f)$. Following Example 25.16 in van der Vaart (1998), for $g \in \dot{\mathcal{P}}$, the paths $f_{t,g}$ defined in (10) of the main text are differentiable in quadratic mean at $t = 0$ with score function g. Note that $\int_{\mathcal{X}} gf = 0$ and $\int_{\mathcal{X}} g^2 f < \infty$ for all $g \in \dot{\mathcal{P}}$. It is convenient to define, for $t \geq 0$, the set $A_t := \{x \in \mathcal{X} : 8t|g(x)| \leq 1\}$, on which we may expand e^{-2tg} easily as a Taylor series. By Hölder's inequality, for $\epsilon \in (0, 1/2)$,

$$\begin{aligned}\int_{A_t^c} f|\log f| &\leq (8t)^{2(1-\epsilon)} \int_{\mathcal{X}} f|g|^{2(1-\epsilon)}|\log f| \\ &\leq (8t)^{2(1-\epsilon)} \biggl\{\int_{\mathcal{X}} g^2 f\biggr\}^{1-\epsilon} \biggl\{\int_{\mathcal{X}} f|\log f|^{1/\epsilon}\biggr\}^{\epsilon} = o(t)\end{aligned}$$

as $t \searrow 0$. Moreover,

$$\int_{A_t^c} f\log(1 + e^{-2tg}) \leq \int_{A_t^c} (\log 2 + 2t|g|)f \leq 16t^2(4\log 2 + 1)\int_{\mathcal{X}} g^2 f.$$

We also have that

$$\begin{aligned}|c(t)^{-1} - 1| &= \biggl|\int_{\mathcal{X}} \biggl(\frac{2}{1 + e^{-2tg}} - 1 - tg\biggr) f\biggr| \\ &\leq \int_{A_t} \biggl|\frac{e^{-2tg} - 1 + 2tg + tg(e^{-2tg} - 1)}{1 + e^{-2tg}}\biggr| f + \int_{A_t^c} (1 + t|g|)f \\ &\leq \frac{16}{3}t^2 \int_{A_t} g^2 f + 72t^2 \int_{A_t^c} g^2 f \leq 72t^2 \int_{\mathcal{X}} g^2 f. \qquad (51)\end{aligned}$$

It follows that

$$\begin{aligned}&\biggl|t^{-1}\{H(f_{t,g}) - H(f)\} + \int_{\mathcal{X}} \{\log f + H(f)\}fg\biggr| \\ &= \biggl|\frac{1}{t}\int_{\mathcal{X}} \biggl\{\biggl(1 - \frac{2c(t)}{1+e^{-2tg}}\biggr)\log f - \frac{2c(t)}{1+e^{-2tg}} \log\biggl(\frac{2c(t)}{1+e^{-2tg}}\biggr) + tg(1 + \log f)\biggr\} f\biggr| \\ &\leq \frac{1}{t}\int_{A_t} f\biggl|\{e^{-2tg} - 1 + 2tg + tg(e^{-2tg} - 1)\}\log f \\ &\qquad\qquad - 2\log\biggl(\frac{2}{1 + e^{-2tg}}\biggr) + tg(1 + e^{-2tg})\biggr| + o(1) \\ &\leq \frac{16}{3}t\int_{\mathcal{X}} g^2 f|\log f| + 22t\int_{\mathcal{X}} g^2 f + o(1) \to 0.\end{aligned}$$

The conclusion (11) in the main text therefore follows from van der Vaart (1998, Theorem 25.21).

We now establish the second part of the theorem. First, by our previous bound on $c(t)$ in (51), for $12t < \{\int_{\mathcal{X}} g^2 f\}^{-1/2}$ we have that

$$\|f_{t,g}\|_\infty \leq 2c(t)\|f\|_\infty \leq \frac{2\|f\|_\infty}{1 - 72t^2 \int_{\mathcal{X}} g^2 f} \leq 4\|f\|_\infty,$$

and $\mu_\alpha(f_{t,g}) \leq 4\mu_\alpha(f)$.

We now study the smoothness properties of $f_{t,g}$. This requires some involved calculations, because we first need to understand corresponding properties of g. To this end, for an m times differentiable function $g : \mathbb{R}^d \to \mathbb{R}$, define

$$M_g^*(x) := \max\biggl\{\max_{t=1,\ldots,m} \|g^{(t)}(x)\|\,, \sup_{y \in B_x^\circ(r_a(x))} \frac{\|g^{(m)}(y) - g^{(m)}(x)\|}{\|y - x\|^{\beta - m}}\biggr\}$$

and

$$D_g := \max\biggl\{1, \sup_{\delta \in (0,\|f\|_\infty)} \frac{\sup_{x: f(x) \geq \delta} M_g^*(x)}{a(\delta)^{m+1}}\biggr\}.$$

Let $\mathcal{J}_m$ denote the set of multisets of elements $\{1,\ldots,d\}$ of cardinality at most m, and for $J = \{j_1,\ldots,j_s\} \in \mathcal{J}_m$, define $g_J(x) := \frac{\partial^s g}{\prod_{\ell=1}^s \partial x_{j_\ell}}(x)$. Moreover, for $i \in \{1,\ldots,s\}$, let $\mathcal{P}_i(J)$ denote the set of partitions of J into i non-empty multisets. As an illustration, if $d = 2$, then

$$\begin{aligned}\mathcal{J}_3 = \bigl\{&\emptyset, \{1\}, \{2\}, \{1,1\}, \{1,2\}, \{2,1\}, \{2,2\}, \\ &\{1,1,1\},\{1,1,2\},\{1,2,1\},\{1,2,2\},\{2,1,1\},\{2,1,2\},\{2,2,1\},\{2,2,2\}\bigr\}.\end{aligned}$$

Moreover, if $J = \{1,1,2\} \in \mathcal{J}_3$, then

$$\mathcal{P}_2(J) = \Bigl\{\bigl\{\{1,1\},\{2\}\bigr\}, \bigl\{\{1,2\},\{1\}\bigr\}, \bigl\{\{1,2\},\{1\}\bigr\}\Bigr\}.$$

Then, by induction, and writing $g^* := g_1 = \log f + H(f)$, it may be shown that

$$g_J^*(x) = \sum_{i=1}^{\mathrm{card}(J)} \frac{(-1)^{i-1}(i-1)!}{f^i} \sum_{\{P_1,\ldots,P_i\} \in \mathcal{P}_i(J)} f_{P_1} \ldots f_{P_i}.$$

Now, the cardinality of $\mathcal{P}_i(J)$ is given by a Stirling's number of the second kind:

$$\mathrm{card}\bigl(\mathcal{P}_i(J)\bigr) = \frac{1}{i!}\sum_{\ell=0}^{i}(-1)^{i-\ell}\binom{i}{\ell}\ell^{\mathrm{card}(J)} =: S\bigl(\mathrm{card}(J), i\bigr),$$

say. Thus, if $\mathrm{card}(J) \leq m$, then

$$|g_J^*(x)| \leq \sum_{i=1}^{\mathrm{card}(J)} (i-1)! S\bigl(\mathrm{card}(J), i\bigr) a(f(x))^i \leq \frac{1}{2} m^{m+1} m! a(f(x))^m. \tag{52}$$

Moreover, if $\|y - x\| \leq r_a(x)$ and $m \geq 1$, then

$$|g_J^*(y) - g_J^*(x)| \leq \sum_{i=1}^{\mathrm{card}(J)} (i-1)! \sum_{\{P_1,\ldots,P_i\} \in \mathcal{P}_i(J)} \biggl\{ \frac{|f_{P_1} \ldots f_{P_i}(y) - f_{P_1} \ldots f_{P_i}(x)|}{f^i(y)} + \frac{|f_{P_1} \ldots f_{P_i}(x)|}{f^i(y)} \biggl| \frac{f^i(y)}{f^i(x)} - 1 \biggr| \biggr\}.$$

Now, by Lemma 2,

$$\biggl| \frac{f^i(y)}{f^i(x)} - 1 \biggr| \leq i \biggl| \frac{f(y)}{f(x)} - 1 \biggr| \biggl(1 + \biggl| \frac{f(y)}{f(x)} - 1 \biggr| \biggr)^{i-1} \leq \Bigl(\frac{71}{56} \Bigr)^{i-1} i \biggl| \frac{f(y)}{f(x)} - 1 \biggr|.$$

Moreover, by induction and Lemma 2 again,

$$|f_{P_1} \ldots f_{P_i}(y) - f_{P_1} \ldots f_{P_i}(x)| \leq 8d^{1/2} \biggl\{ \Bigl(\frac{71}{56} \Bigr)^i - 1 \biggr\} a(f(x))^i f^i(x) \|y - x\|^{\beta - m}.$$

We deduce that (even when $m = 0$),

$$|g_J^*(y) - g_J^*(x)| \leq 8d^{1/2} \Bigl(\frac{71}{41} \Bigr)^m m! (m+1)^{m+2} a(f(x))^{m+1} \|y - x\|^{\beta - m}. \tag{53}$$

Comparing (52) and (53), we see that

$$D_{g^*} \leq 8d^{1/2} \Bigl(\frac{71}{41} \Bigr)^m m! (m+1)^{m+2} =: D. \tag{54}$$

Now let $q(y) := (1 + e^{-2ty})^{-1}$, so that $f_{t,g}(x) = 2c(t) q\bigl(g(x)\bigr) f(x)$. Similar inductive arguments to those used above yield that when $J \in \mathcal{J}_m$ with $m \geq 1$ and g is m times differentiable,

$$(q \circ g)_J(x) = \sum_{i=1}^{\mathrm{card}(J)} q^{(i)}\bigl(g(x)\bigr) \sum_{\{P_1,\ldots,P_i\} \in \mathcal{P}_i(J)} g_{P_1} \ldots g_{P_i}(x),$$

and we now bound the derivatives of q. By induction,

$$q^{(i)}(y) = (2t)^i \sum_{\ell=1}^{i} (-1)^{i-\ell} \frac{a_\ell^{(i)} e^{-2t\ell y}}{(1 + e^{-2ty})^{\ell+1}},$$

where for each $i \in \mathbb{N}$, we have $a_1^{(i)} = 1$, $a_i^{(i)} = i!$ and $a_\ell^{(i)} = \ell(a_\ell^{(i-1)} + a_{\ell-1}^{(i-1)})$ for $\ell \in \{2, \ldots, i-1\}$. Since $\max_{1 \leq \ell \leq i} a_\ell^{(i)} \leq (2i)^{i-1}$ (again by induction), we deduce that

$$(1 + e^{-2ty})|q^{(i)}(y)| \leq 2^{2i-1} i^i t^i. \tag{55}$$

Writing $s := \mathrm{card}(J)$, it follows that

$$\begin{aligned} |(q \circ g)_J(x)| &\leq q\bigl(g(x)\bigr) \sum_{i=1}^{s} 2^{2i-1} i^i t^i S(s,i) a(f(x))^{i(m+1)} D_g^i \\ &\leq q\bigl(g(x)\bigr) s^{s+1} 2^{2s-1} \max(1,t)^s B_s a(f(x))^{s(m+1)} D_g^s, \end{aligned} \tag{56}$$

where $B_s := \sum_{i=1}^{s} S(s,i)$ denotes the sth Bell number. We can now apply the multivariate Leibniz rule, so that for a multi-index $\omega = (\omega_1, \ldots, \omega_d)$ with $|\omega| \leq m$, and for $t \leq 1$ and $m \geq 1$,

$$\begin{aligned} \left| \frac{\partial^\omega f_{t,g^*}(x)}{\partial x^\omega} \right| &= \left| 2c(t) \sum_{\nu : \nu \leq \omega} \binom{\omega}{\nu} \frac{\partial^\nu q\bigl(g^*(x)\bigr)}{\partial x^\nu} \frac{\partial^{\omega - \nu} f(x)}{\partial x^{\omega - \nu}} \right| \\ &\leq 2^{3m-1} m^{m+1} B_m D_{g^*}^m a(f(x))^{m^2+m} f_{t,g^*}(x). \end{aligned} \tag{57}$$

Now, in order to control $\bigl| \frac{\partial^\omega f_{t,g^*}(y)}{\partial x^\omega} - \frac{\partial^\omega f_{t,g^*}(x)}{\partial x^\omega} \bigr|$, we first note that by (53) and (54), we have for $\|y - x\| \leq r_a(x)$, $i \in \mathbb{N}$, $J \in \mathcal{J}_m$ with $\mathrm{card}(J) = s$ and $\{P_1, \ldots, P_i\} \in \mathcal{P}_i(J)$,

$$|g_{P_1}^* \ldots g_{P_i}^*(y) - g_{P_1}^* \ldots g_{P_i}^*(x)| \leq (2D)^i a(f(x))^{i(m+1)} \|y - x\|^{\beta - m}. \tag{58}$$

Thus, by (55), (58), the mean value theorem and Lemma 2, for $t \leq 1$, $\|y - x\| \leq r_a(x)$ and $m \geq 1$,

$$\begin{aligned} &|(q \circ g^*)_J(y) - (q \circ g^*)_J(x)| \\ &\leq \left| \sum_{i=1}^{s} q^{(i)}(g^*(x)) \sum_{\{P_1,\ldots,P_i\} \in \mathcal{P}_i(J)} \{g_{P_1}^* \ldots g_{P_i}^*(y) - g_{P_1}^* \ldots g_{P_i}^*(x)\} \right| \\ &\quad + \left| \sum_{i=1}^{s} \{q^{(i)}(g^*(y)) - q^{(i)}(g^*(x))\} \sum_{\{P_1,\ldots,P_i\} \in \mathcal{P}_i(J)} g_{P_1}^* \ldots g_{P_i}^*(y) \right| \\ &\leq D^m q(g^*(x)) a(f(x))^{m^2+m+1} \|y - x\|^{\beta - m} \\ &\qquad \times \frac{B_m 2^{3m+5} d^{1/2} (m+1)^{m+1} (1 + e^{2tg^*(x)})}{e^{2tg^*(x)} + e^{-2t|g^*(y) - g^*(x)|}} \\ &\leq D^m q(g^*(x)) a(f(x))^{m^2+m+1} \|y - x\|^{\beta - m} B_m 2^{3m+5} d^{1/2} (m+1)^{m+1} \Bigl(\frac{56}{41}\Bigr)^{2t}. \end{aligned} \tag{59}$$

Here, to obtain the final inequality, we used the fact that $\frac{1+a}{b+a} \leq \frac{1}{b}$ for $a, b > 0$ and $b < 1$, and the fact that

$$e^{2t|g^*(y)-g^*(x)|} = \max\biggl\{\Bigl(\frac{f(y)}{f(x)}\Bigr)^{2t}, \Bigl(\frac{f(x)}{f(y)}\Bigr)^{2t}\biggr\} \lesssim \bigl\{1+a(f(x))\|y-x\|^{\beta\wedge 1}\bigr\}^{2t} \lesssim 1$$

for $\|y-x\| \leq r_a(x)$. Using the multivariate Leibniz rule again, together with (56), (59) and Lemma 2, for $t \leq 1$, $\|y-x\| \leq r_a(x)$ and $|\omega| = m \geq 1$,

$$\begin{aligned}
&\left|\frac{\partial^\omega f_{t,g^*}(y)}{\partial x^\omega} - \frac{\partial^\omega f_{t,g^*}(x)}{\partial x^\omega}\right| \\
&\leq 2c(t) \sum_{\nu:\nu\leq\omega} \binom{\omega}{\nu} \biggl\{ \left|\frac{\partial^{\omega-\nu} f(y)}{\partial y^{\omega-\nu}}\right| \left|\frac{\partial^\nu q(g^*(y))}{\partial x^\nu} - \frac{\partial^\nu q(g^*(x))}{\partial x^\nu}\right| \\
&\qquad + \left|\frac{\partial^\nu q(g^*(x))}{\partial x^\nu}\right| \left|\frac{\partial^\nu f(y)}{\partial x^\nu} - \frac{\partial^\nu f(x)}{\partial x^\nu}\right| \biggr\} \\
&\leq 2^{4m+9} d^{1/2} B_m (m+1)^{m+1} D^m a(f(x))^{m^2+m+1} f_{t,g^*}(x)\|y-x\|^{\beta-m} \\
&=: C'_m D^m a(f(x))^{m^2+m+1} f_{t,g^*}(x)\|y-x\|^{\beta-m}. \qquad (60)
\end{aligned}$$

This also holds in the case $m = 0$. Now note that if $12t < \{\int_{\mathcal{X}} (g^*)^2 f\}^{-1/2}$ we have

$$f(x) = \frac{1+e^{-2tg^*(x)}}{2c(t)} f_{t,g^*}(x) \geq \frac{f_{t,g^*}(x)}{4}.$$

Finally, define the function

$$\tilde{a}(\delta) := d^{m/2} C'_m D^m a(\delta/4)^{m^2+m+1}. \qquad (61)$$

Then $\tilde{a} \in \mathcal{A}$ and from (57) and (60), we have $M_{f_{t,g^*},\tilde{a},\beta}(x) \leq \tilde{a}(f_{t,g^*}(x))$. We conclude that for $t < \min\bigl(1, \{144 \int g^2 f\}^{-1/2}\bigr)$, we have that $f_{t,g^*} \in \mathcal{F}_{d,\theta'}$, where $\theta' = (\alpha, \beta, 4\gamma, 4\nu, \tilde{a}) \in \Theta$. The result follows on noting that $f_{t,g_\lambda} = f_{t\lambda,g^*}$. □

References.

Berrett, T. B., Samworth, R. J. and Yuan, M. (2017) Efficient multivariate entropy estimation via k-nearest neighbour distances. *Submitted*.

Götze, F. (1991) On the rate of convergence in the multivariate CLT. *Ann. Prob.*, **19**, 724–739.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Statistical Laboratory
Wilberforce Road
Cambridge
CB3 0WB
United Kingdom
E-mail: r.samworth@statslab.cam.ac.uk
E-mail: t.berrett@statslab.cam.ac.uk
URL: http://www.statslab.cam.ac.uk/~rjs57
URL: http://www.statslab.cam.ac.uk/~tbb26

Department of Statistics
University of Wisconsin–Madison
Medical Sciences Center
1300 University Avenue
Madison, WI 53706
United States of America
E-mail: myuan@stat.wisc.edu
URL: http://pages.stat.wisc.edu/~myuan/