Robust parameter estimation based on the $\mathcal{K}$-divergence: Supplementary Material

Koby Todros

Ben-Gurion University of the Negev

This supplementary material document is organized as follows. In Sections I-III, we provide proofs for Theorems 1-3, respectively. In Section In this supplementary material document we provide proofs for Theorems 1-3. In Section IV we show that the M$\mathcal{K}$DE is Fisher consistent. Finally, in Section V we derive its influence function.

I. Proof of Theorem 1

A. Non-negativity

In Lemma 1 below we show that under the assumption of a strictly positive kernel function

$$\mathcal{K}_h[G||F] = \mathcal{D}[\tilde{G}_h||\tilde{F}_h], \tag{S-1}$$

where $\mathcal{D}[\cdot||\cdot]$ denotes the KLD, $\tilde{G}_h$ and $\tilde{F}_h$ are probability distributions with density functions

$$\tilde{g}_h(\mathbf{r}) \triangleq g(\mathbf{r})\psi_G(\mathbf{r};h) \text{ and } \tilde{f}_h(\mathbf{r}) \triangleq f(\mathbf{r})\psi_F(\mathbf{r};h), \tag{S-2}$$

respectively, $\psi_G(\mathbf{r};h)$ is defined in Eq. (2) and $\psi_F(\mathbf{r};h) \triangleq \frac{(K_h*g)(\mathbf{r})}{\mathbb{E}[(K_h*g)(\mathbf{x});F]}$. We remark that the expectations $\mathbb{E}[(K_h*g)(\mathbf{x});G]$ in Eq. (2) and $\mathbb{E}[(K_h*g)(\mathbf{x});F]$ in $\psi_F(\mathbf{r};h)$ are strictly positive and finite. This is property follows from the boundedness and the strict positiveness of the kernel function $K(\cdot)$. By [s1, Th. 8.6.1] the KLD in (S-1) satisfies:

$$\mathcal{D}[\tilde{G}_h||\tilde{F}_h] \geq 0, \text{ where equality holds } \iff \tilde{G}_h = \tilde{F}_h. \tag{S-3}$$

In Lemma 2, stated below, we prove that

$$\tilde{G}_h = \tilde{F}_h \iff G = F. \tag{S-4}$$

Hence, by (S-1), (S-3) and (S-4) it follows that $\mathcal{K}_h[G||F] \geq 0$, where equality holds if and only if $G = F$. □

Lemma 1. *The equality in (S-1) holds under the assumption of a strictly positive kernel function.*

Proof. The $\mathcal{K}$-divergence (1) can be rewritten as:

$$\mathcal{K}_h[G||F] = \mathbb{E}\left[\psi_G(\mathbf{x},h)\log\frac{g(\mathbf{x})\psi_G(\mathbf{x},h)}{f(\mathbf{x})\psi_F(\mathbf{x},h)};G\right] - \mathbb{E}\left[\psi_G(\mathbf{x},h)\log\frac{\psi_G(\mathbf{x},h)}{\psi_F(\mathbf{x},h)};G\right] + \log\mathbb{E}\left[\psi_G(\mathbf{x},h);F\right], \tag{S-5}$$

where $\psi_G(\mathbf{r};h)$ and $\psi_F(\mathbf{r};h)$ are defined in Eq. (2) and below Eq. (S-2), respectively. Note that since the kernel function is strictly positive, then $\psi_G(\mathbf{r};h)$ and $\psi_F(\mathbf{r};h)$ are strictly positive as well, which implies that the quotient

$\psi_G(\mathbf{r},h)/\psi_F(\mathbf{r},h)$ is well defined. Also notice that $\tilde{g}_h(\mathbf{r}) \triangleq g(\mathbf{r})\psi_G(\mathbf{r};h)$ and $\tilde{f}_h(\mathbf{r}) \triangleq f(\mathbf{r})\psi_F(\mathbf{r};h)$ are non-negative functions that integrate to 1. Hence, these functions are viable density functions. Therefore, the expectation

$$\mathbb{E}\left[\psi_G(\mathbf{x},h)\log\frac{g(\mathbf{x})\psi_G(\mathbf{x},h)}{f(\mathbf{x})\psi_F(\mathbf{x},h)};G\right] = \mathbb{E}\left[\log\frac{\tilde{g}_h(\mathbf{x})}{\tilde{f}_h(\mathbf{x})};\tilde{G}_h\right] \triangleq \mathcal{D}[\tilde{G}_h||\tilde{F}_h], \tag{S-6}$$

where $\tilde{G}_h$ and $\tilde{F}_h$ are the probability distributions associated with $\tilde{g}_h(\mathbf{r})$ and $\tilde{f}_h(\mathbf{r})$, respectively. Additionally, using the definitions of $\psi_G(\mathbf{r};h)$ and $\psi_F(\mathbf{r};h)$, one can verify that

$$\mathbb{E}\left[\psi_G(\mathbf{x},h)\log\frac{\psi_G(\mathbf{x},h)}{\psi_F(\mathbf{x},h)};G\right] = -\log\mathbb{E}\left[\psi_G(\mathbf{x},h);F\right]. \tag{S-7}$$

Hence, the relation (S-1) follows directly from (S-5)-(S-7). □

Lemma 2. *The relation in (S-4) holds under the assumption of a strictly positive kernel function.*

Proof.

$$\begin{aligned}
\tilde{G}_h = \tilde{F}_h \quad &\Longleftrightarrow \quad \tilde{g}_h(\mathbf{r}) = \tilde{f}_h(\mathbf{r}) \;\; \lambda - a.e., \\
&\overset{(a)}{\Longleftrightarrow} \quad g(\mathbf{r})\psi_G(\mathbf{r};h) = f(\mathbf{r})\psi_F(\mathbf{r};h) \;\; \lambda - a.e., \\
&\overset{(b)}{\Longleftrightarrow} \quad \frac{g(\mathbf{r})}{\mathbb{E}[(K_h * g)(\mathbf{x});G]} = \frac{f(\mathbf{r})}{\mathbb{E}[(K_h * g)(\mathbf{x});F]} \;\; \lambda - a.e., \\
&\overset{(c)}{\Longleftrightarrow} \quad g(\mathbf{r}) = f(\mathbf{r}) \;\; \lambda - a.e., \\
&\Longleftrightarrow \quad G = F,
\end{aligned} \tag{S-8}$$

where (a) follows from (S-2) and (b) stems from the definitions of $\psi_G(\mathbf{r};h)$ and $\psi_F(\mathbf{r};h)$ and the strict-positiveness of the convolution $(K_h * g)(\mathbf{r})$, that arise from the assumption that the kernel function $K(\cdot)$ is strictly positive. We shall now prove the equivalence in (c). If $\frac{g(\mathbf{r})}{\mathbb{E}[(K_h*g)(\mathbf{x});G]} = \frac{f(\mathbf{r})}{\mathbb{E}[(K_h*g)(\mathbf{x});F]}$ $\lambda - a.e.$, then integration of both sides of the equality yields $\mathbb{E}[(K_h * g)(\mathbf{x});G] = \mathbb{E}[(K_h * g)(\mathbf{x});F]$, implying that $g(\mathbf{r}) = f(\mathbf{r})$ $\lambda - a.e$. The converse is trivial. □

B. Relation to the KLD

By Eqs. (1) and (2) it follows that

$$\lim_{h\to\infty}\mathcal{K}_h[G||F] = \frac{\lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})g(\mathbf{r})\log\frac{g(\mathbf{r})}{f(\mathbf{r})}d\lambda(\mathbf{r})}{\lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r})} + \log\frac{\lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})f(\mathbf{r})d\lambda(\mathbf{r})}{\lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r})}, \tag{S-9}$$

where $\eta_h(\mathbf{r}) \triangleq \int_{\mathbb{R}^p} K\left(\frac{\mathbf{r}-\boldsymbol{s}}{h}\right) g(\boldsymbol{s})d\lambda(\boldsymbol{s})$. Since the kernel function $K(\cdot)$ is bounded, continuous and strictly positive it follows from the dominated convergence theorem (DCT) [s11, Cor. 2.3.12] that

$$\lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r}) = \lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})f(\mathbf{r})d\lambda(\mathbf{r}) = K(0) > 0, \tag{S-10}$$

Additionally, if $\mathbb{E}[|\log\frac{g(\mathbf{x})}{f(\mathbf{x})}|;G]$ is finite, the DCT also implies that

$$\lim_{h\to\infty}\int_{\mathbb{R}^p}\eta_h(\mathbf{r})g(\mathbf{r})\log\frac{g(\mathbf{r})}{f(\mathbf{r})}d\lambda(\mathbf{r}) = K(0)\times\mathbb{E}\left[\log\frac{g(\mathbf{x})}{f(\mathbf{x})};G\right]. \tag{S-11}$$

Therefore, the relation in Eq. (3) follows directly from (S-9), (S-10), (S-11) and the definition of the KLD. □

C. Small bandwidth asymptotics

By Eqs. (1) and (2) it follows that

$$\lim_{h\to 0}\mathcal{K}_h[G||F]=\frac{\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})g(\mathbf{r})\log\frac{g(\mathbf{r})}{f(\mathbf{r})}d\lambda(\mathbf{r})}{\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r})}+\log\frac{\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})f(\mathbf{r})d\lambda(\mathbf{r})}{\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r})},\tag{S-12}$$

where $\zeta_h(\mathbf{r})\triangleq(K_h*g)(\mathbf{r})$. By the triangle and Cauchy-Schwartz's inequalities, it follows that:

$$\begin{aligned}\left|\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r})-\int_{\mathbb{R}^p}g^2(\mathbf{r})d\lambda(\mathbf{r})\right| &= \left|\int_{\mathbb{R}^p}(\zeta_h(\mathbf{r})-g(\mathbf{r}))g(\mathbf{r})d\lambda(\mathbf{r})\right| \\ &\leq \sqrt{\int_{\mathbb{R}^p}|\zeta_h(\mathbf{r})-g(\mathbf{r})|^2d\lambda(\mathbf{r})}\sqrt{\int_{\mathbb{R}^p}g^2(\mathbf{r})d\lambda(\mathbf{r})}.\end{aligned}\tag{S-13}$$

Therefore, since $K(\cdot)$ is integrable, $\int_{\mathbb{R}^p}K(\mathbf{r})d\lambda(\mathbf{r})=1$ and $g(\cdot)$ is square integrable, it follows from [s3, Th. 8.14] that

$$\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})g(\mathbf{r})d\lambda(\mathbf{r})=\int_{\mathbb{R}^p}g^2(\mathbf{r})d\lambda(\mathbf{r}).\tag{S-14}$$

Similarly, since $f(\cdot)$ is square integrable and $\mathbb{E}[|\log\frac{g(\mathbf{x})}{f(\mathbf{x})}|^2;G]<\infty$ it can be shown that

$$\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})f(\mathbf{r})d\lambda(\mathbf{r})=\int_{\mathbb{R}^p}g(\mathbf{r})f(\mathbf{r})d\lambda(\mathbf{r}).\tag{S-15}$$

and

$$\lim_{h\to 0}\int_{\mathbb{R}^p}\zeta_h(\mathbf{r})g(\mathbf{r})\log\frac{g(\mathbf{r})}{f(\mathbf{r})}d\lambda(\mathbf{r})=\int_{\mathbb{R}^p}g^2(\mathbf{r})\log\frac{g(\mathbf{r})}{f(\mathbf{r})}d\lambda(\mathbf{r}).\tag{S-16}$$

Hence, relation (4) follows directly from (S-12), (S-14)-(S-16) and the definition of $\bar{\psi}_G(\cdot)$ stated below (4). □

D. Invariance

Define the random vector:

$$\mathbf{y}\triangleq\mathbf{a}(\mathbf{x}),\tag{S-17}$$

where $\mathbf{a}:\mathbb{R}^p\to\mathbb{R}^p$ is a continuously differentiable one-to-one mapping. In the following, $G_{\mathbf{x}}$, $F_{\mathbf{x}}$, $g_{\mathbf{x}}(\cdot)$ and $f_{\mathbf{x}}(\cdot)$ will denote the probability distributions of $\mathbf{x}$ and their associate density functions. The probability distributions of $\mathbf{y}$ and their densities will be denoted by $G_{\mathbf{y}}$, $F_{\mathbf{y}}$, $g_{\mathbf{y}}(\cdot)$ and $f_{\mathbf{y}}(\cdot)$. In Lemma 3 below, we show that

$$\mathcal{K}_h[G_{\mathbf{y}}||F_{\mathbf{y}}]=\frac{\int_{\mathbb{R}^p}\zeta_{\mathbf{a},h}(\boldsymbol{t})g_{\mathbf{x}}(\boldsymbol{t})\log\frac{g_{\mathbf{x}}(\boldsymbol{t})}{f_{\mathbf{x}}(\boldsymbol{t})}d\lambda(\boldsymbol{t})}{\int_{\mathbb{R}^p}\zeta_{\mathbf{a},h}(\boldsymbol{t})g_{\mathbf{x}}(\boldsymbol{t})d\lambda(\boldsymbol{t})}+\log\frac{\int_{\mathbb{R}^p}\zeta_{\mathbf{a},h}(\boldsymbol{t})f_{\mathbf{x}}(\boldsymbol{t})d\lambda(\boldsymbol{t})}{\int_{\mathbb{R}^p}\zeta_{\mathbf{a},h}(\boldsymbol{t})g_{\mathbf{x}}(\boldsymbol{t})d\lambda(\boldsymbol{t})},\tag{S-18}$$

where

$$\zeta_{\mathbf{a},h}(\boldsymbol{t})\triangleq\int_{\mathbb{R}^p}K_h\left(\mathbf{a}(\boldsymbol{t})-\mathbf{a}(\boldsymbol{\tau})\right)g_{\mathbf{x}}(\boldsymbol{\tau})d\lambda(\boldsymbol{\tau}).\tag{S-19}$$

Hence, by Eqs. (1), (2), (S-18) and (S-19), it follows that the equality

$$\mathcal{K}_h[G_{\mathbf{y}}||F_{\mathbf{y}}]=\mathcal{K}_h[G_{\mathbf{x}}||F_{\mathbf{x}}]\tag{S-20}$$

holds when

$$K_h\left(\mathbf{a}(\boldsymbol{t})-\mathbf{a}(\boldsymbol{\tau})\right)=K_h\left(\boldsymbol{t}-\boldsymbol{\tau}\right).\tag{S-21}$$

Notice that the relation in (S-21) is satisfied when $\mathbf{a}(\mathbf{x}) = \mathbf{x} + \mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^p$ is a deterministic shift parameter. Additionally, (S-21) holds when $\mathbf{a}(\mathbf{x}) = \mathbf{U}\mathbf{x}$, such that $\mathbf{U}$ is a deterministic unitary matrix, and the kernel function is spherical, i.e., $K(\mathbf{x}) = S(\|\mathbf{x}\|)$, where $S(\cdot)$ is a strictly positive scalar function. □

Lemma 3. *The relation in (S-18) holds under the transformation in (S-17).*

Proof. Let $\mathbf{b}(\cdot)$ denote the inverse of $\mathbf{a}(\cdot)$. By the change of variables formulae [s4, Ch. 2.2.5], it follows that

$$g_{\mathbf{y}}(\mathbf{r}) = g_{\mathbf{x}}(\mathbf{b}(\mathbf{r}))|\det[\nabla \mathbf{b}(\mathbf{r})]| \quad \text{and} \quad f_{\mathbf{y}}(\mathbf{r}) = f_{\mathbf{x}}(\mathbf{b}(\mathbf{r}))|\det[\nabla \mathbf{b}(\mathbf{r})]|. \tag{S-22}$$

Since $\mathbf{a}(\cdot)$ is continuously differentiable, then also $\mathbf{b}(\cdot)$. Therefore, let $\boldsymbol{t} \triangleq \mathbf{b}(\mathbf{r})$. By [s11, Th. 4.4.6], it follows that

$$|\det[\nabla \mathbf{b}(\mathbf{r})]| d\lambda(\mathbf{r}) = d\lambda(\boldsymbol{t}). \tag{S-23}$$

The relation in (S-18) can now be easily verified using (1), (S-22) and (S-23). □

II. Proof of Theorem 2

By Eqs. (1) and (2), it follows that minimization of $\mathcal{K}_h[G||F_{\boldsymbol{\theta}}]$ over $\boldsymbol{\Theta}$ amounts to maximization of the deterministic objective function:

$$\bar{\mathcal{J}}_h(\boldsymbol{\theta}) \triangleq \mathbb{E}[\psi_G(\mathbf{x}, h) \log f(\mathbf{x}; \boldsymbol{\theta}); G] - \log \mathbb{E}[(K_h * g)(\mathbf{x}); F_{\boldsymbol{\theta}}]. \tag{S-24}$$

Therefore, under Assumption (A-2), $\bar{\mathcal{J}}_h(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_h^*$ (9). In Proposition 1, stated below, we show that under (A-1) and (A-3)-(A-5), the random objective function $\mathcal{J}_h(\boldsymbol{\theta})$ (7) converges uniformly in probability to $\bar{\mathcal{J}}_h(\boldsymbol{\theta})$ as $N \to \infty$. Furthermore, Assumptions (A-3) and (A-5) imply that $\mathcal{J}_h(\boldsymbol{\theta})$ must be continuous over the compact parameter space $\boldsymbol{\Theta}$ w.p.1. Hence, by [s5, Th. 3.4] we conclude that the convergence in (11) holds. □

Proposition 1. *Assume that conditions (A-1) and (A-3)-(A-5) are satisfied. Then,*

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\mathcal{J}_h(\boldsymbol{\theta}) - \bar{\mathcal{J}}_h(\boldsymbol{\theta})| \xrightarrow[N \to \infty]{p} 0. \tag{S-25}$$

Proof. Using Eqs. (7), (S-24) and the triangle inequality, one can verify that:

$$|\mathcal{J}_h(\boldsymbol{\theta}) - \bar{\mathcal{J}}_h(\boldsymbol{\theta})| \leq A(\boldsymbol{\theta}) + B(\boldsymbol{\theta}), \tag{S-26}$$

where

$$A(\boldsymbol{\theta}) \triangleq \left| \frac{\sum_{n=1}^{N} \sum_{m \neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m) \log f(\mathbf{x}_n; \boldsymbol{\theta})}{\sum_{n=1}^{N} \sum_{m \neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m)} - \frac{\mathbb{E}[K_h(\mathbf{x} - \mathbf{x}') \log f(\mathbf{x}; \boldsymbol{\theta}); G \times G]}{\mathbb{E}[K_h(\mathbf{x} - \mathbf{x}'); G \times G]} \right|, \tag{S-27}$$

$\mathbf{x}'$ is an independent copy of $\mathbf{x}$, such that the product $G \times G$ denotes their joint probability distribution,

$$B(\boldsymbol{\theta}) \triangleq |\log T(\boldsymbol{\theta})|, \tag{S-28}$$

$$T(\boldsymbol{\theta}) \triangleq \frac{1}{N} \sum_{n=1}^{N} \frac{c(\mathbf{x}_n; \boldsymbol{\theta})}{\mathbb{E}[c(\mathbf{x}; \boldsymbol{\theta}); G]} \tag{S-29}$$

and

$$c(\mathbf{r}; \boldsymbol{\theta}) \triangleq \mathbb{E}[K_h(\mathbf{x}'' - \mathbf{r}); F_{\boldsymbol{\theta}}]. \tag{S-30}$$

Note that since $K_h(\cdot)$ is strictly positive, the statistic $T(\boldsymbol{\theta})$ (S-29) must be strictly positive over $\boldsymbol{\Theta}$ w.p.1. In Lemmas 4 and 5, stated below, we show that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} A(\boldsymbol{\theta}) \xrightarrow[N\to\infty]{p} 0 \tag{S-31}$$

and

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} B(\boldsymbol{\theta}) \xrightarrow[N\to\infty]{p} 0. \tag{S-32}$$

Hence, the relation in (S-25) follows directly from (S-26), (S-31) and (S-32). □

Lemma 4. *Relation (S-31) holds under Assumptions (A-1) and (A-3).*

Proof. Using (S-27) and the triangle inequality, one can verify that

$$A(\boldsymbol{\theta}) = |A_1(\boldsymbol{\theta}) + A_2(\boldsymbol{\theta})| \leq |A_1(\boldsymbol{\theta})| + |A_2(\boldsymbol{\theta})|, \tag{S-33}$$

where

$$A_1(\boldsymbol{\theta}) \triangleq \frac{\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{m\neq n}^{N} Z_h(\mathbf{x}_n, \mathbf{x}_m; \boldsymbol{\theta}) - \mathbb{E}[Z_h(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}); G\times G]}{\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m)}, \tag{S-34}$$

$$A_2(\boldsymbol{\theta}) \triangleq \mathbb{E}[Z_h(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}); G\times G]\left(\frac{1}{\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m)} - \frac{1}{\mathbb{E}\left[K_h(\mathbf{x}-\mathbf{x}'); G\times G\right]}\right), \tag{S-35}$$

and

$$Z_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta}) \triangleq K_h(\mathbf{r} - \boldsymbol{s}) \log f(\mathbf{r}; \boldsymbol{\theta}). \tag{S-36}$$

Since $K_h(\cdot)$ is symmetric and bounded, it follows from [s8, Th. 5.4.A] that

$$\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m) \xrightarrow[N\to\infty]{w.p.1} \mathbb{E}\left[K_h(\mathbf{x}-\mathbf{x}'); G\times G\right]. \tag{S-37}$$

Furthermore, the series

$$\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{m\neq n}^{N} Z_h(\mathbf{x}_n, \mathbf{x}_m; \boldsymbol{\theta}) = \frac{2}{N(N-1)}\sum_{n=1}^{N}\sum_{m=n+1}^{N} W_h(\mathbf{x}_n, \mathbf{x}_m; \boldsymbol{\theta}), \tag{S-38}$$

where

$$W_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta}) \triangleq \frac{1}{2}\left(Z_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta}) + Z_h(\boldsymbol{s}, \mathbf{r}; \boldsymbol{\theta})\right) \tag{S-39}$$

is a symmetrized version of $Z_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta})$, i.e., $W_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta}) = W_h(\boldsymbol{s}, \mathbf{r}; \boldsymbol{\theta})$, and

$$\mathbb{E}[W_h(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}); G\times G] = \mathbb{E}[Z_h(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}); G\times G]. \tag{S-40}$$

We shall assume that the expectation

$$\mathbb{E}[|W_h(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})|; G\times G] < \infty \tag{S-41}$$

and that there exist functions $b : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}_+$ and $v : \mathbb{R}_+ \to \mathbb{R}_+$, such that $\mathbb{E}[b(\mathbf{x}, \mathbf{x}'); G\times G] < \infty$, $v(\cdot)$ is continuous at the origin, $v(0) = 0$ and for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, it holds that

$$|W_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta}) - W_h(\mathbf{r}, \boldsymbol{s}; \boldsymbol{\theta}')| \leq b(\mathbf{r}, \boldsymbol{s}) v(\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|) \quad \lambda\times\lambda - a.e. \tag{S-42}$$

Under these assumptions, it follows from [s9, Cor. 4.1] that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{2}{N(N-1)}\sum\nolimits_{n=1}^{N}\sum\nolimits_{m=n+1}^{N}W_h(\mathbf{x}_n,\mathbf{x}_m;\boldsymbol{\theta})-\mathbb{E}[W_h(\mathbf{x},\mathbf{x}';\boldsymbol{\theta});G\times G]\right|\xrightarrow[N\to\infty]{p}0. \tag{S-43}$$

and $\mathbb{E}[W_h(\mathbf{x},\mathbf{x}';\boldsymbol{\theta});G\times G]$ is continuous over $\boldsymbol{\Theta}$. Hence, it follows from (S-38), (S-40) and (S-43) that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N(N-1)}\sum\nolimits_{n=1}^{N}\sum\nolimits_{m=n+1}^{N}Z_h(\mathbf{x}_n,\mathbf{x}_m;\boldsymbol{\theta})-\mathbb{E}[Z_h(\mathbf{x},\mathbf{x}';\boldsymbol{\theta});G\times G]\right|\xrightarrow[N\to\infty]{p}0. \tag{S-44}$$

and $\mathbb{E}[Z_h(\mathbf{x},\mathbf{x}';\boldsymbol{\theta});G\times G]$ is continuous over $\boldsymbol{\Theta}$. Therefore, by (S-34), (S-35), (S-37), (S-44), the continuity of $\mathbb{E}[Z_h(\mathbf{x},\mathbf{x}';\boldsymbol{\theta});G\times G]$, the compactness of $\boldsymbol{\Theta}$ and Mann-Wald's Theorem , we conclude that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|A_1(\boldsymbol{\theta})|\xrightarrow[N\to\infty]{p}0\quad\text{and}\quad\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|A_2(\boldsymbol{\theta})|\xrightarrow[N\to\infty]{p}0. \tag{S-45}$$

Hence, the relation in (S-31) follows directly from (S-33) and (S-45).

To complete the proof, we need to show now that the conditions in (S-41) and (S-42) hold. First, using (S-36) and (S-40) one can easily verify that since the kernel function $K_h(\cdot)$ is strictly positive and bounded, the condition in (S-41) is satisfied under Assumption (A-4). Additionally, by (S-36) and (S-39) it follows that

$$\begin{aligned}|W_h(\mathbf{r},\boldsymbol{s};\boldsymbol{\theta})-W_h(\mathbf{r},\boldsymbol{s};\boldsymbol{\theta}')| &= \frac{1}{2}K_h(\mathbf{r}-\boldsymbol{s})|\log\rho(\mathbf{r},\boldsymbol{\theta},\boldsymbol{\theta}')+\log\rho(\boldsymbol{s},\boldsymbol{\theta},\boldsymbol{\theta}')|\\ &\le \frac{1}{2}K_h(\mathbf{r}-\boldsymbol{s})|\log\rho(\mathbf{r},\boldsymbol{\theta},\boldsymbol{\theta}')|+\frac{1}{2}K_h(\mathbf{r}-\boldsymbol{s})|\log\rho(\boldsymbol{s},\boldsymbol{\theta},\boldsymbol{\theta}')|\\ &\le \frac{1}{2}K_h(\mathbf{r}-\boldsymbol{s})(u(\mathbf{r})+u(\boldsymbol{s}))v(\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|),\end{aligned} \tag{S-46}$$

where $\rho(\mathbf{r},\boldsymbol{\theta},\boldsymbol{\theta}')\triangleq f(\mathbf{r};\boldsymbol{\theta})/f(\mathbf{r};\boldsymbol{\theta}')$ and the last inequality in (S-46) follows from Assumption (A-5). Hence, the function $b(\cdot,\cdot)$ in (S-46) is given by:

$$b(\mathbf{r},\boldsymbol{s})=\frac{1}{2}K_h(\mathbf{r}-\boldsymbol{s})(u(\mathbf{r})+u(\boldsymbol{s})). \tag{S-47}$$

We need to show that in this case the expectation $\mathbb{E}[b(\mathbf{x},\mathbf{x}');G\times G]$ is indeed finite. In Assumption (A-5) it is stated that $\mathbb{E}[\psi_G(\mathbf{x},h)u(\mathbf{x});G]<\infty$. Hence, by Eq. (2) it follows that $\mathbb{E}[b(\mathbf{x},\mathbf{x}');G\times G]=\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')u(\mathbf{x});G\times G]<\infty$ when $\mathbb{E}[\psi_G(\mathbf{x},h)u(\mathbf{x});G]<\infty$. □

Lemma 5. *Relation (S-32) holds under assumptions Assumptions (A-1) and (A-3)*

Proof. Let $\mu_T(\boldsymbol{\theta})\triangleq\mathbb{E}[T(\boldsymbol{\theta});P_{T(\boldsymbol{\theta})}]$. By (S-29), and the assumption that $\{\mathbf{x}_n\}_{n=1}^N$ are identically distributed, we have that $\mu_T(\boldsymbol{\theta})=1$. Hence, the mean-value theorem [s6, Th. 3.4] implies that

$$\log T(\boldsymbol{\theta})=\log\mu_T(\boldsymbol{\theta})+\left.\frac{d\log T}{dT}\right|_{T=T^*(\boldsymbol{\theta})}(T(\boldsymbol{\theta})-\mu_T(\boldsymbol{\theta}))=\frac{T(\boldsymbol{\theta})-1}{T^*(\boldsymbol{\theta})}, \tag{S-48}$$

where $T^*(\boldsymbol{\theta})$ lies in the line segment connecting $T(\boldsymbol{\theta})$ and $\mu_T(\boldsymbol{\theta})=1$. Note that since $T(\boldsymbol{\theta})>0$ w.p. 1, then $T^*(\boldsymbol{\theta})$ must be strictly positive w.p. 1. Therefore, by (S-28) and (S-48)

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}B(\boldsymbol{\theta})\le\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|T(\boldsymbol{\theta})-1|\times(\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}T^*(\boldsymbol{\theta}))^{-1} \tag{S-49}$$

where $(\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}T^*(\boldsymbol{\theta}))^{-1}<\infty$ w.p.1. Using (S-29), one can verify that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|T(\boldsymbol{\theta})-1|\le\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\frac{1}{\mathbb{E}[c(\mathbf{x};\boldsymbol{\theta});G]}\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}c(\mathbf{x}_n;\boldsymbol{\theta})-\mathbb{E}[c(\mathbf{x};\boldsymbol{\theta});G]\right|. \tag{S-50}$$

Under Assumption (A-3), the function $c(\mathbf{r};\boldsymbol{\theta})$ (S-30) is continuous over $\boldsymbol{\Theta}$ λ-a.e. Additionally, since the kernel function $K_h(\cdot)$ is bounded, it follows from (S-30) that $c(\mathbf{r};\boldsymbol{\theta})$ is bounded over $\mathbb{R}^p \times \boldsymbol{\Theta}$. Hence, since $\{\mathbf{x}_n\}_{n=1}^N$ are i.i.d. samples from G, it follows from the uniform weak law of large numbers [s7] that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{N}\sum_{n=1}^{N} c(\mathbf{x}_n;\boldsymbol{\theta}) - \mathbb{E}[c(\mathbf{x};\boldsymbol{\theta});G] \right| \xrightarrow[N\to\infty]{p} 0. \tag{S-51}$$

and that $\mathbb{E}[c(\mathbf{x};\boldsymbol{\theta});G]$ is continuous over the parameter space $\boldsymbol{\Theta}$. The latter continuity property, and Assumption (A-1) (compactness of $\boldsymbol{\Theta}$) imply that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{\mathbb{E}[c(\mathbf{x};\boldsymbol{\theta});G]} < \infty. \tag{S-52}$$

Hence, by (S-50)-(S-52) we conclude that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |T(\boldsymbol{\theta}) - 1| \xrightarrow[N\to\infty]{p} 0. \tag{S-53}$$

We now analyze the convergence of $\inf\limits_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} T^*(\boldsymbol{\theta})$ in (S-49). Note that

$$|\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} T^*(\boldsymbol{\theta}) - 1| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |T^*(\boldsymbol{\theta}) - 1| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |T(\boldsymbol{\theta}) - 1|, \tag{S-54}$$

where the second inequality in (S-54) follows from the property that $T^*(\boldsymbol{\theta})$ lies in the line segment connecting $T(\boldsymbol{\theta})$ and $\mu_T(\boldsymbol{\theta}) = 1$. Hence, by (S-53) we conclude that

$$\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} T^*(\boldsymbol{\theta}) \xrightarrow[N\to\infty]{p} 1. \tag{S-55}$$

Therefore, the relation in (S-32) follows directly from (S-49), (S-53), (S-55) and Mann-Wald's Theorem [s10]. □

III. Proof of Theorem 3

Remark 1. *Throughout the proof we shall assume that integration and differentiation operations can be interchanged. Using [s12, Th. 2.40] it can be shown that this assumption is justified under conditions (B-3), (B-4) and the boundedness of the kernel function $K_h(\cdot)$.*

By Assumptions (B-1) and (B-2) the estimator $\hat{\boldsymbol{\theta}}_h$ is weakly consistent and the estimand $\boldsymbol{\theta}_h^*$ lies in the interior of $\boldsymbol{\Theta}$, which is assumed to be compact. Therefore, $\hat{\boldsymbol{\theta}}_h$ lies in an open neighbourhood $\mathcal{U} \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_h^*$ with sufficiently high probability as N gets large, i.e., it does not lie on the boundary of $\boldsymbol{\Theta}$. Hence, $\hat{\boldsymbol{\theta}}_h$ is a maximum point of the objective function $\mathcal{J}_h(\boldsymbol{\theta})$ (7) whose gradient satisfies:

$$\nabla \mathcal{J}_h(\hat{\boldsymbol{\theta}}_h) = \mathbf{0}. \tag{S-56}$$

Using (7) one can verify that

$$\nabla \mathcal{J}_h(\boldsymbol{\theta}) = b^{-1}(\boldsymbol{\theta})\mathbf{a}(\boldsymbol{\theta}), \tag{S-57}$$

where

$$\mathbf{a}(\boldsymbol{\theta}) \triangleq \frac{1}{(N-1)N^2} \sum\nolimits_{n=1}^{N} \sum\nolimits_{m\neq n}^{N} \sum\nolimits_{l=1}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m)\mathbf{h}(\mathbf{x}_n, \mathbf{x}_l;\boldsymbol{\theta}), \tag{S-58}$$

$$b(\boldsymbol{\theta}) \triangleq \frac{1}{(N-1)N^2} \sum\nolimits_{n=1}^{N} \sum\nolimits_{m\neq n}^{N} \sum\nolimits_{l=1}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m) d(\mathbf{x}_l;\boldsymbol{\theta}), \tag{S-59}$$

$$\mathbf{h}(\mathbf{r},\boldsymbol{t};\boldsymbol{\theta}) \triangleq \mathbf{q}(\mathbf{r};\boldsymbol{\theta}) d(\boldsymbol{t};\boldsymbol{\theta}) - \mathbf{w}(\boldsymbol{t};\boldsymbol{\theta}), \tag{S-60}$$

$$d(\boldsymbol{t};\boldsymbol{\theta}) \triangleq \mathbb{E}[K_h(\mathbf{x}-\boldsymbol{t});F_{\boldsymbol{\theta}}], \tag{S-61}$$

$$\mathbf{w}(\boldsymbol{t};\boldsymbol{\theta}) \triangleq \mathbb{E}[K_h(\mathbf{x}-\boldsymbol{t})\mathbf{q}(\mathbf{x};\boldsymbol{\theta});F_{\boldsymbol{\theta}}] \tag{S-62}$$

and $\mathbf{q}(\boldsymbol{\theta})$ is the score function defined in Eq. (17). Therefore, by the strict positivity of the kernel function $K_h(\cdot)$, it follows that the equality in (S-56) is equivalent to

$$\mathbf{a}(\hat{\boldsymbol{\theta}}_h) = \mathbf{0}. \tag{S-63}$$

Using Assumptions (B-3), (B-4) and the dominated convergence Theorem [s11, Cor. 2.3.12], one can verify that $\mathbf{a}(\boldsymbol{\theta})$ is continuous over $\boldsymbol{\Theta}$ w.p.1. Hence, the mean-value theorem [s6, Th. 3.4] implies that

$$\mathbf{0} = \mathbf{a}(\hat{\boldsymbol{\theta}}_h) = \mathbf{a}(\boldsymbol{\theta}^*_h) + \mathbf{F}(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}^*_h), \tag{S-64}$$

where

$$\mathbf{F}(\boldsymbol{\theta}) \triangleq \frac{d\mathbf{a}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \frac{1}{(N-1)N^2}\sum\nolimits_{n=1}^{N}\sum\nolimits_{m\neq n}^{N}\sum\nolimits_{l=1}^{N} K_h(\mathbf{x}_n-\mathbf{x}_m)\frac{d\mathbf{h}(\mathbf{x}_n,\mathbf{x}_l;\boldsymbol{\theta})}{d\boldsymbol{\theta}} \tag{S-65}$$

and $\tilde{\boldsymbol{\theta}}$ lies in the line segment connecting $\hat{\boldsymbol{\theta}}_h$ and $\boldsymbol{\theta}^*_h$.

In Proposition 2, stated below, we show that

$$\mathbf{F}(\tilde{\boldsymbol{\theta}}) \xrightarrow[N\to\infty]{p} \eta(\boldsymbol{\theta}^*_h,h)\mathbf{C}(\boldsymbol{\theta}^*_h,h), \tag{S-66}$$

where

$$\eta(\boldsymbol{\theta},h) \triangleq \mathbb{E}[\psi_G(\mathbf{x},h);F_{\boldsymbol{\theta}}] \times \mathbb{E}^2[K_h(\mathbf{x}-\mathbf{x}');G\times G] \tag{S-67}$$

and the matrix function $\mathbf{C}(\boldsymbol{\theta},h)$, defined in Eq. (16), is non-singular by assumption. Note that strict positivity and finiteness of $\eta(\boldsymbol{\theta}^*_h,h)$ follows from the assumption that the kernel function $K_h(\cdot)$ is strictly positive and bounded. Hence, by Mann-Wald's Theorem [s10]

$$\mathbf{F}^{-1}(\tilde{\boldsymbol{\theta}}) \xrightarrow[N\to\infty]{p} \eta^{-1}(\boldsymbol{\theta}^*_h,h)\mathbf{C}^{-1}(\boldsymbol{\theta}^*_h,h), \tag{S-68}$$

which implies that $\mathbf{F}(\tilde{\boldsymbol{\theta}})$ is invertible with sufficiently high probability as N gets large. Therefore, by (S-64) the equality

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}^*_h) = -\mathbf{F}^{-1}(\tilde{\boldsymbol{\theta}})\sqrt{N}\mathbf{a}(\boldsymbol{\theta}^*_h) \tag{S-69}$$

holds with sufficiently large probability as N gets large. Furthermore, by Proposition 3 stated below

$$\sqrt{N}\mathbf{a}(\boldsymbol{\theta}^*_h) \xrightarrow[N\to\infty]{d} \mathcal{N}(\mathbf{0},\eta^2(\boldsymbol{\theta}^*_h,h)\mathbf{D}(\boldsymbol{\theta}^*_h,h)), \tag{S-70}$$

where $\mathbf{D}(\boldsymbol{\theta},h)$ is defined in Eq. (18). Thus, by (S-68)-(S-70) and Slutskey's Theorem, the relation in (14) holds. □

Proposition 2. *The relation in (S-66) holds under assumptions (B-1)-(B-8).*

Proof. First, in Lemma 6 stated below we prove that

$$\mathbf{F}(\tilde{\boldsymbol{\theta}}) \xrightarrow[N\to\infty]{p} \bar{\mathbf{F}}(\boldsymbol{\theta}^*_h), \tag{S-71}$$

where

$$\bar{\mathbf{F}}(\boldsymbol{\theta}) \triangleq \mathbb{E}\left[K_h(\mathbf{x}-\mathbf{x}')\frac{d\mathbf{h}(\mathbf{x},\mathbf{x}'';\boldsymbol{\theta})}{d\boldsymbol{\theta}};G\times G\times G\right]. \tag{S-72}$$

Next, in Lemma 8, we show that

$$\bar{\mathbf{F}}(\boldsymbol{\theta}_h^*) = \eta(\boldsymbol{\theta}_h^*, h)\mathbf{C}(\boldsymbol{\theta}_h^*, h). \tag{S-73}$$

□

Lemma 6. *The relation in (S-71) holds under assumptions (B-1)-(B-8).*

Proof. Notice that

$$\begin{aligned} \|\mathbf{F}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\boldsymbol{\theta}_h^*)\| &= \|\mathbf{F}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\tilde{\boldsymbol{\theta}}) + \bar{\mathbf{F}}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\boldsymbol{\theta}_h^*)\| \\ &\leq \|\mathbf{F}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\tilde{\boldsymbol{\theta}})\| + \|\bar{\mathbf{F}}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\boldsymbol{\theta}_h^*)\| \\ &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{F}(\boldsymbol{\theta}) - \bar{\mathbf{F}}(\boldsymbol{\theta})\| + \|\bar{\mathbf{F}}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\boldsymbol{\theta}_h^*)\|. \end{aligned} \tag{S-74}$$

In Lemma 7, stated below, we show that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{F}(\boldsymbol{\theta}) - \bar{\mathbf{F}}(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{p} 0 \tag{S-75}$$

and that $\bar{\mathbf{F}}(\boldsymbol{\theta})$ is continuous over $\boldsymbol{\Theta}$. Now, recall that $\tilde{\boldsymbol{\theta}}$ lies in the line segment connecting $\hat{\boldsymbol{\theta}}_h$ and $\boldsymbol{\theta}_h^*$. Furthermore, by Assumption (B-1) we have that $\hat{\boldsymbol{\theta}}_h \xrightarrow[N\to\infty]{p} \boldsymbol{\theta}_h^*$. Therefore, since $\bar{\mathbf{F}}(\boldsymbol{\theta})$ is continuous, it follows from Mann-Wald's Theorem that

$$\|\bar{\mathbf{F}}(\tilde{\boldsymbol{\theta}}) - \bar{\mathbf{F}}(\boldsymbol{\theta}_h^*)\| \xrightarrow[N\to\infty]{p} 0. \tag{S-76}$$

Hence, the relation in (S-71) follows directly from (S-74)-(S-76). □

Lemma 7. *The relation in (S-75) holds under Assumptions (B-2)-(B-8).*

Proof. Using (S-60) and (S-65), one can verify that $\mathbf{F}(\boldsymbol{\theta})$ can be written as:

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{A}_1(\boldsymbol{\theta})\mathbf{A}_2(\boldsymbol{\theta}) + \mathbf{A}_3(\boldsymbol{\theta})\mathbf{A}_4(\boldsymbol{\theta}) - \mathbf{A}_5\mathbf{A}_6(\boldsymbol{\theta}), \tag{S-77}$$

where

$$\begin{aligned} \mathbf{A}_1(\boldsymbol{\theta}) &\triangleq \frac{1}{(N-1)N}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m)\mathbf{H}(\mathbf{x}_n;\boldsymbol{\theta}), \\ \mathbf{A}_2(\boldsymbol{\theta}) &\triangleq \frac{1}{N}\sum_{l=1}^{N} d(\mathbf{x}_l;\boldsymbol{\theta}), \\ \mathbf{A}_3(\boldsymbol{\theta}) &\triangleq \frac{1}{(N-1)N}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m)\mathbf{q}(\mathbf{x}_n;\boldsymbol{\theta}), \\ \mathbf{A}_4(\boldsymbol{\theta}) &\triangleq \frac{1}{N}\sum_{l=1}^{N} \mathbf{w}^T(\mathbf{x}_l;\boldsymbol{\theta}), \\ \mathbf{A}_5 &\triangleq \frac{1}{(N-1)N}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n - \mathbf{x}_m), \\ \mathbf{A}_6(\boldsymbol{\theta}) &\triangleq \frac{1}{N}\sum_{l=1}^{N} \boldsymbol{\Gamma}(\mathbf{x}_l;\boldsymbol{\theta}), \end{aligned} \tag{S-78}$$

$\mathbf{q}(\mathbf{r};\boldsymbol{\theta})$ and $\mathbf{H}(\mathbf{r};\boldsymbol{\theta})$ are the gradient and Hessian, respectively, of the log-likelihood function defined in Eq. (17), $d(\mathbf{r};\boldsymbol{\theta})$ and $\mathbf{w}(\mathbf{r};\boldsymbol{\theta})$ are given in Eqs. (S-61) and (S-62), respectively, and $\boldsymbol{\Gamma}(\mathbf{r};\boldsymbol{\theta}) \triangleq \frac{\partial \mathbf{w}(\mathbf{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. Additionally, by (S-60) and (S-72), it follows that

$$\bar{\mathbf{F}}(\boldsymbol{\theta}) = \bar{A}_1(\boldsymbol{\theta})\bar{A}_2(\boldsymbol{\theta}) + \bar{A}_3(\boldsymbol{\theta})\bar{A}_4(\boldsymbol{\theta}) - \bar{A}_5\bar{A}_6(\boldsymbol{\theta}), \tag{S-79}$$

where

$$\begin{aligned}
\bar{\mathbf{A}}_1(\boldsymbol{\theta}) &\triangleq \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{H}(\mathbf{x};\boldsymbol{\theta}); G\times G], \\
\bar{\mathbf{A}}_2(\boldsymbol{\theta}) &\triangleq \mathbb{E}[d(\mathbf{x};\boldsymbol{\theta}); G], \\
\bar{\mathbf{A}}_3(\boldsymbol{\theta}) &\triangleq \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{q}(\mathbf{x};\boldsymbol{\theta}); G\times G], \\
\bar{\mathbf{A}}_4(\boldsymbol{\theta}) &\triangleq \mathbb{E}^T[\mathbf{w}(\mathbf{x};\boldsymbol{\theta}); G], \\
\bar{\mathbf{A}}_5 &\triangleq \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}'); G\times G], \\
\bar{\mathbf{A}}_6(\boldsymbol{\theta}) &\triangleq \mathbb{E}[\boldsymbol{\Gamma}(\mathbf{x};\boldsymbol{\theta}); G].
\end{aligned} \tag{S-80}$$

Therefore, using (S-77), (S-79), the identity

$$\mathbf{AB}-\mathbf{CD} = (\mathbf{A}-\mathbf{C})(\mathbf{B}-\mathbf{D}) + (\mathbf{A}-\mathbf{C})\mathbf{D} + \mathbf{C}(\mathbf{B}-\mathbf{D}),$$

that holds for matrices pairs $(\mathbf{A},\mathbf{C})$ and $(\mathbf{B},\mathbf{D})$ with identical within pair dimensions, and the triangle inequality, one can verify that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{F}(\boldsymbol{\theta}) - \bar{\mathbf{F}}(\boldsymbol{\theta})\| \leq B_1 + B_2 + B_3, \tag{S-81}$$

where

$$\begin{aligned}
B_1 &\triangleq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_1(\boldsymbol{\theta}) - \bar{\mathbf{A}}_1(\boldsymbol{\theta})\| \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_2(\boldsymbol{\theta}) - \bar{\mathbf{A}}_2(\boldsymbol{\theta})\| \\
&+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_1(\boldsymbol{\theta}) - \bar{\mathbf{A}}_1(\boldsymbol{\theta})\| \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\bar{\mathbf{A}}_2(\boldsymbol{\theta})\| \\
&+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\bar{\mathbf{A}}_1(\boldsymbol{\theta})\| \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_2(\boldsymbol{\theta}) - \bar{\mathbf{A}}_2(\boldsymbol{\theta})\|,
\end{aligned} \tag{S-82}$$

$$\begin{aligned}
B_2 &\triangleq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_2(\boldsymbol{\theta}) - \bar{\mathbf{A}}_2(\boldsymbol{\theta})\| \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_3(\boldsymbol{\theta}) - \bar{\mathbf{A}}_3(\boldsymbol{\theta})\| \\
&+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_2(\boldsymbol{\theta}) - \bar{\mathbf{A}}_2(\boldsymbol{\theta})\| \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\bar{\mathbf{A}}_3(\boldsymbol{\theta})\| \\
&+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\bar{\mathbf{A}}_2(\boldsymbol{\theta})\| \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_3(\boldsymbol{\theta}) - \bar{\mathbf{A}}_3(\boldsymbol{\theta})\|
\end{aligned} \tag{S-83}$$

and

$$\begin{aligned}
B_3 &\triangleq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_4(\boldsymbol{\theta}) - \bar{\mathbf{A}}_4(\boldsymbol{\theta})\| \times \|\mathbf{A}_5 - \bar{\mathbf{A}}_5\| \\
&+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_4(\boldsymbol{\theta}) - \bar{\mathbf{A}}_4(\boldsymbol{\theta})\| \times \|\bar{\mathbf{A}}_5\| \\
&+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\bar{\mathbf{A}}_4(\boldsymbol{\theta})\| \times \|\mathbf{A}_5 - \bar{\mathbf{A}}_5\|.
\end{aligned} \tag{S-84}$$

By the definitions of $\mathbf{A}_2(\boldsymbol{\theta})$, $\mathbf{A}_4(\boldsymbol{\theta})$, $\mathbf{A}_6(\boldsymbol{\theta})$ in (S-78), the definitions of $\bar{\mathbf{A}}_2(\boldsymbol{\theta})$, $\bar{\mathbf{A}}_4(\boldsymbol{\theta})$, $\bar{\mathbf{A}}_6(\boldsymbol{\theta})$ in (S-80), Assumptions (B-3), (B-4) and the uniform weak law of large numbers [s7], it follows that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_2(\boldsymbol{\theta}) - \bar{\mathbf{A}}_2(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{p} 0, \;\; \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_4(\boldsymbol{\theta}) - \bar{\mathbf{A}}_4(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{p} 0, \;\; \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_6(\boldsymbol{\theta}) - \bar{\mathbf{A}}_6(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{p} 0 \tag{S-85}$$

and $\mathbf{A}_2(\boldsymbol{\theta})$, $\mathbf{A}_4(\boldsymbol{\theta})$ and $\mathbf{A}_6(\boldsymbol{\theta})$ are continuous over $\boldsymbol{\Theta}$. Furthermore, since the kernel function $K_h(\cdot)$ is bounded and symmetric, it follows from [s8, Th. 5.4.A] that

$$\|\mathbf{A}_5 - \bar{\mathbf{A}}_5\| \xrightarrow[N\to\infty]{p} 0. \tag{S-86}$$

Now, similarly to the proof of equality (S-43), in Lemma 4, it can be shown that when Assumptions (B-3)-(B-8) are satisfied it follows from [s9, Cor. 4.1] that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_1(\boldsymbol{\theta}) - \bar{\mathbf{A}}_1(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{p} 0, \;\; \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}_3(\boldsymbol{\theta}) - \bar{\mathbf{A}}_3(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{p} 0, \tag{S-87}$$

and $\mathbf{A}_1(\boldsymbol{\theta})$ and $\mathbf{A}_3(\boldsymbol{\theta})$ are continuous over $\boldsymbol{\Theta}$. Therefore, the relation in (S-75) follows directly from (S-81)-(S-87), the continuities of $\mathbf{A}_1(\boldsymbol{\theta})$, $\mathbf{A}_2(\boldsymbol{\theta})$, $\mathbf{A}_3(\boldsymbol{\theta})$, $\mathbf{A}_4(\boldsymbol{\theta})$ and $\mathbf{A}_6(\boldsymbol{\theta})$ and the compactness of $\boldsymbol{\Theta}$ that follows from Assumption (B-2). Additionally, by (S-77), the continuity of $\mathbf{F}(\boldsymbol{\theta})$ is a consequence of the continuities of $\mathbf{A}_1(\boldsymbol{\theta})$, $\mathbf{A}_2(\boldsymbol{\theta})$, $\mathbf{A}_3(\boldsymbol{\theta})$, $\mathbf{A}_4(\boldsymbol{\theta})$ and $\mathbf{A}_6(\boldsymbol{\theta})$. □

Lemma 8. *The relation in (S-73) holds.*

Proof. Using Eqs. (S-72), (S-79), (S-80) and the definition of $\psi_G(\mathbf{r}, h)$ in Eq. (2), one can verify that

$$\begin{aligned} \bar{\mathbf{F}}(\boldsymbol{\theta}) &= \Big(\mathbb{E}\left[\psi_G(\mathbf{x}, h)\mathbf{H}(\mathbf{x}; \boldsymbol{\theta}); G\right] \times \mathbb{E}\left[\psi_G(\mathbf{x}, h); F_{\boldsymbol{\theta}}\right] \\ &+ \mathbb{E}\left[\psi_G(\mathbf{x}, h)\mathbf{q}(\mathbf{x}; \boldsymbol{\theta}); G\right] \times \mathbb{E}\left[\psi_G(\mathbf{x}, h)\mathbf{q}^T(\mathbf{x}; \boldsymbol{\theta}); F_{\boldsymbol{\theta}}\right] \\ &- \mathbb{E}\left[\psi_G(\mathbf{x}, h)\frac{\nabla^2_{\boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta})}{f(\mathbf{x}; \boldsymbol{\theta})}; F_{\boldsymbol{\theta}}\right] \Big) \mathbb{E}^2[K_h(\mathbf{x} - \mathbf{x}'); G \times G]. \end{aligned} \tag{S-88}$$

Notice that

$$\frac{\nabla^2_{\boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta})}{f(\mathbf{x}; \boldsymbol{\theta})} = \mathbf{H}(\mathbf{x}; \boldsymbol{\theta}) + \mathbf{q}(\mathbf{x}; \boldsymbol{\theta})\mathbf{q}^T(\mathbf{x}; \boldsymbol{\theta}). \tag{S-89}$$

Additionally, by (S-99) we have that $\mathbb{E}[K_h(\mathbf{x} - \mathbf{x}')\mathbf{h}(\mathbf{x}, \mathbf{x}''; \boldsymbol{\theta}^*_h)); G \times G \times G] = \mathbf{0}$. Hence, using (S-60)-(S-62), and the definition of $\psi_G(\mathbf{r}, h)$ in Eq. (2), one can verify that

$$\mathbb{E}\left[\psi_G(\mathbf{x}, h)\mathbf{q}(\mathbf{x}; \boldsymbol{\theta}^*_h); F_{\boldsymbol{\theta}^*_h}\right] = \mathbb{E}\left[\psi_G(\mathbf{x}, h)\mathbf{q}(\mathbf{x}; \boldsymbol{\theta}^*_h); G\right] \times \mathbb{E}\left[\psi_G(\mathbf{x}, h); F_{\boldsymbol{\theta}^*_h}\right]. \tag{S-90}$$

Using again the definition of $\psi_G(\mathbf{r}, h)$ in Eq. (2), it can be easily shown that for any scalar function $v(\cdot)$, such that the expectation $\mathbb{E}\left[\psi_G(\mathbf{x}, h)v(\mathbf{x}); F_{\boldsymbol{\theta}}\right]$ is finite, it holds that

$$\frac{\mathbb{E}\left[\psi_G(\mathbf{x}, h)v(\mathbf{x}); F_{\boldsymbol{\theta}}\right]}{\mathbb{E}\left[\psi_G(\mathbf{x}, h); F_{\boldsymbol{\theta}}\right]} = \mathbb{E}\left[\psi_F(\mathbf{x}, h)v(\mathbf{x}); F_{\boldsymbol{\theta}}\right], \tag{S-91}$$

where $\psi_F(\cdot, \cdot)$ is defined below Eq. (16). Therefore, by (S-92) and (S-91) we conclude that

$$\mathbb{E}\left[\psi_G(\mathbf{x}, h)\mathbf{q}(\mathbf{x}; \boldsymbol{\theta}^*_h); G\right] = \mathbb{E}\left[\psi_F(\mathbf{x}, h)\mathbf{q}(\mathbf{x}; \boldsymbol{\theta}^*_h); F_{\boldsymbol{\theta}^*_h}\right]. \tag{S-92}$$

Hence, relations (S-88)-(S-92) imply that $\bar{\mathbf{F}}(\boldsymbol{\theta}^*_h) = \eta(\boldsymbol{\theta}^*_h, h)\mathbf{C}(\boldsymbol{\theta}^*_h, h)$. □

Proposition 3. *The relation in (S-70) holds under conditions (B-2)-(B-5).*

Proof. Using (S-58), one can verify that

$$\sqrt{N}\mathbf{a}(\boldsymbol{\theta}) = \frac{N-2}{N}\sqrt{N}\mathbf{a}_1(\boldsymbol{\theta}) + \frac{1}{\sqrt{N}}\mathbf{a}_2(\boldsymbol{\theta}) + \frac{1}{\sqrt{N}}\mathbf{a}_3(\boldsymbol{\theta}), \tag{S-93}$$

where

$$\mathbf{a}_1(\boldsymbol{\theta}) \triangleq \frac{1}{(N-2)(N-1)N}\sum_{n=1}^{N}\sum_{m\neq n}^{N}\sum_{l\neq n,m}^{N} K_h(\mathbf{x}_n-\mathbf{x}_m)\mathbf{h}(\mathbf{x}_n,\mathbf{x}_l;\boldsymbol{\theta}), \tag{S-94}$$

$$\mathbf{a}_2(\boldsymbol{\theta}) \triangleq \frac{1}{(N-1)N}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n-\mathbf{x}_m)\mathbf{h}(\mathbf{x}_n,\mathbf{x}_n;\boldsymbol{\theta}), \tag{S-95}$$

$$\mathbf{a}_3(\boldsymbol{\theta}) \triangleq \frac{1}{(N-1)N}\sum_{n=1}^{N}\sum_{m\neq n}^{N} K_h(\mathbf{x}_n-\mathbf{x}_m)\mathbf{h}(\mathbf{x}_n,\mathbf{x}_m;\boldsymbol{\theta}). \tag{S-96}$$

First, we shall analyze convergence in probability of $\mathbf{a}_2(\boldsymbol{\theta})$ and $\mathbf{a}_3(\boldsymbol{\theta})$. Using Assumptions (B-4), (B-5), the boundedness of the kernel function and [s8, Th. 5.4.A], it can be shown that

$$\mathbf{a}_2(\boldsymbol{\theta}) \xrightarrow[N\to\infty]{p} \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{h}(\mathbf{x},\mathbf{x};\boldsymbol{\theta}); G\times G] < \infty \tag{S-97}$$

and

$$\mathbf{a}_3(\boldsymbol{\theta}) \xrightarrow[N\to\infty]{p} \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{h}(\mathbf{x},\mathbf{x}';\boldsymbol{\theta}); G\times G] < \infty. \tag{S-98}$$

Next, we shall analyze convergence in distribution of $\sqrt{N}\mathbf{a}_1(\boldsymbol{\theta}_h^*)$. We begin by calculating the expected value of $\mathbf{a}_1(\boldsymbol{\theta}_h^*)$. As shown at the beginning of Sec. II, $\boldsymbol{\theta}_h^*$ is the maximizer of the deterministic objective $\bar{\mathcal{J}}_h(\boldsymbol{\theta})$ defined in Eq. (S-24). Hence, by Assumption (B-2) and Eqs. (S-24), (S-60), (S-61) and (S-94) it follows that

$$\begin{aligned}\mathbb{E}[\mathbf{a}_1(\boldsymbol{\theta}_h^*); P_{\mathbf{a}_1(\boldsymbol{\theta}_h^*)}] &= \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{h}(\mathbf{x},\mathbf{x}'';\boldsymbol{\theta}_h^*)); G\times G\times G] \\ &= \nabla\bar{\mathcal{J}}_h(\boldsymbol{\theta}_h^*) \times \mathbb{E}^2\left[K_h(\mathbf{x}-\mathbf{x}'); G\times G\right] \times \mathbb{E}\left[d(\mathbf{x}'';\boldsymbol{\theta}_h^*); G\right] = \mathbf{0}.\end{aligned} \tag{S-99}$$

Now, define the statistic

$$T \triangleq \boldsymbol{\beta}^T\mathbf{a}_1(\boldsymbol{\theta}_h^*), \tag{S-100}$$

where $\boldsymbol{\beta}\in\mathbb{R}^m$ is an arbitrary deterministic coefficient vector. In the following, we shall express T as a normalized U-statistic [s8, Ch. 5]. Using (S-94) and (S-100), one can verify that

$$T = \frac{U}{6}, \tag{S-101}$$

where

$$U \triangleq \frac{6}{(N-2)(N-1)N}\sum_{c} w(\mathbf{x}_{i_1},\mathbf{x}_{i_2},\mathbf{x}_{i_3}), \tag{S-102}$$

$$w(\mathbf{r},\boldsymbol{s},\boldsymbol{t}) \triangleq \boldsymbol{\beta}^T\boldsymbol{\xi}(\mathbf{r},\boldsymbol{s},\boldsymbol{t}), \tag{S-103}$$

$$\boldsymbol{\xi}(\mathbf{r},\boldsymbol{s},\boldsymbol{t}) \triangleq \mathbf{z}(\mathbf{r},\boldsymbol{s},\boldsymbol{t}) + \mathbf{z}(\mathbf{r},\boldsymbol{t},\boldsymbol{s}) + \mathbf{z}(\boldsymbol{s},\mathbf{r},\boldsymbol{t}) + \mathbf{z}(\boldsymbol{s},\boldsymbol{t},\mathbf{r}) + \mathbf{z}(\boldsymbol{t},\mathbf{r},\boldsymbol{s}) + \mathbf{z}(\boldsymbol{t},\boldsymbol{s},\mathbf{r}), \tag{S-104}$$

$$\mathbf{z}(\mathbf{r},\boldsymbol{s},\boldsymbol{t}) \triangleq K_h(\mathbf{r}-\boldsymbol{s})\mathbf{h}(\mathbf{r},\boldsymbol{t};\boldsymbol{\theta}_h^*) \tag{S-105}$$

and $\sum_{c}$ denotes the summation over the $\binom{N}{3}$ combinations of distinct elements $\{i_1, i_2, i_3\}$ from $\{1,\ldots,N\}$. Note that U is a U-statistic with symmetric kernel $w(\cdot,\cdot,\cdot)$. Also note that by (S-99)-(S-101), U is a zero-mean statistic. Therefore, assume that

$$\mathbb{E}[w^2(\mathbf{x},\mathbf{x}',\mathbf{x}''); G\times G\times G] < \infty \tag{S-106}$$

and

$$\xi_1 > 0, \tag{S-107}$$

where

$$\xi_1 \triangleq \text{var}[w_1(\mathbf{x}); G], \tag{S-108}$$

$$w_1(\mathbf{r}) \triangleq \mathbb{E}[w(\mathbf{r}, \mathbf{x}', \mathbf{x}''); G \times G] = \boldsymbol{\beta}^T \boldsymbol{\zeta}(\mathbf{r}), \tag{S-109}$$

and

$$\boldsymbol{\zeta}(\mathbf{r}) \triangleq \mathbb{E}[\boldsymbol{\xi}(\mathbf{r}, \mathbf{x}', \mathbf{x}''); G \times G]. \tag{S-110}$$

By [s8, Th. 5.5.1-A], it follows that

$$\sqrt{N} U \xrightarrow[N \to \infty]{d} \mathcal{N}(0, 9\xi_1). \tag{S-111}$$

In Lemma 9, stated below, we show that

$$\xi_1 = 4\eta^2(\boldsymbol{\theta}_h^*, h) \boldsymbol{\beta}^T \mathbf{D}(\boldsymbol{\theta}_h^*, h) \boldsymbol{\beta}, \tag{S-112}$$

where $\mathbf{D}(\boldsymbol{\theta}, h)$ and $\eta(\boldsymbol{\theta}, h)$ are defined in Eqs. (18) and (S-67), respectively. Hence, by (S-99)-(S-101), (S-111) and the Cramér-Wold Device [s13, Th. 11.2.3] we conclude that

$$\sqrt{N} \mathbf{a}_1(\boldsymbol{\theta}_h^*) \xrightarrow[N \to \infty]{d} \mathcal{N}(\mathbf{0}, \eta^2(\boldsymbol{\theta}_h^*, h) \mathbf{D}(\boldsymbol{\theta}_h^*, h)). \tag{S-113}$$

Thus, the relation in (S-70) follows directly from (S-93), (S-97), (S-98), (S-113) and Slutskey's Theorem [s11, Th. 9.1.6]. To complete the proof, we need to show that the assumptions in (S-106) and (S-107) are satisfied. By Eq. (S-103) the inequality in (S-106) holds when $\mathbb{E}[\|\boldsymbol{\xi}(\mathbf{x}, \mathbf{x}', \mathbf{x}'')\|^2; G \times G \times G] < \infty$. Using Eqs. (S-104), (S-105), the Cauchy-Schwartz and the triangle inequalities, the boundedness of the kernel function $K_h(\cdot)$ and Assumption (B-5), one can verify that the latter inequality indeed holds. The inequality in (S-107) is satisfied when $\mathbf{D}(\boldsymbol{\theta}_h^*, h)$ is positive definite. Again, we note that $0 < \eta(\boldsymbol{\theta}_h^*, h) < \infty$ since the kernel function $K_h(\cdot)$ is strictly positive and bounded. □

Lemma 9. *The equality in (S-112) holds.*

Proof. Using (S-99)-(S-103), (S-109) and (S-110), one can verify that $\mathbb{E}[w_1(\mathbf{x}); G] = 0$. Hence, by (S-108)-(S-110), it follows that

$$\xi_1 = \boldsymbol{\beta}^T \mathbb{E}\left[\boldsymbol{\zeta}(\mathbf{x}) \boldsymbol{\zeta}^T(\mathbf{x}); G\right] \boldsymbol{\beta}. \tag{S-114}$$

Furthermore, using (S-60), (S-104), (S-105), (S-110) and the symmetry property of the kernel function $K_h(\cdot)$, one can verify that

$$\begin{aligned} \boldsymbol{\zeta}(\mathbf{r}) &= 2\Big(\mathbb{E}\left[K_h(\mathbf{x}' - \mathbf{r}); G\right] \times \mathbb{E}\left[\mathbf{h}(\mathbf{r}, \mathbf{x}'; \boldsymbol{\theta}_h^*); G\right] \\ &+ \mathbb{E}\left[K_h(\mathbf{x}' - \mathbf{r}) \mathbf{h}(\mathbf{x}', \mathbf{x}''; \boldsymbol{\theta}_h^*); G \times G\right] \\ &+ \mathbb{E}\left[K_h(\mathbf{x}' - \mathbf{x}'') \mathbf{h}(\mathbf{x}', \mathbf{r}; \boldsymbol{\theta}_h^*); G \times G\right]\Big) = 2\eta(\boldsymbol{\theta}_h^*, h) \mathbf{v}(\mathbf{r}, \boldsymbol{\theta}_h^*, h), \end{aligned} \tag{S-115}$$

where $\mathbf{v}(\mathbf{r}, \boldsymbol{\theta}, h)$ is defined below Eq. (18). Hence, the relation in (S-112) follows directly from Eqs. (18), (S-114) and (S-115). □

IV. Fisher consistency

Proposition 4. *Under Assumption (A-2), $\hat{\boldsymbol{\theta}}_h$ (8) is a Fisher consistent estimator [s14] of $\boldsymbol{\theta}_h^*$ (9), i.e., it can be represented as a statistical functional of the empirical probability distribution $\mathbf{S}[\hat{G}]$ that satisfies $\mathbf{S}[G] = \boldsymbol{\theta}_h^*$.*

Proof. First, we show that $\hat{\boldsymbol{\theta}}_h$ can be represented as a statistical functional of the empirical probability distribution $\hat{G} \triangleq \frac{1}{N}\sum_{n=1}^{N}\delta_{\mathbf{x}_n}$, where $\delta_{\mathbf{r}}$ denotes a Dirac measure [s3] concentrated at $\mathbf{r}$. Using Eq. (7), one can verify that the objective function

$$\mathcal{J}_h(\boldsymbol{\theta}) = \frac{\mathbb{E}[\tilde{K}_h(\mathbf{x}-\mathbf{x}')\log f(\mathbf{x};\boldsymbol{\theta}); \hat{G}\times\hat{G}]}{\mathbb{E}[\tilde{K}_h(\mathbf{x}-\mathbf{x}'); \hat{G}\times\hat{G}]} - \log\mathbb{E}[\tilde{K}_h(\mathbf{x}-\mathbf{x}'); \hat{G}\times F_{\boldsymbol{\theta}}] \triangleq H[\hat{G};\boldsymbol{\theta}]. \tag{S-116}$$

where $\tilde{K}_h(\mathbf{r}) \triangleq K_h(\mathbf{r}) - K_h(0)\mathbb{1}_0(\mathbf{r})$ and $\mathbb{1}_0(\cdot)$ denotes the indicator function of 0. Therefore, by (8) it follows that

$$\hat{\boldsymbol{\theta}}_h = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} H[\hat{G};\boldsymbol{\theta}] \triangleq \mathbf{S}[\hat{G}]. \tag{S-117}$$

Next, we prove Fisher consistency of $\hat{\boldsymbol{\theta}}_h$. By (S-24) and (S-116) it follows that

$$H[G;\boldsymbol{\theta}] = \bar{\mathcal{J}}_h(\boldsymbol{\theta}), \tag{S-118}$$

which as shown below (S-24) is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_h^*$ when Assumption (A-2) is satisfied. Therefore, by (S-117) we conclude that

$$\boldsymbol{\theta}_h^* = \mathbf{S}[G], \tag{S-119}$$

implying that $\hat{\boldsymbol{\theta}}_h$ is a Fisher consistent estimator of $\boldsymbol{\theta}_h^*$. □

V. Influence function

Define the contaminated probability distribution

$$G_\epsilon \triangleq (1-\epsilon)F_{\boldsymbol{\theta}_0} + \epsilon\delta_{\mathbf{r}}, \tag{S-120}$$

where $0 \le \epsilon \le 1$, $\mathbf{r} \in \mathbb{R}^p$ and $\delta_{\mathbf{r}}$ is the Dirac probability measure at $\mathbf{r}$. The influence function of a Fisher consistent estimator with statistical functional $\mathbf{S}[\cdot]$ at $F_{\boldsymbol{\theta}_0}$ is defined as:

$$\mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0) \triangleq \lim_{\epsilon\to 0}\frac{\mathbf{S}[G_\epsilon]-\mathbf{S}[F_{\boldsymbol{\theta}_0}]}{\epsilon} = \left.\frac{d\mathbf{S}[G_\epsilon]}{d\epsilon}\right|_{\epsilon=0}. \tag{S-121}$$

Proposition 5. *Assume that $\hat{\boldsymbol{\theta}}_h$ is Fisher consistent. Furthermore, assume that conditions (B-3) and (B-4), stated in Theorem 3, are satisfied. Then, the influence function of $\hat{\boldsymbol{\theta}}_h$ takes the form:*

$$\mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0,h) = \bar{\mathbf{D}}^{-1}(\boldsymbol{\theta}_0,h)\bar{\mathbf{c}}(\mathbf{r},\boldsymbol{\theta}_0,h)\bar{\psi}_F(\mathbf{r},\boldsymbol{\theta}_0,h), \tag{S-122}$$

where $\bar{\mathbf{D}}(\boldsymbol{\theta},h) \triangleq \mathbb{E}[\bar{\psi}_F(\mathbf{x},\boldsymbol{\theta},h)\bar{\mathbf{c}}(\mathbf{x},\boldsymbol{\theta},h)\bar{\mathbf{c}}^T(\mathbf{x},\boldsymbol{\theta},h); F_{\boldsymbol{\theta}}]$, $\bar{\mathbf{c}}(\mathbf{r},\boldsymbol{\theta},h) \triangleq \mathbf{q}(\mathbf{r},\boldsymbol{\theta}) - \mathbb{E}[\bar{\psi}_F(\mathbf{x},\boldsymbol{\theta},h)\mathbf{q}(\mathbf{x},\boldsymbol{\theta}); F_{\boldsymbol{\theta}}]$, $\mathbf{q}(\mathbf{r},\boldsymbol{\theta})$ is the score-function defined in (17), $\bar{\psi}_F(\mathbf{r},\boldsymbol{\theta},h) \triangleq (K_h * f)(\mathbf{r};\boldsymbol{\theta})/\mathbb{E}[(K_h * f)(\mathbf{x};\boldsymbol{\theta}); F_{\boldsymbol{\theta}}]$ and it is assumed that $\bar{\mathbf{D}}(\boldsymbol{\theta},h)$ is non-singular.

Proof. Throughout the proof we shall assume that integration and differentiation operations can be interchanged. Using [s12, Th. 2.40] it can be shown that this assumption is justified under conditions (B-3), (B-4) and the boundedness of the kernel function $K_h(\cdot)$.

By Eq. (S-92), the definition of $\psi_G(\cdot,\cdot)$ in Eq. (2), the definition of $\psi_F(\cdot,\cdot,\cdot)$ stated below Eq. (16) and the Fisher consistency of $\hat{\boldsymbol{\theta}}_h$ it follows that

$$\frac{\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{q}(\mathbf{x};\mathbf{S}[G_\epsilon]);G_\epsilon\times G_\epsilon]}{\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}');G_\epsilon\times G_\epsilon]}-\frac{\int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});G_\epsilon]\tilde{\mathbf{q}}(\mathbf{x};\mathbf{S}[G_\epsilon]))d\lambda(\boldsymbol{\tau})}{\int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});G_\epsilon]f(\mathbf{x};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau})}=\mathbf{0}, \tag{S-123}$$

where $\tilde{\mathbf{q}}(\mathbf{x};\boldsymbol{\theta})\triangleq\nabla_{\boldsymbol{\theta}}f(\mathbf{x};\boldsymbol{\theta})$. Under the assumption of a symmetric kernel function, it follows from (S-120) that the terms comprising (S-123) take the forms:

$$\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{q}(\mathbf{x};\mathbf{S}[G_\epsilon]);G_\epsilon\times G_\epsilon]=(1-\epsilon)^2\mathbf{a}(\epsilon)+\epsilon(1-\epsilon)\left(\mathbf{b}(\mathbf{r},\epsilon)+\mathbf{q}(\mathbf{r};\mathbf{S}[G_\epsilon])c(\mathbf{r})\right)+\epsilon^2\mathbf{q}(\mathbf{r};\mathbf{S}[G_\epsilon])K_h(\mathbf{0}), \tag{S-124}$$

$$\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}');G_\epsilon\times G_\epsilon]=(1-\epsilon)^2d+2\epsilon(1-\epsilon)c(\mathbf{r})+\epsilon^2K_h(\mathbf{0}), \tag{S-125}$$

$$\int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});G_\epsilon]\tilde{\mathbf{q}}(\mathbf{x};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau})=(1-\epsilon)\mathbf{w}(\epsilon)+\epsilon\mathbf{z}(\mathbf{r},\epsilon) \tag{S-126}$$

and

$$\int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});G_\epsilon]f(\boldsymbol{\tau};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau})=(1-\epsilon)u(\epsilon)+\epsilon v(\mathbf{r},\epsilon), \tag{S-127}$$

where

$$\mathbf{a}(\epsilon)\triangleq\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{q}(\mathbf{x};\mathbf{S}[G_\epsilon]);F_{\boldsymbol{\theta}_0}\times F_{\boldsymbol{\theta}_0}],\quad \mathbf{b}(\mathbf{r},\epsilon)\triangleq\mathbb{E}[K_h(\mathbf{x}-\mathbf{r})\mathbf{q}(\mathbf{x};\mathbf{S}[G_\epsilon]);F_{\boldsymbol{\theta}_0}], \tag{S-128}$$

$$c(\mathbf{r})\triangleq\mathbb{E}[K_h(\mathbf{x}-\mathbf{r});F_{\boldsymbol{\theta}_0}],\quad d\triangleq\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}');F_{\boldsymbol{\theta}_0}\times F_{\boldsymbol{\theta}_0}], \tag{S-129}$$

$$\mathbf{w}(\epsilon)\triangleq\int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});F_{\boldsymbol{\theta}_0}]\tilde{\mathbf{q}}(\mathbf{x};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau})],\quad \mathbf{z}(\mathbf{r},\epsilon)\triangleq\int_{\mathbb{R}^p}K_h(\mathbf{r}-\boldsymbol{\tau})\tilde{\mathbf{q}}(\mathbf{x};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau})], \tag{S-130}$$

$$u(\epsilon)\triangleq\int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});F_{\boldsymbol{\theta}_0}]f(\boldsymbol{\tau};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau})\quad\text{and}\quad v(\mathbf{r},\epsilon)\triangleq\int_{\mathbb{R}^p}K_h(\mathbf{r}-\boldsymbol{\tau})f(\boldsymbol{\tau};\mathbf{S}[G_\epsilon])d\lambda(\boldsymbol{\tau}). \tag{S-131}$$

Hence, by (S-123)-(S-127) we obtain that

$$\boldsymbol{\alpha}(\mathbf{r},\epsilon)-\boldsymbol{\beta}(\mathbf{r},\epsilon)=\mathbf{0}, \tag{S-132}$$

where

$$\boldsymbol{\alpha}(\mathbf{r},\epsilon)\triangleq\frac{(1-\epsilon^2)\mathbf{a}(\epsilon)+\epsilon(1-\epsilon)(\mathbf{b}(\mathbf{r},\epsilon)+c(\mathbf{r})\mathbf{q}(\mathbf{r};\mathbf{S}[G_\epsilon]))+\epsilon^2\mathbf{q}(\mathbf{r};\mathbf{S}[G_\epsilon])K_h(\mathbf{0})}{(1-\epsilon)^2d+2\epsilon(1-\epsilon)c(\mathbf{r})+\epsilon^2K_h(0)} \tag{S-133}$$

and

$$\boldsymbol{\beta}(\mathbf{r},\epsilon)\triangleq\frac{(1-\epsilon)\mathbf{w}(\epsilon)+\epsilon\mathbf{z}(\mathbf{r},\epsilon)}{(1-\epsilon)u(\epsilon)+\epsilon v(\mathbf{r},\epsilon)}. \tag{S-134}$$

From (S-132) we obtain that

$$\left.\frac{\partial\boldsymbol{\alpha}(\mathbf{r},\epsilon)}{\partial\epsilon}\right|_{\epsilon=0}-\left.\frac{\boldsymbol{\beta}(\mathbf{r},\epsilon)}{\partial\epsilon}\right|_{\epsilon=0}=\mathbf{0}. \tag{S-135}$$

Using (S-133) and (S-134) one can verify that

$$\left.\frac{\partial\boldsymbol{\alpha}(\mathbf{r},\epsilon)}{\partial\epsilon}\right|_{\epsilon=0}=\frac{\dot{\mathbf{a}}(0)+\mathbf{b}(\mathbf{r},0)+\mathbf{q}(\mathbf{r};\boldsymbol{\theta}_0)c(\mathbf{r})}{d}-\frac{2\mathbf{a}(0)c(\mathbf{r})}{d^2} \tag{S-136}$$

and

$$\left.\frac{\partial\boldsymbol{\beta}(\mathbf{r},\epsilon)}{\partial\epsilon}\right|_{\epsilon=0}=\frac{\dot{\mathbf{w}}(0)+\mathbf{z}(\mathbf{r},0)}{u(0)}-\frac{\mathbf{w}(0)(\dot{u}(0)+v(0))}{u^2(0)}, \tag{S-137}$$

where $\dot{\mathbf{a}}(0) \triangleq \left.\frac{d\mathbf{a}(\epsilon)}{d\epsilon}\right|_{\epsilon=0}$, $\dot{\mathbf{w}}(0) \triangleq \left.\frac{d\mathbf{w}(\epsilon)}{d\epsilon}\right|_{\epsilon=0}$ and $\dot{u}(0) \triangleq \left.\frac{du(\epsilon)}{d\epsilon}\right|_{\epsilon=0}$. By the chain-rule for derivatives we obtain that

$$\dot{\mathbf{a}}(0) = \left.\frac{d\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{q}(\mathbf{x};\mathbf{S}[G_\epsilon]);F_{\boldsymbol{\theta}_0}\times F_{\boldsymbol{\theta}_0}]}{d\mathbf{S}[G_\epsilon]}\right|_{\epsilon=0} \left.\frac{d\mathbf{S}[G_\epsilon]}{d\epsilon}\right|_{\epsilon=0}. \tag{S-138}$$

Note that

$$\frac{d\mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{q}(\mathbf{x};\mathbf{S}[G_\epsilon]);F_{\boldsymbol{\theta}_0}\times F_{\boldsymbol{\theta}_0}]}{d\mathbf{S}[G_\epsilon]} = \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{H}(\mathbf{x};\mathbf{S}[G_\epsilon]);F_{\boldsymbol{\theta}_0}\times F_{\boldsymbol{\theta}_0}], \tag{S-139}$$

where $\mathbf{H}(\cdot,\cdot)$ is the Hessian defined in Eq. (17). Also note that the Fisher consistency of $\hat{\boldsymbol{\theta}}_h$ implies that

$$\mathbf{S}[G_\epsilon] = \boldsymbol{\theta}_0 \;\text{ for }\; \epsilon = 0 \tag{S-140}$$

and therefore,

$$\mathbf{H}(\mathbf{x};\mathbf{S}[G_\epsilon])|_{\epsilon=0} = \mathbf{H}(\mathbf{x};\boldsymbol{\theta}_0). \tag{S-141}$$

Additionally, by (S-121) it follows that $\left.\frac{d\mathbf{S}[G_\epsilon]}{d\epsilon}\right|_{\epsilon=0} = \mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0,h)$. Hence, we conclude that

$$\dot{\mathbf{a}}(0) = \mathbf{A}\times\mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0,h). \tag{S-142}$$

where $\mathbf{A} \triangleq \mathbb{E}[K_h(\mathbf{x}-\mathbf{x}')\mathbf{H}(\mathbf{x};\boldsymbol{\theta}_0);F_{\boldsymbol{\theta}_0}\times F_{\boldsymbol{\theta}_0}]$. Similarly, it can be shown that

$$\dot{\mathbf{w}}(0) = \mathbf{B}\times\mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0,h) \tag{S-143}$$

and

$$\dot{u}(0) = \boldsymbol{\eta}^T\times\mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0,h), \tag{S-144}$$

where the matrix $\mathbf{B} \triangleq \int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});F_{\boldsymbol{\theta}_0}]\tilde{\mathbf{H}}(\mathbf{r};\boldsymbol{\theta}_0)d\lambda(\boldsymbol{\tau})$, with $\tilde{\mathbf{H}}(\mathbf{x};\boldsymbol{\theta}) \triangleq \nabla^2_{\boldsymbol{\theta}}f(\mathbf{x};\boldsymbol{\theta})$ and the vector $\boldsymbol{\eta} \triangleq \int_{\mathbb{R}^p}\mathbb{E}[K_h(\mathbf{x}-\boldsymbol{\tau});F_{\boldsymbol{\theta}_0}]\tilde{\mathbf{q}}(\mathbf{r};\boldsymbol{\theta}_0)d\lambda(\boldsymbol{\tau})$. Therefore, by (S-135)-(S-137) and (S-142)-(S-144) we obtain that

$$\mathbf{IF}(\mathbf{r};\boldsymbol{\theta}_0,h) = \mathbf{J}^{-1}(\boldsymbol{\theta}_0)\mathbf{K}(\mathbf{r},\boldsymbol{\theta}_0), \tag{S-145}$$

where

$$\mathbf{J}(\boldsymbol{\theta}) \triangleq \frac{\mathbf{A}}{d} - \frac{\mathbf{B}}{u(0)} + \frac{\mathbf{w}(0)\boldsymbol{\eta}^T}{u^2(0)} = \frac{\mathbf{A}-\mathbf{B}}{d} + \frac{\mathbf{w}(0)\boldsymbol{\eta}^T}{d^2} \tag{S-146}$$

and

$$\begin{aligned}\mathbf{K}(\mathbf{r},\boldsymbol{\theta}) \;&\triangleq\; -\frac{\mathbf{b}(\mathbf{r},0)+c(r)\mathbf{q}(\mathbf{r};\boldsymbol{\theta})}{d} + \frac{2c(\mathbf{r})\mathbf{a}(0)}{d^2} + \frac{\mathbf{z}(\mathbf{r},0)}{u(0)} - \frac{v(\mathbf{r},0)\mathbf{w}(0)}{u^2(0)} \\ &=\; -\frac{\mathbf{b}(\mathbf{r},0)+c(r)\mathbf{q}(\mathbf{r};\boldsymbol{\theta})-\mathbf{z}(\mathbf{r},0)}{d} + \frac{2c(\mathbf{r})\mathbf{a}(0)-v(\mathbf{r},0)\mathbf{w}(0)}{d^2},\end{aligned} \tag{S-147}$$

where the last equalities in (S-146) and (S-147) follow from the definitions of d and $u(\cdot)$ in (S-129) and (S-131), respectively, and relation (S-140) according to which $d = u(0)$. Using the definitions of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{w}(\cdot)$, $\boldsymbol{\eta}$, d and $u(\cdot)$ stated above, relation (S-140) and the identity

$$\tilde{\mathbf{H}}(\mathbf{x};\boldsymbol{\theta}) = f(\mathbf{x};\boldsymbol{\theta})\left(\mathbf{H}(\mathbf{x};\boldsymbol{\theta}) + \mathbf{q}(\mathbf{x};\boldsymbol{\theta})\mathbf{q}^T(\mathbf{x};\boldsymbol{\theta})\right),$$

that follows directly from the definition of $\tilde{\mathbf{H}}(\cdot;\cdot)$ below (S-144), one can verify that

$$\mathbf{J}(\boldsymbol{\theta}) = -\bar{\mathbf{D}}(\boldsymbol{\theta},h), \tag{S-148}$$

where $\bar{\mathbf{D}}(\boldsymbol{\theta}, h)$ is defined below Eq. (S-122). Now, by (S-123), (S-140) and the definitions of $\mathbf{a}(\cdot)$, $\mathbf{w}(\cdot)$, d and $u(\cdot)$ it follows that

$$\mathbf{a}(0) = \mathbf{w}(0). \tag{S-149}$$

Therefore, by (S-147), (S-149), the definitions of $\mathbf{b}(\cdot,\cdot)$, $c(\cdot)$, $\mathbf{z}(\cdot,\cdot)$, $\mathbf{a}(\cdot)$, $v(\cdot,\cdot)$ and $\mathbf{w}(\cdot)$ stated above, it can be shown that

$$\mathbf{K}(\mathbf{r}, \boldsymbol{\theta}) = -\bar{\mathbf{c}}(\mathbf{r}, \boldsymbol{\theta}_0, h)\bar{\psi}_F(\mathbf{r}, \boldsymbol{\theta}_0, h), \tag{S-150}$$

where $\bar{\mathbf{c}}(\mathbf{r}, \boldsymbol{\theta}_0, h)$ and $\bar{\psi}_F(\mathbf{r}, \boldsymbol{\theta}_0, h)$ are defined below Eq. (S-122). Hence the equality in (S-122) follows directly from (S-145), (S-148) and (S-150). □

References

[s1] T. M. Cover, *Elements of information theory,* John Wiley & Sons, 1999.

[s2] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.

[s3] G. B. Folland, *Real Analysis,* John Wiley and Sons, 1984.

[s4] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* third edition, John Wiley & Sons, 2003.

[s5] H. White, *Estimation, inference and specification analysis,* Cambridge university press, 1996.

[s6] C. H. Edwards, *Advanced calculus of several variables,* Courier Corporation, 2012.

[s7] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics,* vol. 4, pp. 2111-2245, 1994.

[s8] R. J. Serfling, *Approximation theorems of mathematical statistics.* John Wiley & Sons, 1980.

[s9] W. K. Newey, "Uniform convergence in probability and stochastic equicontinuity," *Econometrica,* pp. 1161-1167, 1991.

[s10] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.,* vol. 14, pp. 217-226, 1943.

[s11] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.

[s12] M. Giaquinta and G. Modica, *Mathematical Analysis: An Introduction to Functions of Several Variables.* Birkhäuser, p. 88, Boston, 2009.

[s13] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses.* Springer Texts in Statistics, 2005.

[s14] D. R. Cox and D. V. Hinkley, *Theoretical Statistics,* Chapman & Hall, 1974.