Supplementary material for: Ensemble Estimation of Multivariate f-Divergence

Kevin R. Moon
Department of EECS
University of Michigan
Ann Arbor, MI
krmoon@umich.edu

Alfred O. Hero III
Department of EECS
University of Michigan
Ann Arbor, MI
hero@eecs.umich.edu

1 Assumptions

We use the same assumptions on the densities and the functional as in [1] and [2]. They are

- $(\mathcal{A}.0)$: Assume that $k_i = k_0 M_i^\beta$ with $0 < \beta < 1$, that $M_2 = \alpha_{frac} T$ with $0 < \alpha_{frac} < 1$.
- $(\mathcal{A}.1)$: Assume there exist constants $\epsilon_0, \epsilon_\infty$ such that $0 < \epsilon_0 \leq f_i(x) \leq \epsilon_\infty < \infty$, $\forall x \in S$.
- $(\mathcal{A}.2)$: Assume that the densities f_i have continuous partial derivatives of order d in the interior of $\mathcal{S}$ that are upper bounded.
- $(\mathcal{A}.3)$: Assume that g has derivatives $g^{(j)}$ of order $j = 1, \dots, \max\{\lambda, d\}$ where $\lambda\beta > 1$.
- $(\mathcal{A}.4)$: Assume that $\left|g^{(j)}\left(f_1(x)/f_2(x)\right)\right|$, $j = 0, \dots, \max\{\lambda, d\}$ are strictly upper bounded for $\epsilon_0 \leq f_i(x) \leq \epsilon_\infty$.
- $(\mathcal{A}.5)$: Let $\epsilon \in (0,1)$, $\delta \in (2/3, 1)$, and $\mathcal{C}(k) = \exp\left(-3k^{(1-\delta)}\right)$. For fixed ϵ, define $p_{l,i} = (1-\epsilon)\epsilon_0 \frac{k_i - 1}{M_i}$, $p_{u,i} = (1+\epsilon)\epsilon_\infty \frac{k_i-1}{M_i}$, $q_{l,i} = \frac{k_i - 1}{M_i \bar{c} D^d}$, and $q_{u,i} = (1+\epsilon)\epsilon_\infty$ where D is the diameter of the support S. Let $\mathbf{P}_i$ be a beta distributed random variable with parameters k_i and $M_i - k_i + 1$. Define $p_l = \frac{p_{l,1}}{p_{u,2}}$ and $p_u = \frac{p_{u,1}}{p_{l,2}}$. Assume that for $U(L) = g(L)$, $g^{(3)}(L)$, and $g^{(\lambda)}(L)$,
  - (i) $\mathbb{E}\left[\sup_{L \in (p_l, p_u)} \left|U\left(L \frac{\mathbf{P}_2}{\mathbf{P}_1}\right)\right|\right] = G_1 < \infty$,
  - (ii) $\sup_{L \in \left(\frac{q_{l,1}}{q_{u,2}}, \frac{q_{u,1}}{q_{l,2}}\right)} \left|U(L)\right| \mathcal{C}(k_1) \mathcal{C}(k_2) = G_2 < \infty$,
  - (iii) $\mathbb{E}\left[\sup_{L \in \left(\frac{q_{l,1}}{p_{u,2}}, \frac{q_{u,1}}{p_{l,2}}\right)} \left|U(L\mathbf{P}_2)\right| \mathcal{C}(k_1)\right] = G_3 < \infty$,
  - (iv) $\mathbb{E}\left[\sup_{L \in \left(\frac{p_{l,1}}{q_{u,2}}, \frac{p_{u,1}}{q_{l,2}}\right)} \left|U\left(\frac{L}{\mathbf{P}_1}\right)\right| \mathcal{C}(k_2)\right] = G_4 < \infty$, $\forall M_i$.

Densities for which assumptions $(\mathcal{A}.0) - (\mathcal{A}.5)$ hold include the truncated Gaussian distribution and the Beta distribution on the unit cube. Functions for which the assumptions hold include $g(L) = -\ln L$ and $g(L) = L^\alpha$.

2 Proof of Theorem 2

We use the following lemma which is proved in [3]:

Lemma 1. *Let the random variables $\{\mathbf{Y}_{M,i}\}_{i=1}^N$ belong to a zero mean, unit variance, interchangeable process for all values of M. Assume that $Cov(\mathbf{Y}_{M,1}, \mathbf{Y}_{M,2})$ and $Cov(\mathbf{Y}_{M,1}^2, \mathbf{Y}_{M,2}^2)$ are*

$O(1/M)$. *Then the random variable*

$$\mathbf{S}_{N,M} = \frac{\sum_{i=1}^{N} \mathbf{Y}_{M,i}}{\sqrt{\mathbb{V}\left[\sum_{i=1}^{N} \mathbf{Y}_{M,i}\right]}} \tag{1}$$

converges in distribution to a standard normal random variable.

For simplicity, let $M_1 = M_2 = M$ and $\hat{\mathbf{L}}_{k(l)} := \hat{\mathbf{L}}_{k(l),k(l)}$. Define

$$\mathbf{Y}_{M,i} = \frac{\sum_{l \in \bar{l}} w(l) g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right) - \mathbb{E}\left[\sum_{l \in \bar{l}} w(l) g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right]}{\sqrt{\mathbb{V}\left[\sum_{l \in \bar{l}} w(l) g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right]}}.$$

Then from Eq. 1, we have that

$$\mathbf{S}_{N,M} = \frac{\hat{\mathbf{G}}_w - \mathbb{E}\left[\hat{\mathbf{G}}_w\right]}{\sqrt{\mathbb{V}\left[\hat{\mathbf{G}}_w\right]}}.$$

Thus it is sufficient to show from Lemma 1 that $Cov(\mathbf{Y}_{M,1}, \mathbf{Y}_{M,2})$ and $Cov(\mathbf{Y}_{M,1}^2, \mathbf{Y}_{M,2}^2)$ are $O(1/M)$. To do this, it is necessary to show that the denominator of $\mathbf{Y}_{M,i}$ converges to a nonzero constant or to zero sufficiently slowly. Note that the numerator and denominator of $\mathbf{Y}_{M,i}$ are, respectively,

$$\begin{aligned} &\sum_{l \in \bar{l}} w(l) g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right) - \mathbb{E}\left[\sum_{l \in \bar{l}} w(l) g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right] \\ &\quad = \sum_{l \in \bar{l}} w(l) \left(g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right) - \mathbb{E}\left[g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right]\right), \end{aligned} \tag{2}$$

$$\begin{aligned} &\sqrt{\mathbb{V}\left[\sum_{l \in \bar{l}} w(l) g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right]} \\ &= \sqrt{\sum_{l \in \bar{l}} \sum_{l' \in \bar{l}} w(l) w(l') Cov\left(g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right), g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_i)\right)\right)}. \end{aligned} \tag{3}$$

Therefore, to bound $Cov(\mathbf{Y}_{M,1}, \mathbf{Y}_{M,2})$, we require bounds on the quantity $Cov\left[g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right), g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_j)\right)\right]$.

Some preliminary work is required before we can directly tackle this quantity. Define $\mathcal{M}(\mathbf{Z}) := \mathbf{Z} - \mathbb{E}\mathbf{Z}$, $\hat{\mathbf{F}}_{k(l)}(\mathbf{Z}) := \hat{\mathbf{L}}_{k(l)}(\mathbf{Z}) - \mathbb{E}_{\mathbf{Z}}\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})\right)$, and $\hat{\mathbf{e}}_{i,k(l)}(\mathbf{Z}) := \hat{\mathbf{f}}_{i,k(l)}(\mathbf{Z}) - \mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{i,k(l)}(\mathbf{Z})$. By forming a Taylor series expansion of $g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})\right)$ around $\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})$, we get

$$g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})\right) = \sum_{i=0}^{\lambda-1} \frac{g^{(i)}\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})\right)}{i!} \hat{\mathbf{F}}_{k(l)}^{i}(\mathbf{Z}) + \frac{g^{(\lambda)}\left(\xi_{\mathbf{Z}}\right)}{\lambda!} \hat{\mathbf{F}}_{k(l)}^{\lambda}(\mathbf{Z}),$$

where $\xi_{\mathbf{Z}} \in \left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{F}}_{k(l)}(\mathbf{Z}), \hat{\mathbf{F}}_{k(l)}(\mathbf{Z})\right)$. Let $\Psi(\mathbf{Z}) = g^{(\lambda)}\left(\xi_{\mathbf{Z}}\right)/\lambda!$ and

$$\begin{aligned} \mathbf{p}_i^{(l)} &:= \mathcal{M}\left(g\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right), \\ \mathbf{q}_i^{(l)} &:= \mathcal{M}\left(g'\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}\left(\mathbf{X}_i\right)\right)\hat{\mathbf{F}}_{k(l)}\left(\mathbf{X}_i\right)\right), \\ \mathbf{r}_i^{(l)} &:= \mathcal{M}\left(\sum_{j=2}^{\lambda-1} \frac{g^{(j)}\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}\left(\mathbf{X}_i\right)\right)}{j!}\hat{\mathbf{F}}_{k(l)}^{j}\left(\mathbf{X}_i\right)\right), \\ \mathbf{s}_i^{(l)} &:= \mathcal{M}\left(\Psi\left(\mathbf{X}_i\right)\hat{\mathbf{F}}_{k(l)}^{\lambda}\left(\mathbf{X}_i\right)\right). \end{aligned}$$

Then

$$Cov\left[g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right), g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_j)\right)\right]$$

$$= \mathbb{E}\left[\left(\mathbf{p}_i^{(l)} + \mathbf{q}_i^{(l)} + \mathbf{r}_i^{(l)} + \mathbf{s}_i^{(l)}\right)\left(\mathbf{p}_j^{(l')} + \mathbf{q}_j^{(l')} + \mathbf{r}_j^{(l')} + \mathbf{s}_j^{(l')}\right)\right]. \tag{4}$$

To obtain expressions for $\hat{\mathbf{F}}^i_{k(l)}(\mathbf{Z})$, we expand $\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})$ around $\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})$ and $\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})$:

$$\begin{aligned}
\frac{\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} &= \frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} + \frac{\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} - \mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})\frac{\hat{\mathbf{e}}_{2,k(l)}(\mathbf{Z})}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^2} \\
&\quad - \frac{\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}_{2,k(l)}(\mathbf{Z})}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^2} + \mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})\frac{\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z})}{2\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^3} \\
&\quad + \frac{\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z})}{2\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^3} + o\left(\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z}) + \hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z})\right) \qquad (5) \\
&= \frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} + h(\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z}), \hat{\mathbf{e}}_{2,k(l)}(\mathbf{Z})).
\end{aligned}$$

Let $\mathbf{h}(\mathbf{Z}) = h(\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z}), \hat{\mathbf{e}}_{2,k(l)}(\mathbf{Z}))$. Thus $\hat{\mathbf{F}}_{k(l)}(\mathbf{Z}) = \frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} - \mathbb{E}_{\mathbf{Z}}\hat{\mathbf{L}}_{k(l)}(\mathbf{Z}) + \mathbf{h}(\mathbf{Z})$. By the binomial theorem,

$$\hat{\mathbf{F}}^q_{k(l)}(\mathbf{Z}) = \sum_{j=0}^{q} a_{q,j}\left(\frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} - \mathbb{E}_{\mathbf{Z}}\hat{\mathbf{L}}_{k(l)}(\mathbf{Z})\right)^{q-j}\mathbf{h}^j(\mathbf{Z}), \tag{6}$$

where $a_{q,j}$ is the binomial coefficient. Using a Taylor series expansion of $\frac{1}{x}$ about $\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})$,

$$\begin{aligned}
\mathbb{E}_{\mathbf{Z}}\frac{1}{\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})} &= \mathbb{E}_{\mathbf{Z}}\left[\frac{1}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})} - \frac{\hat{\mathbf{e}}_{2,k_2}}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})\right)^2} + \frac{\hat{\mathbf{e}}^2_{2,k_2}}{2\xi_{2,\mathbf{Z}}}\right] \\
&= \frac{1}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})} + \frac{\left(\mathbb{V}_{\mathbf{Z}}\left[\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})\right]\right)}{2\xi_{2,\mathbf{Z}}} \\
&= \frac{1}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})} + c_{3,2}(\mathbf{Z})\left(\frac{1}{k_2}\right), \qquad (7)
\end{aligned}$$

where $\xi_{2,\mathbf{Z}} \in \left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k_2}(\mathbf{Z}), \hat{\mathbf{f}}_{2,k_2}(\mathbf{Z})\right)$ from the mean value thoerem and we use the fact that the variance of the kernel density estimate converges to zero with rate $\frac{1}{M\sigma}$ where $\sigma = O\left(\frac{k(l)}{M}\right)$. Sricharan et al [4] showed that for a truncated uniform kernel density estimator with bandwidth $(k/M)^{1/d}$, $\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{i,k(l)}(\mathbf{Z}) = f_i(\mathbf{Z}) + \sum_{j=1}^{d} c_{i,j,k(l)}(\mathbf{Z})\left(\frac{k(l)}{M}\right)^{j/d} + o\left(\frac{k(l)}{M}\right) = f_i(\mathbf{Z}) + c_{1,i}(\mathbf{Z}, k(l), M) = f_i(\mathbf{Z}) + o(1)$. It can then be shown that the k-nn density estimator converges to a truncated uniform kernel density estimator [5]. Thus the result holds for the k-nn density estimator as well. Combining this with Eq. 7 gives

$$
\begin{aligned}
& \left(\frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} - \mathbb{E}_{\mathbf{Z}}\hat{\mathbf{L}}_{k_1,k_2}(\mathbf{Z})\right)^q \\
= & \left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})c_{3,2}(\mathbf{Z})\left(\frac{1}{k(l)}\right)\right)^q \qquad (8)\\
= & \left(f_1(\mathbf{Z})c_{3,2}(\mathbf{Z})\left(\frac{1}{k(l)}\right) + \sum_{j=1}^{d} c_{1,j,k(l)}\left(\frac{k(l)}{M}\right)^{\frac{j}{d}}\left(\frac{1}{k(l)}\right) + o\left(\frac{1}{M}\right)\right)^q \\
= & 1_{\{q=1\}}c_3(\mathbf{Z})\left(\frac{1}{k(l)}\right) + 1_{\{q\geq 2\}}O\left(\frac{1}{k(l)^q}\right) + o\left(\frac{1}{M}\right) =: b_{q,k(()}\mathbf{Z}). \qquad (9)
\end{aligned}
$$

Combining Eqs. 5, 6, and 9 gives

$$
\begin{aligned}
\hat{\mathbf{F}}^q_{k(l)}(\mathbf{Z}) = \; & b_{q,k(l)}(\mathbf{Z}) + b^{1_{\{q\geq 2\}}}_{q-1,k(l)}(\mathbf{Z})a_{q,1}\mathbf{h}(\mathbf{Z}) + 1_{\{q\geq 2\}}b^{1_{\{q\geq 3\}}}_{q-2,k(l)}(\mathbf{Z})a_{q,2}\mathbf{h}^2(\mathbf{Z}) \\
& + 1_{\{q\geq 3\}}b^{1_{\{q\geq 4\}}}_{q-3,k(l)}\mathbf{Z})O\left(\mathbf{h}^3(\mathbf{Z})\right) \qquad (10)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{h}(\mathbf{Z}) = \; & \frac{\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})}{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})} - \frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^2}\hat{\mathbf{e}}_{2,k(l)}(\mathbf{Z}) - \frac{\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}_{2,k(l)}(\mathbf{Z})}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^2} \\
& + \frac{\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})}{2\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^3}\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z}) + o\left(\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z})\right), \\
\mathbf{h}^2(\mathbf{Z}) = \; & \frac{\hat{\mathbf{e}}^2_{1,k(l)}(\mathbf{Z})}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^2} + \frac{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{1,k(l)}(\mathbf{Z})\right)^2}{\left(\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{f}}_{2,k(l)}(\mathbf{Z})\right)^4}\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z}) \\
& + O\left(\hat{\mathbf{e}}_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}_{2,k_2}(\mathbf{Z}) + \hat{\mathbf{e}}^3_{2,k(l)}(\mathbf{Z})\right), \\
O\left(\mathbf{h}^3(\mathbf{Z})\right) = \; & O\left(\hat{\mathbf{e}}^3_{1,k(l)}(\mathbf{Z}) + \hat{\mathbf{e}}^3_{2,k(l)}(\mathbf{Z}) + \hat{\mathbf{e}}^2_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^2_{2,k(l)}(\mathbf{Z})\right).
\end{aligned}
$$

We now obtain bounds on the expected value of products of the $\hat{\mathbf{e}}_{i,k(l)}$ terms:

Lemma 2. *Let $l, l' \in \bar{l}$ be fixed, $M_1 = M_2 = M$, and $k(l) = l\sqrt{M}$. Let $\gamma(z)$ be an arbitrary function with $\sup_z |\gamma(z)| < \infty$. Let $\mathbf{Z}$ be a realization of the density f_2 independent of $\hat{\mathbf{f}}_{i,k(l)}$ and $\hat{\mathbf{f}}_{i,k(l')}$ for $i = 1, 2$. Then,*

$$
\mathbb{E}\left[\gamma(\mathbf{Z})\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{Z})\right] = \begin{cases} 1_{\{q=2\}}\left(c_{2,i}(\gamma(z))\left(\frac{1}{k(l)}\right) + o\left(\frac{1}{k(l)}\right)\right) + 1_{\{q\geq 3\}}O\left(\frac{1}{k(l)^{\frac{q}{2}}}\right), & q \geq 2 \\ 0, & q = 1, \end{cases} \qquad (11)
$$

$$
\mathbb{E}\left[\gamma(\mathbf{Z})\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^r_{i,k(l')}(\mathbf{Z})\right] = \begin{cases} O\left(\frac{1}{k(l)^{\frac{q}{2}}k(l')^{\frac{r}{2}}}\right), & q + r \geq 2 \\ 0, & otherwise \end{cases} \qquad (12)
$$

$$
\begin{aligned}
& \mathbb{E}\left[\gamma(\mathbf{Z})\hat{\mathbf{e}}^q_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^{q'}_{1,k(l')}(\mathbf{Z})\hat{\mathbf{e}}^r_{2,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^{r'}_{2,k(l')}(\mathbf{Z})\right] \\
= & \begin{cases} 0, & q + q' = 1 \text{ or } r + r' = 1 \\ O\left(\frac{1}{k(l)^{\frac{q+r}{2}}k(l')^{\frac{q'+r'}{2}}}\right), & otherwise \end{cases} \qquad (13)
\end{aligned}
$$

$$
\mathbb{E}\left[\gamma(\mathbf{Z})\hat{\mathbf{F}}^q_{k(l)}(\mathbf{Z})\right] = 1_{\{q=1\}}O\left(\frac{1}{k(l)}\right) + 1_{\{q\geq 2\}}O\left(\frac{1}{k(l)^{\frac{q}{2}}}\right). \qquad (14)
$$

Proof. For $i = 2$, Eq. 11 is given and proved as Lemma 5 in [4] where the density estimator is a truncated uniform kernel density estimator with bandwidth $(k(l)/M)^{1/d}$. The proof uses concentration inequalities to bound $\mathbb{E}_{\mathbf{Z}}\hat{\mathbf{e}}^q_{2,k(l)}(\mathbf{Z})$ in terms of $k(l)$. Then since the truncated uniform kernel density estimator converges to the k-nn estimator, it holds for the k-nn estimator as well. For $i = 1$, the proof follows the same procedure but results in a different constant.

Equation 12 is proved in a similar manner. Let $S_l(X) := \left\{Y \in \mathcal{S} : ||X - Y||_\infty \leq \left(k(l)/M\right)^{1/d}/2\right\}$, $V_l(X) := \int_{S_l(X)} dz$, $U_{i,l}(X) := Pr(\mathbf{Z} \in S_l(X))$ where $\mathbf{Z}$ is drawn from f_i, and $\mathbf{1}_{i,l}(X)$ denote the number of samples from the ith distribution that fall in $S_l(X)$; i.e. the number of samples from$\{\mathbf{Y}_1, \dots, \mathbf{Y}_M\}$ if $i = 1$ or $\{\mathbf{X}_{N+1}, \dots, \mathbf{X}_{N+M}\}$ if $i = 2$ that fall in $S_l(X)$. The uniform kernel density estimator is then

$$\tilde{\mathbf{f}}_{i,k(l)}(X) = \frac{\mathbf{1}_{i,l}(X)}{MV_l(X)}.$$

Let $\natural_{i,l}(X)$ denote the event $(1 - p_{k(l)})MU_{i,l}(X) < \mathbf{1}_{i,l}(X) < (1 + p_{k(l)})MU_{i,l}(X)$, where $p_{k(l)} = 1/k(l)^{\delta/2}$. It can be shown [4] using standard Chernoff inequalities that $Pr(\natural^C_{i,l}(X)) = O\left(e^{-p^2_{k(l)}k(l)}\right)$ and that under the event $\natural_{i,l}(X)$, $\hat{\mathbf{e}}_{i,k(l)} = O(1/(k^{\delta/2}))$. Thus

$$\begin{aligned}
\mathbb{E}\left[\gamma(\mathbf{Z})\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^r_{i,k(l')}(\mathbf{Z})\right] &= \mathbb{E}\left[\gamma(\mathbf{Z})1_{\natural_{i,l}(X)\cap\natural_{i,l'}(X)}\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^r_{i,k(l')}(\mathbf{Z})\right] \\
&\quad + \mathbb{E}\left[\gamma(\mathbf{Z})1_{\{\natural_{i,l}(X)\cap\natural_{i,l'}(X)\}^C}\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^r_{i,k(l')}(\mathbf{Z})\right] \\
&= O\left(\frac{1}{k(l)^{\frac{q}{2}}k(l')^{\frac{r}{2}}}\right),
\end{aligned}$$

where we use the fact that δ can be chosen arbitrarily close to 1.

For Eq. 13, note that due to conditional independence and Eq. 12,

$$\begin{aligned}
&\mathbb{E}\left[\gamma(\mathbf{Z})\hat{\mathbf{e}}^q_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^{q'}_{1,k(l')}(\mathbf{Z})\hat{\mathbf{e}}^r_{2,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^{r'}_{2,k(l')}(\mathbf{Z})\right] \\
= \; &\mathbb{E}\left[\gamma(\mathbf{Z})\mathbb{E}_{\mathbf{Z}}\left[\hat{\mathbf{e}}^q_{1,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^{q'}_{1,k(l')}(\mathbf{Z})\right]\mathbb{E}_{\mathbf{Z}}\left[\hat{\mathbf{e}}^r_{2,k(l)}(\mathbf{Z})\hat{\mathbf{e}}^{r'}_{2,k(l')}(\mathbf{Z})\right]\right] \\
= \; &\mathbb{E}\left[\gamma(\mathbf{Z})\left(O\left(\frac{1}{k(l)^{\frac{q+r}{2}}k(l')^{\frac{q'+r'}{2}}}\right)\right)\right] \\
= \; &O\left(\frac{1}{k(l)^{\frac{q+r}{2}}k(l')^{\frac{q'+r'}{2}}}\right).
\end{aligned}$$

Equation 14 is obtained by applying Eqs. 11 and 12 to Eq. 10. $\square$

The following lemma provides bounds on the covariance between the $\hat{\mathbf{F}}^q_{k(l)}(\mathbf{Z})$ terms:

Lemma 3. *Let $l, l' \in \bar{l}$ be fixed, $M_1 = M_2 = M$, and $k(l) = l\sqrt{M}$. Let $\gamma_1(x), \gamma_2(x)$ be arbitrary functions with 1 partial derivative wrt x and $\sup_x |\gamma_i(x)| < \infty$, $i = 1, 2$. Let $\mathbf{X}_i$ and $\mathbf{X}_j$ be realizations of the density f_2 independent of $\hat{\mathbf{f}}_{1,k(l)}$, $\hat{\mathbf{f}}_{1,k(l')}$, $\hat{\mathbf{f}}_{2,k(l)}$, and $\hat{\mathbf{f}}_{2,k(l')}$ and independent of each other when $i \neq j$. Then*

$$Cov\left[\gamma_1(\mathbf{X}_i)\hat{\mathbf{F}}^q_{k(l)}(\mathbf{X}_i), \gamma_2(\mathbf{X}_j)\hat{\mathbf{F}}^r_{k(l')}(\mathbf{X}_j)\right] = \begin{cases} o(1), & i = j \\ 1_{\{q=1,r=1\}}c_8\left(\gamma_1(x), \gamma_2(x)\right)\left(\frac{1}{M}\right) + o\left(\frac{1}{M}\right), & i \neq j. \end{cases}$$

Proof. Throughout the following, assume that $\mathbf{X}$ and $\mathbf{Y}$ are realizations of the density f_2 independent of each other and $\hat{\mathbf{f}}_{1,k(l)}$, $\hat{\mathbf{f}}_{1,k(l')}$, $\hat{\mathbf{f}}_{2,k(l)}$, and $\hat{\mathbf{f}}_{2,k(l')}$. First consider the case where $i = j$. By Cauchy-Schwarz and Eq. 11,

$$Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{X}), \gamma_2(\mathbf{X})\hat{\mathbf{e}}^r_{i,k(l')}(\mathbf{X})\right] = O\left(\frac{1}{M^{\frac{q+r}{4}}}\right). \tag{15}$$

By Eq. 12 and Eq. 13,

$$Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}^q_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}^r_{2,k(l)}(\mathbf{X}),\gamma_2(\mathbf{X})\hat{\mathbf{e}}^{q'}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}^{r'}_{2,k(l')}(\mathbf{X})\right]=O\left(\frac{1}{M^{\frac{q+r+q'+r'}{4}}}\right). \quad (16)$$

Applying Eqs. 15 and 16 to Eq. 10 completes the proof for this case.

We'll now prove the case where $i\neq j$. Define $\Psi(l,l')=\left\{\|X-Y\|_1\geq 2\left(\frac{\max(k(l),k(l'))}{M}\right)^{\frac{1}{d}}\right\}$. For a fixed pair of points $\{X,Y\}\in\Psi(l,l')$,

$$Cov\left[\hat{\mathbf{e}}^q_{i,k(l)}(X),\hat{\mathbf{e}}^r_{i,k(l')}(Y)\right]=1_{\{q=r=1\}}\left(\frac{-f_i(X)f_i(Y)}{M}\right)+o\left(\frac{1}{M}\right). \quad (17)$$

This can be shown in the same way as in the proof of Lemma 6 in [4] for a truncated uniform kernel density estimator. This is done by recognizing that for $\{X,Y\}\in\Psi(l,l')$, the functions $\mathbf{1}_{i,l}(X)$ and $\mathbf{1}_{i,l'}(Y)$ are distributed jointly as a multinomial random variable with parameters M, $U_{i,l}(X)$, $U_{i,l'}(Y)$ and $1-U_{i,l}(X)-U_{i,l'}(Y)$. Equation 17 is then established by using the concentration inequality for the high probability event of $\natural_{i,l}(X)\cap\natural_{i,l'}(Y)$ and then relating the functions $\mathbf{1}_{i,l}(X)$ and $\mathbf{1}_{i,l'}(Y)$ to two binomial random variables with parameters $\{U_{i,l}(X),M-q\}$ and $\{U_{i,l'}(Y),M-r\}$, respectively. Note that the relationship holds whether $l=l'$ or $l\neq l'$. For fixed $\{X,Y\}\in\Psi(l,l')^C$, Cauchy-Schwarz and Eq. 11 give

$$Cov\left[\hat{\mathbf{e}}^q_{i,k(l)}(X),\hat{\mathbf{e}}^r_{i,k(l')}(Y)\right]=O\left(\frac{1}{k(l)^{\frac{q}{2}}k(l')^{\frac{r}{2}}}\right). \quad (18)$$

From Eqs. 17 and 18, we have that

$$Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}^q_{i,k(l)}(\mathbf{X}),\gamma_2(\mathbf{Y})\hat{\mathbf{e}}^r_{i,k(l')}(\mathbf{Y})\right]=1_{\{q=r=1\}}c_{7,i}(\gamma_1(x),\gamma_2(x))\left(\frac{1}{M}\right)+o\left(\frac{1}{M}\right). \quad (19)$$

This is proved in the same way as in the proof of Lemma 8 in [4] by splitting the covariances into the cases where $\{\mathbf{X},\mathbf{Y}\}\in\Psi(l,l')$ and $\{\mathbf{X},\mathbf{Y}\}\in\Psi(l,l')^C$. For the first case, the bound falls clearly from Eq. 17. For the second case, the bound holds with Eq. 18 since $\int_{\Psi(l,l')^C}dy=2^d\frac{\max(k(l),k(l'))}{M}$.

Now let $E_0=\{s,q,t,r\geq 1\}$, $E_{1,1}=\{s=0,q\geq 2,t\geq 1,r\geq 1\}\cup\{s\geq 1,q\geq 1,t=0,r\geq 2\}$, and $E_{1,2}=\{s\geq 2,q=0,t\geq 1,r\geq 1\}\cup\{s\geq 1,q\geq 1,t\geq 2,r=0\}$. For fixed X,Y, we have by Eqs. 11 and 12 and conditional independence when E_0, $E_{1,1}$, or $E_{1,2}$ hold that

$$\begin{aligned}&Cov\left[\gamma_1(X)\hat{\mathbf{e}}^s_{1,k(l)}(X)\hat{\mathbf{e}}^q_{2,k(l)}(X),\gamma_2(Y)\hat{\mathbf{e}}^t_{1,k(l')}(Y)\hat{\mathbf{e}}^r_{2,k(l')}(Y)\right]\\
&= \mathbb{E}\left[\gamma_1(X)\hat{\mathbf{e}}^s_{1,k(l)}(X)\hat{\mathbf{e}}^q_{2,k(l)}(X)\gamma_2(Y)\hat{\mathbf{e}}^t_{1,k(l')}(Y)\hat{\mathbf{e}}^r_{2,k(l')}(Y)\right]\\
&\quad -\mathbb{E}\left[\gamma_1(X)\hat{\mathbf{e}}^s_{1,k(l)}(X)\hat{\mathbf{e}}^q_{2,k(l)}(X)\right]\mathbb{E}\left[\gamma_2(Y)\hat{\mathbf{e}}^t_{1,k(l')}(Y)\hat{\mathbf{e}}^r_{2,k(l')}(Y)\right]\\
&= \gamma_1(X)\gamma_2(Y)\mathbb{E}\left[\hat{\mathbf{e}}^s_{1,k(l)}(X)\hat{\mathbf{e}}^t_{1,k(l')}(Y)\right]\mathbb{E}\left[\hat{\mathbf{e}}^q_{2,k(l)}(X)\hat{\mathbf{e}}^r_{2,k(l')}(Y)\right]\\
&\quad +1_{\{q,r,s,t\neq 1\}}O\left(\frac{1}{k(l)^{\frac{q+s}{2}}k(l')^{\frac{r+t}{2}}}\right). \end{aligned} \quad (20)$$

Now $\mathbb{E}\left[\hat{\mathbf{e}}^s_{i,k(l)}(X)\hat{\mathbf{e}}^t_{i,k(l')}(Y)\right]=Cov\left[\hat{\mathbf{e}}^s_{i,k(l)}(X),\hat{\mathbf{e}}^t_{i,k(l')}(Y)\right]+\mathbb{E}\left[\hat{\mathbf{e}}^s_{i,k(l)}(X)\right]\mathbb{E}\left[\hat{\mathbf{e}}^t_{i,k(l')}(Y)\right]$. By Eqs. 11, 17, and 18 this gives (when $s,t\geq 1$)

$$\begin{aligned}\mathbb{E}\left[\hat{\mathbf{e}}^s_{i,k(l)}(X)\hat{\mathbf{e}}^t_{i,k(l')}(Y)\right] &= 1_{\{s,t\geq 2\}}O\left(\frac{1}{k(l)^{\frac{s}{2}}k(l')^{\frac{t}{2}}}\right)\\
&\quad+\begin{cases}1_{\{s=t=1\}}\left(\frac{-f_i(X)f_i(Y)}{M}\right)+o\left(\frac{1}{M}\right), & \{X,Y\}\in\Psi(l,l')\\ O\left(\frac{1}{k(l)^{\frac{s}{2}}k(l')^{\frac{t}{2}}}\right), & \{X,Y\}\in\Psi(l,l')^c.\end{cases}\end{aligned} \quad (21)$$

Now

$$\mathbb{E}\left[Cov_{\mathbf{X},\mathbf{Y}}\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}_{1,k(l)}^s(\mathbf{X})\hat{\mathbf{e}}_{2,k(l)}^q(\mathbf{X}),\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l')}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k(l')}^r(\mathbf{Y})\right]\right]=I_1+I_2,$$

where

$$I_1=\mathbb{E}\left[1_{\{\mathbf{X},\mathbf{Y}\}\in\Psi(l,l')^C}\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})Cov_{\mathbf{X},\mathbf{Y}}\left[\hat{\mathbf{e}}_{1,k(l)}{}^s(\mathbf{X})\hat{\mathbf{e}}_{2,k(l)}^q(\mathbf{X}),\hat{\mathbf{e}}_{1,k(l')}^t(Y)\hat{\mathbf{e}}_{2,k(l')}^r(\mathbf{Y})\right]\right],$$

$$I_2=\mathbb{E}\left[1_{\{\mathbf{X},\mathbf{Y}\}\in\Psi(l,l')}\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})Cov_{\mathbf{X},\mathbf{Y}}\left[\hat{\mathbf{e}}_{1,k(l)}{}^s(\mathbf{X})\hat{\mathbf{e}}_{2,k(l)}^q(\mathbf{X}),\hat{\mathbf{e}}_{1,k(l')}^t(Y)\hat{\mathbf{e}}_{2,k(l')}^r(\mathbf{Y})\right]\right].$$

Combining Eqs. 20 and 21 gives

$$\begin{aligned}
I_1 &= \mathbb{E}\left[1_{\{\mathbf{X},\mathbf{Y}\}\in\Psi(l,l')^C}\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})O\left(\frac{1}{k(l)^{\frac{s+q}{2}}k(l')^{\frac{r+t}{2}}}\right)\right]\\
&= \int\left(O\left(\frac{1}{k(l)^{\frac{s+q}{2}}k(l')^{\frac{r+t}{2}}}\right)(\gamma_1(x)\gamma_2(x)+o(1))\right)\left(\int_{\{x,y\}\in\Psi(l,l')^C}dy\right)dx\\
&= \int\left(O\left(\frac{1}{k(l)^{\frac{s+q}{2}}k(l')^{\frac{r+t}{2}}}\right)(\gamma_1(x)\gamma_2(x)+o(1))\right)\left(2^d\frac{\max(k(l),k(l'))}{M}\right)dx\\
&= o\left(\frac{1}{M}\right).
\end{aligned}$$

Similarly,

$$I_2=\mathbb{E}\left[1_{\{\mathbf{X},\mathbf{Y}\}\in\Psi(l,l')}\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})o\left(\frac{1}{M}\right)\right]=o\left(\frac{1}{M}\right),$$

and so

$$Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}_{1,k(l)}^s(\mathbf{X})\hat{\mathbf{e}}_{2,k(l)}^q(\mathbf{X}),\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l')}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k(l')}^r(\mathbf{Y})\right]=o\left(\frac{1}{M}\right). \tag{22}$$

Assume now that neither E_0, $E_{1,1}$, nor $E_{1,2}$. If either $q,r=0$ or $s,t=0$ and the remaining exponents are nonzero, then the left hand side of Eq. 22 reduces to Eq. 19. For the other cases, suppose that $s,q=0$ and $t,r\geq 2$ as an example. Then we have that

$$\begin{aligned}
Cov\left[\gamma_1(\mathbf{X}),\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l')}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k(l')}^r(\mathbf{Y})\right] &= \mathbb{E}\left[\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k_1}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k_2}^r(\mathbf{Y})\right]\\
&\quad -\mathbb{E}\left[\gamma_1(\mathbf{X})\right]\mathbb{E}\left[\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k_1}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k_2}^r(\mathbf{Y})\right]\\
&= \mathbb{E}\left[\gamma_1(\mathbf{X})\right]\mathbb{E}\left[\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k_1}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k_2}^r(\mathbf{Y})\right]\\
&\quad -\mathbb{E}\left[\gamma_1(\mathbf{X})\right]\mathbb{E}\left[\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k_1}^t(\mathbf{Y})\hat{\mathbf{e}}_{2,k_2}^r(\mathbf{Y})\right]\\
&= 0.
\end{aligned}$$

The same result follows for all other cases.

Finally, applying Eqs. 19 and 22 to Eq. 10 gives

$$\begin{aligned}
&Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{F}}_{k(l)}^q(\mathbf{X}),\gamma_2(\mathbf{Y})\hat{\mathbf{F}}_{k(l')}^r(\mathbf{Y})\right]=o\left(\frac{1}{M}\right)+1_{\{q=1,r=1\}}\left(\frac{1}{M}\right)\times\\
&\left(c_{7,1}\left(\frac{\gamma_1(x)}{\mathbb{E}_X\hat{\mathbf{f}}_{2,k(l)}(x)},\frac{\gamma_2(x)}{\mathbb{E}_X\hat{\mathbf{f}}_{2,k(l')}(x)}\right)+c_{7,2}\left(\frac{-\gamma_1(x)\mathbb{E}_X\hat{\mathbf{f}}_{1,k(l)}(x)}{\mathbb{E}_X\hat{\mathbf{f}}_{2,k(l)}(x)},\frac{-\gamma_2(x)\mathbb{E}_X\hat{\mathbf{f}}_{1,k(l')}(x)}{\mathbb{E}_X\hat{\mathbf{f}}_{2,k(l')}(x)}\right)\right)\\
&=1_{\{q=1,r=1\}}c_8\left(\gamma_1(x),\gamma_2(x)\right)\left(\frac{1}{M}\right)+o\left(\frac{1}{M}\right).
\end{aligned}$$

Note that this holds even if $l=l'$. □

The following lemma is required to bound the $\Psi(\mathbf{Z})$ term.

Lemma 4. *Assume that U(x) is any arbitrary functional which satisfies*

$$
\begin{array}{ll}
(i) & \mathbb{E}\left[\sup_{L\in(p_l,p_u)}\left|U\left(L\frac{\mathbf{p}_2}{\mathbf{p}_1}\right)\right|\right]=G_1<\infty, \\
(ii) & \sup_{L\in\left(\frac{q_{l,1}}{q_{u,2}},\frac{q_{u,1}}{q_{l,2}}\right)}\left|U\left(L\right)\right|\mathcal{C}\left(k_1\right)\mathcal{C}\left(k_2\right)=G_2<\infty, \\
(iii) & \mathbb{E}\left[\sup_{L\in\left(\frac{q_{l,1}}{p_{u,2}},\frac{q_{u,1}}{p_{l,2}}\right)}\left|U\left(L\mathbf{p}_2\right)\right|\mathcal{C}\left(k_1\right)\right]=G_3<\infty, \\
(iv) & \mathbb{E}\left[\sup_{L\in\left(\frac{p_{l,1}}{q_{u,2}},\frac{p_{u,1}}{q_{l,2}}\right)}\left|U\left(\frac{L}{\mathbf{p}_1}\right)\right|\mathcal{C}\left(k_2\right)\right]=G_4<\infty.
\end{array}
$$

Let $\mathbf{Z}$ *be* $\mathbf{X}_i$ *for some fixed* $i\in\{1,\dots,N\}$ *and* $\xi_{\mathbf{Z}}$ *be any random variable which almost surely lies in* $(L(\mathbf{Z}),\hat{\mathbf{L}}_{k_1,k_2}(\mathbf{Z}))$. *Then* $\mathbb{E}|U(\xi_{\mathbf{Z}})|<\infty$.

Proof. This is a version of Lemma 9 in [4] modified to apply to functionals of the likelihood ratio. Because of assumption $\mathcal{A}.1$, it is sufficient to show that the conditional expectation $\mathbb{E}\left[|U(\xi_Z)|\,|\,\mathbf{X}_1,\dots,\mathbf{X}_N\right]<\infty$.

First, some properties of k-NN density estimators are required. Let $\mathbf{S}_{k_i,i}(Z)=\left\{Y:d(Z,Y)\le\mathbf{d}_{Z,i}^{(k_i)}\right\}$ where $\mathbf{d}_{Z,i}^{(k_i)}$ is the distance to the k_ith nearest neighbor of Z from the corresponding set of samples. Then let $\mathbf{P}_i(Z)=\int_{\mathbf{S}_{k_i,i}(Z)}f_i(x)dx$ which has a beta distribution with parameters k_i and M_i-k_i+1 [6]. Let $A_i(Z)$ be the event that $\mathbf{P}_i(Z)<\left(\frac{\sqrt{6}}{k_i^{\delta/2}}+1\right)\frac{k_i-1}{M_i}$. It has been shown that $Pr\left(A_i(Z)^C\right)=\Theta\left(\mathcal{C}\left(k_i\right)\right)$ and that under $A_i(Z)$ [3, 5],

$$\frac{p_{l,i}}{\mathbf{P}_i(Z)}<\hat{\mathbf{f}}_{i,k_i}(Z)<\frac{p_{u,i}}{\mathbf{P}_i(Z)}.$$

It has also been shown that under $A_i(Z)^C$ [3, 5],

$$q_{l,i}<\hat{\mathbf{f}}_{i,k_i}(Z)<q_{u,i}.$$

Let $A(Z)=A_1(Z)\cap A_2(Z)$ and note that $A_1(Z)$ and $A_2(Z)$ are independent events. Thus since $\hat{\mathbf{L}}_{k_1,k_2}(Z)=\frac{\hat{\mathbf{f}}_{1,k_1}(Z)}{\hat{\mathbf{f}}_{2,k_2}(Z)}$, we have that under $A(Z)$,

$$p_l\frac{\mathbf{P}_2(Z)}{\mathbf{P}_1(Z)}<\hat{\mathbf{L}}_{k_1,k_2}(Z)<p_u\frac{\mathbf{P}_2(Z)}{\mathbf{P}_1(Z)}.$$

Now let $Q_1(Z)=A_1(Z)^C\cap A_2(Z)^C$, $Q_2(Z)=A_1(Z)^C\cap A_2(Z)$, and $Q_3(Z)=A_1(Z)\cap A_2(Z)^C$. Then due to independence and the fact that the $Q_i(Z)$s are disjoint,

$$
\begin{aligned}
A(Z)^C &= A_1(Z)^C\cup A_2(Z)^C=Q_1(Z)\cup Q_2(Z)\cup Q_3(Z),\\
\implies Pr\left(A(Z)^C\right) &= Pr\left(Q_1(Z)\right)+Pr\left(Q_2(Z)\right)+Pr\left(Q_3(Z)\right)\\
&\le \mathcal{C}\left(k_1\right)\mathcal{C}\left(k_2\right)+\mathcal{C}\left(k_1\right)+\mathcal{C}\left(k_2\right).
\end{aligned}
$$

Then under $Q_1(Z)$, $Q_2(Z)$, and $Q_3(Z)$, respectively,

$$\frac{q_{l,1}}{q_{u,2}}<\hat{\mathbf{L}}_{k_1,k_2}(Z)<\frac{q_{u,1}}{q_{l,2}},$$

$$\frac{q_{l,1}\mathbf{P}_2(Z)}{p_{u,2}}<\hat{\mathbf{L}}_{k_1,k_2}(Z)<\frac{q_{u,1}\mathbf{P}_2(Z)}{p_{l,2}},$$

$$\frac{p_{l,1}}{\mathbf{P}_1(Z)q_{u,2}}<\hat{\mathbf{L}}_{k_1,k_2}(Z)<\frac{p_{u,1}}{\mathbf{P}_1(Z)q_{l,2}}.$$

Conditioning on $\mathbf{X}_1, \dots, \mathbf{X}_N$ gives

$$\begin{aligned}
\mathbb{E}\left[|U(\xi_Z)|\right] &= \mathbb{E}\left[1_{A(Z)}\left|U(\xi_Z)\right|\right] + \mathbb{E}\left[1_{Q_1(Z)}\left|U(\xi_Z)\right|\right] + \mathbb{E}\left[1_{Q_2(Z)}\left|U(\xi_Z)\right|\right] + \mathbb{E}\left[1_{Q_3(Z)}\left|U(\xi_Z)\right|\right] \\
&\le Pr(A(Z))\mathbb{E}\left[\sup_{L\in(p_l,p_u)}\left|U\left(L\frac{\mathbf{P}_2(Z)}{\mathbf{P}_1(Z)}\right)\right|\right] + Pr(Q_1(Z))\sup_{L\in\left(\frac{q_{l,1}}{q_{u,2}},\frac{q_{u,1}}{q_{l,2}}\right)}\left|U\left(L\right)\right| \\
&\quad + Pr(Q_1(Z))\mathbb{E}\left[\sup_{L\in\left(\frac{q_{l,1}}{p_{u,2}},\frac{q_{u,1}}{p_{l,2}}\right)}\left|U\left(L\mathbf{P}_2(Z)\right)\right|\right] \\
&\quad + Pr(Q_1(Z))\mathbb{E}\left[\sup_{L\in\left(\frac{p_{l,1}}{q_{u,2}},\frac{p_{u,1}}{q_{l,2}}\right)}\left|U\left(\frac{L}{\mathbf{P}_1(Z)}\right)\right|\right] \\
&\le \mathbb{E}\left[\sup_{L\in(p_l,p_u)}\left|U\left(L\frac{\mathbf{P}_2(Z)}{\mathbf{P}_1(Z)}\right)\right|\right] + \sup_{L\in\left(\frac{q_{l,1}}{q_{u,2}},\frac{q_{u,1}}{q_{l,2}}\right)}\left|U\left(L\right)\right|\mathcal{C}\left(k_1\right)\mathcal{C}\left(k_2\right) \\
&\quad + \mathbb{E}\left[\sup_{L\in\left(\frac{q_{l,1}}{p_{u,2}},\frac{q_{u,1}}{p_{l,2}}\right)}\left|U\left(L\mathbf{P}_2(Z)\right)\right|\mathcal{C}\left(k_1\right)\right] \\
&\quad + \mathbb{E}\left[\sup_{L\in\left(\frac{p_{l,1}}{q_{u,2}},\frac{p_{u,1}}{q_{l,2}}\right)}\left|U\left(\frac{L}{\mathbf{P}_1(Z)}\right)\right|\mathcal{C}\left(k_2\right)\right] \\
&= G_1 + G_2 + G_3 + G_4 < \infty.
\end{aligned}$$

□

The next lemma gives the last result necessary to bound the covariance of $\mathbf{Y}_{M,1}$ and $\mathbf{Y}_{M,2}$.

Lemma 5. *Let $l, l' \in \bar{l}$ be fixed, $M_1 = M_2 = M$, $k(l) = l\sqrt{M}$, and $\hat{\mathbf{L}}_{k(l)} = \hat{\mathbf{L}}_{k(l),k(l)}$. Let $\mathbf{X}_i$ and $\mathbf{X}_j$ be realizations of the density f_2 independent of $\hat{\mathbf{f}}_{1,k_1}$ and $\hat{\mathbf{f}}_{2,k_2}$ and independent of each other when $i \neq j$. Then*

$$Cov\left[g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right), g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_j)\right)\right]$$

$$= \begin{cases} \mathbb{E}\left[\mathbf{p}_i^{(l)}\mathbf{p}_i^{(l')}\right] + o(1), & i = j \\ c_8\left(g'\left(\mathbb{E}_X\hat{\mathbf{L}}_{k(l)}(x)\right), g'\left(\mathbb{E}_X\hat{\mathbf{L}}_{k(l')}(x)\right)\right)\left(\frac{1}{M}\right) + o\left(\frac{1}{M}\right), & i \neq j. \end{cases} \tag{23}$$

Proof. Consider the case where $i = j$. Then applying Lemma 3 to Eq. 4 gives

$$Cov\left[g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right), g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_j)\right)\right] = \mathbb{E}\left[\mathbf{p}_i^{(l)}\mathbf{p}_i^{(l')}\right] + o(1).$$

Note that $\mathbb{E}\left[\mathbf{p}_i^{(l)}\mathbf{p}_i^{(l')}\right] = O(1)$ since $\mathbf{p}_i^{(l)} = \mathcal{M}\left(g\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right) = \mathcal{M}\left(g\left(L(\mathbf{X}_i)\right)\right) + o(1)$.

Now let $i \neq j$. Since $\mathbf{X}_i$ and $\mathbf{X}_j$ are independent, $\mathbb{E}\left[\mathbf{p}_i^{(l)}\left(\mathbf{p}_j^{(l')} + \mathbf{q}_j^{(l')} + \mathbf{r}_j^{(l')} + \mathbf{s}_j^{(l')}\right)\right] = 0$. Applying Lemma 3 gives

$$\begin{aligned}
\mathbb{E}\left[\mathbf{q}_i^{(l)}\mathbf{q}_j^{(l')}\right] &= c_8\left(g'\left(\mathbb{E}_X\hat{\mathbf{L}}_{k(l)}(x)\right), g'\left(\mathbb{E}_X\hat{\mathbf{L}}_{k(l')}(x)\right)\right)\left(\frac{1}{M}\right) + o\left(\frac{1}{M}\right), \\
\mathbb{E}\left[\mathbf{q}_i^{(l)}\mathbf{r}_j^{(l')}\right] &= o\left(\frac{1}{M}\right), \\
\mathbb{E}\left[\mathbf{r}_i^{(l)}\mathbf{r}_j^{(l')}\right] &= o\left(\frac{1}{M}\right).
\end{aligned}$$

We use Cauchy-Schwarz and Lemma 2 to get

$$\mathbb{E}\left[g'\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}\left(\mathbf{X}_i\right)\right)\hat{\mathbf{F}}_{k(l)}\left(\mathbf{X}_i\right)\Psi\left(\mathbf{X}_j\right)\hat{\mathbf{F}}^{\lambda}_{k(l')}\left(\mathbf{X}_j\right)\right]$$

$$\begin{aligned}
&\leq \sqrt{\mathbb{E}\left[\Psi^2\left(\mathbf{X}_j\right)\right]\mathbb{E}\left[\left(g'\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}\left(\mathbf{X}_i\right)\right)\hat{\mathbf{F}}_{k(l)}\left(\mathbf{X}_i\right)\right)^2\hat{\mathbf{F}}^{2\lambda}_{k(l')}\left(\mathbf{X}_j\right)\right]} \\
&\leq \sqrt{\mathbb{E}\left[\Psi^2\left(\mathbf{X}_j\right)\right]\sqrt{\mathbb{E}\left[\left(g'\left(\mathbb{E}_{\mathbf{X}_i}\hat{\mathbf{L}}_{k(l)}\left(\mathbf{X}_i\right)\right)\hat{\mathbf{F}}_{k(l)}\left(\mathbf{X}_i\right)\right)^4\right]\mathbb{E}\left[\hat{\mathbf{F}}^{4\lambda}_{k(l')}\left(\mathbf{X}_j\right)\right]}} \\
&= \sqrt{\mathbb{E}\left[\Psi^2\left(\mathbf{X}_j\right)\right]\sqrt{O\left(\frac{1}{k(l)^2}\right)O\left(\frac{1}{k(l')^{2\lambda}}\right)}} \\
&= \sqrt{\mathbb{E}\left[\Psi^2\left(\mathbf{X}_j\right)\right]}o\left(\frac{1}{k(l')^{\lambda/2}}\right).
\end{aligned}$$

Lemma 4 and assumption $(\mathcal{A}.5)$ implies that $\mathbb{E}\left[\Psi^2\left(\mathbf{X}_j\right)\right] = O(1)$ and from assumption $(\mathcal{A}.3)$, $o\left(\frac{1}{k(l')^{\lambda/2}}\right) = o\left(\frac{1}{M}\right)$. This implies that $\mathbb{E}\left[\mathbf{q}_i^{(l)}\mathbf{s}_j^{(l')}\right] = o\left(\frac{1}{M}\right)$. Similarly, $\mathbb{E}\left[\mathbf{r}_i^{(l)}\mathbf{s}_j^{(l')}\right] = o\left(\frac{1}{M}\right)$ and $\mathbb{E}\left[\mathbf{s}_i^{(l)}\mathbf{s}_j^{(l')}\right] = o\left(\frac{1}{M}\right)$. Combining these results with Eq. 4 completes the proof. □

Applying Lemma 5 to Eqs. 2 and 3 shows that $Cov(\mathbf{Y}_{M,i},\mathbf{Y}_{M,j}) = O\left(\frac{1}{M}\right)$.

For the covariance of $\mathbf{Y}^2_{M,i}$ and $\mathbf{Y}^2_{M,j}$, we only need to consider the numerator since we previously showed that $\mathbb{V}\left[\sum_{l\in\bar{l}}w(l)g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right] = O(1)+o(1)$. Assume WLOG that $i=1$ and $j=2$ and let $h_l = \mathbb{E}\left[g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_i)\right)\right]$. The numerator of the covariance is then

$$\begin{aligned}
&\sum_{l\in\bar{l}}\sum_{l'\in\bar{l}}\sum_{j\in\bar{l}}\sum_{j'\in\bar{l}}Cov\left[\left(g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_1)\right)-h_l\right)\left(g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_1)\right)-h_{l'}\right),\right. \\
&\qquad\left.\left(g\left(\hat{\mathbf{L}}_{k(j)}(\mathbf{X}_2)\right)-h_j\right)\left(g\left(\hat{\mathbf{L}}_{k(j')}(\mathbf{X}_2)\right)-h_{j'}\right)\right] \\
&= \sum_{l\in\bar{l}}\sum_{l'\in\bar{l}}\sum_{j\in\bar{l}}\sum_{j'\in\bar{l}}Cov\left[\left(\mathbf{p}_1^{(l)}+\mathbf{q}_1^{(l)}+\mathbf{r}_1^{(l)}+\mathbf{s}_1^{(l)}\right)\left(\mathbf{p}_1^{(l')}+\mathbf{q}_1^{(l')}+\mathbf{r}_1^{(l')}+\mathbf{s}_1^{(l')}\right),\right. \\
&\qquad\left.\left(\mathbf{p}_2^{(j)}+\mathbf{q}_2^{(j)}+\mathbf{r}_2^{(j)}+\mathbf{s}_2^{(j)}\right)\left(\mathbf{p}_2^{(j')}+\mathbf{q}_2^{(j')}+\mathbf{r}_2^{(j')}+\mathbf{s}_2^{(j')}\right)\right].
\end{aligned}$$

Let $d_l(x) = (g(x)-h_l)^2$. Then for the case where $l=l'$ and $j=j'$, we have

$$Cov\left[d_l\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_1)\right), d_j\left(\hat{\mathbf{L}}_{k(j)}(\mathbf{X}_2)\right)\right] = O\left(\frac{1}{M}\right).$$

This follows from Lemma 5.

For the general case, note that due to the independence of $\mathbf{X}_1$ and $\mathbf{X}_2$,

$$\begin{aligned}
Cov\left[\mathbf{p}_1^{(l)}\mathbf{p}_1^{(l')}, \left(g\left(\hat{\mathbf{L}}_{k(j)}(\mathbf{X}_2)\right)-h_j\right)\left(g\left(\hat{\mathbf{L}}_{k(j')}(\mathbf{X}_2)\right)-h_{j'}\right)\right] &= 0, \\
Cov\left[\left(g\left(\hat{\mathbf{L}}_{k(l)}(\mathbf{X}_1)\right)-h_l\right)\left(g\left(\hat{\mathbf{L}}_{k(l')}(\mathbf{X}_1)\right)-h_{l'}\right), \mathbf{p}_2^{(j)}\mathbf{p}_2^{(j')}\right] &= 0.
\end{aligned}$$

To bound the remaining terms, we require the following Lemma:

Lemma 6. *Let $\gamma_1(x)$, $\gamma_2(x)$ be arbitrary functions with 1 partial derivative wrt x and $\sup_x|\gamma_i(x)| < \infty$, $i=1,2$. Let $l,l',j,j' \in \bar{l}$ be fixed, $M_1 = M_2 = M$, $k(l) = l\sqrt{M}$. Let $\mathbf{X}$ and $\mathbf{Y}$ be realizations of the density f_2 independent of $\hat{\mathbf{f}}_{i,k(l)}$, $\hat{\mathbf{f}}_{i,k(l')}$, $\hat{\mathbf{f}}_{i,k(j)}$, and $\hat{\mathbf{f}}_{i,k(j')}$, $i=1,2$. If $q,r,s,t\geq 0$ and the cases $\{t=0,r=0\}$ or $\{q=0,s=0\}$ do not hold, then*

$$Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{F}}^s_{k(l)}(\mathbf{X})\hat{\mathbf{F}}^q_{k(l')}(\mathbf{X}), \gamma_2(\mathbf{Y})\hat{\mathbf{F}}^t_{k(j)}(\mathbf{Y})\hat{\mathbf{F}}^r_{k(j')}(\mathbf{Y})\right] = O\left(\frac{1}{M}\right).$$

Proof. Under certain conditions, by Cauchy-Schwarz and Eq. 13 we have

$$\begin{aligned}
& Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}^{s}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}^{q}_{2,k(l)}(\mathbf{X})\hat{\mathbf{e}}^{s'}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}^{q'}_{2,k(l')}(\mathbf{X}),\right. \\
& \left.\gamma_2(\mathbf{Y})\hat{\mathbf{e}}^{t}_{1,k(j)}(\mathbf{Y})\hat{\mathbf{e}}^{r}_{2,k(j)}(\mathbf{Y})\hat{\mathbf{e}}^{t'}_{1,k(j')}(\mathbf{Y})\hat{\mathbf{e}}^{r'}_{1,k(j')}(\mathbf{Y})\right] \\
& \leq \mathbb{E}\left[\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\sqrt{\mathbb{V}_{\mathbf{X}}\left[\hat{\mathbf{e}}^{s}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}^{q}_{2,k(l)}(\mathbf{X})\hat{\mathbf{e}}^{s'}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}^{q'}_{2,k(l')}(\mathbf{X})\right]}\right. \\
& \left.\times\sqrt{\mathbb{V}_{\mathbf{Y}}\left[\hat{\mathbf{e}}^{t}_{1,k(j)}(\mathbf{Y})\hat{\mathbf{e}}^{r}_{2,k(j)}(\mathbf{Y})\hat{\mathbf{e}}^{t'}_{1,k(j')}(\mathbf{Y})\hat{\mathbf{e}}^{r'}_{1,k(j')}(\mathbf{Y})\right]}\right] \\
& = O\left(\frac{1}{k(l)^{\frac{s+q}{2}}k(l')^{\frac{s'+q'}{2}}k(j)^{\frac{t+r}{2}}k(j')^{\frac{t'+r'}{2}}}\right) = O\left(\frac{1}{M^{\frac{s+q+s'+q'+t+r+t'+r'}{4}}}\right). \qquad (24)
\end{aligned}$$

Note that the exponents q, s, r, t are not the same as in the statement of the lemma. The conditions under which this expression holds are as follows: (1) There must be at least one positive exponent on both sides of the arguments in the covariance. (2) $\{s+s'+t+t' \neq 1\} \cap \{q+q'+r+r' \neq 1\}$. If neither case holds, this reduces to Eq. 19. If only one holds, then the covariance is zero.

Note that if $s+q+s'+q'+t+r+t'+r' \geq 4$, Eq. 24 becomes $O\left(\frac{1}{M}\right)$. Now consider the case where $\{\{s+s'+t+t'=3\} \cap \{s,s',t,t' \leq 1\} \cap \{q,q',r,r'=0\}\} \cup \{\{q+q'+r+r'=3\} \cap \{q,q',r,r' \leq 1\} \cap \{s,s',t,t'=0\}\}$. Assume WLOG that $s,s',t=1$. Then Eq. 24 becomes $O\left(\frac{1}{M^{\frac{3}{4}}}\right)$ which does not decay fast enough to use Lemma 1. However, we can use the fact that $k(l) = O(k(l'))$ to obtain a bound of $O\left(\frac{1}{M}\right)$. By Markov's inequality and Eqs. 11 and 12, for fixed $\nu > 0$,

$$\begin{aligned}
Pr\left(\left|\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X}) - \hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\right| > \nu\right) & \leq \frac{\mathbb{E}\left[\left(\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X}) - \hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\right)^4\right]}{\nu^4} \\
& = O\left(\frac{1}{M}\right).
\end{aligned}$$

Let H be the event that $\left|\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X}) - \hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\right| \leq 1$. This gives

$$\begin{aligned}
& Cov\left[\gamma_1(\mathbf{X})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(l')}(\mathbf{X}), \gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] \\
= \; & \mathbb{E}\left[\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] \\
= \; & \mathbb{E}\left[\mathbf{1}_H\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] \\
& + \mathbb{E}\left[\mathbf{1}_{H^C}\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] \\
\leq \; & \mathbb{E}\left[\mathbf{1}_H\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}^2_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] + \mathbb{E}\left[\mathbf{1}_H\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] \\
& + \mathbb{E}\left[\mathbf{1}_{H^C}\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] \\
= \; & O\left(\frac{1}{M}\right). \qquad (25)
\end{aligned}$$

The final step for the first two terms comes from Eq. 19. The final step for the third term comes from the fact that $Pr(H^C) = O\left(\frac{1}{M}\right)$ and the fact that $\mathbb{E}\left[\gamma_1(\mathbf{X})\gamma_2(\mathbf{Y})\hat{\mathbf{e}}_{1,k(l)}(\mathbf{X})\hat{\mathbf{e}}_{1,k(l')}(\mathbf{X})\hat{\mathbf{e}}_{1,k(j)}(\mathbf{Y})\right] = o(1)$ by Eq. 24. Applying Eqs. 19, 22, 24, and 25 to Eq. 10 completes the proof. □

From Lemma 6, it is clear that

$$\begin{aligned}
& Cov\left[\left(\mathbf{p}_1^{(l)}+\mathbf{q}_1^{(l)}+\mathbf{r}_1^{(l)}+\mathbf{s}_1^{(l)}\right)\left(\mathbf{p}_1^{(l')}+\mathbf{q}_1^{(l')}+\mathbf{r}_1^{(l')}+\mathbf{s}_1^{(l')}\right),\right. \\
& \left.\left(\mathbf{p}_2^{(j)}+\mathbf{q}_2^{(j)}+\mathbf{r}_2^{(j)}+\mathbf{s}_2^{(j)}\right)\left(\mathbf{p}_2^{(j')}+\mathbf{q}_2^{(j')}+\mathbf{r}_2^{(j')}+\mathbf{s}_2^{(j')}\right)\right] = O\left(\frac{1}{M}\right) \\
& \implies Cov\left[\mathbf{Y}^2_{M,i}, \mathbf{Y}^2_{M,j}\right] = O\left(\frac{1}{M}\right).
\end{aligned}$$

Then by Lemma 1, $\mathbf{S}_{N,M} = \frac{\hat{\mathbf{G}}_w - \mathbb{E}[\hat{\mathbf{G}}_w]}{\sqrt{\mathbb{V}[\hat{\mathbf{G}}_w]}}$ converges in distribution to a standard normal random variable.

References

[1] K. R. Moon and A. O. Hero III, "Ensemble estimation of multivariate f-divergence," in *IEEE International Symposium on Information Theory*, pp. 356–360, 2014.

[2] K. R. Moon and A. O. Hero III, "Ensemble estimation of multivariate f-divergence," *CoRR*, vol. abs/1404.6230, 2014.

[3] K. Sricharan, R. Raich, and A. O. Hero, "Estimation of nonlinear functionals of densities with confidence," *IEEE Trans. Information Theory*, vol. 58, no. 7, pp. 4135–4159, 2012.

[4] K. Sricharan, D. Wei, and A. O. Hero III, "Ensemble estimators for multivariate entropy estimation," *IEEE Trans. Information Theory*, vol. 59, no. 7, pp. 4374–4388, 2013.

[5] K. Sricharan, *Neighborhood graphs for estimation of density functionals*. PhD thesis, Univ. Michigan, 2012.

[6] Y. Mack and M. Rosenblatt, "Multivariate k-nearest neighbor density estimates," *Journal of Multivariate Analysis*, vol. 9, no. 1, pp. 1–15, 1979.