Supplementary Material to
“Simulation–based Predition”

Eunji Lim and Peter W. Glynn

EC.1. Proofs of Theorems 1, 3, and 4

Throughout this section and the next two, we will use the following lemmas. Lemma 3 is due to Ibragimov (1962).

Lemma 1. Let $V = (V_i : i = 0, 1, \cdots)$ be a sequence of random vectors, each of which takes values in $S \subset \mathbb{R}^p$. Let $h : S \to \mathbb{R}$ be a measurable function. If V is strongly mixing with the strong mixing coefficients $(\alpha_n : n = 1, 2, \cdots)$, then $(h(V_i) : i = 0, 1, \cdots)$ is also strongly mixing with the strong mixing coefficients $(\alpha'_n : n = 1, 2, \cdots)$ satisfying $\alpha'_n \leq \alpha_n$ for $n = 1, 2, \cdots$.

Proof of Lemma 1. For $0 \leq i \leq j \leq \infty$, let $\mathcal{G}_i^j = \sigma(h(V_i), h(V_{i+1}), \cdots, h(V_j))$. Then $\mathcal{G}_i^j \subset \sigma(V_i, V_{i+1}, \cdots, V_j)$. Hence, $\alpha'_n \leq \alpha_n$ for $n = 1, 2, \cdots$. □

Lemma 2. Let $X = (X_i : i = 0, 1, \cdots)$ be a strictly stationary sequence of random vectors, each of which takes values in $S \subset \mathbb{R}^p$. Let $h : S \to \mathbb{R}$ be a measurable function. Then $(h(X_i) : i = 0, 1, \cdots)$ is also strictly stationary.

Proof of Lemma 2. Suppose $X = (X_i : i = 0, 1, \cdots)$ is strictly staionary. By the stationarity of X, we have

$$\begin{aligned}
&\mathbb{P}[(h(X_{i_1}), \cdots, h(X_{i_k})) \in B_1 \times \cdots \times B_k] \\
&= \mathbb{P}[(X_{i_1}, \cdots, X_{i_k}) \in (h^{-1}(B_1), \cdots, h^{-1}(B_k))] \\
&= \mathbb{P}[(X_{i_1+j}, \cdots, X_{i_k+j}) \in (h^{-1}(B_1), \cdots, h^{-1}(B_k))] \\
&= \mathbb{P}[(h(X_{i_1+j}), \cdots, h(X_{i_k+j})) \in B_1 \times \cdots B_k]
\end{aligned}$$

for any nonnegative integer j, positive integer k, nonnegative integers $i_1, \cdots, i_k$ satisfying $i_1 < i_2 < \cdots < i_k$, and Borel measurable $B_1, \cdots, B_k \subset \mathbb{R}$. Hence, $(h(X_i) : i = 0, 1, \cdots)$ is strictly stationary. □

Lemma 3 (Ibragimov, 1962)**.** Let $W = (W_i : i = 0, 1, \cdots)$ be a strictly stationary sequence of real–valued random variables satisfying $\mathbb{E}[W_0] = 0$. Suppose W is strongly mixing with the strong mixing coefficients $(\alpha_n : n = 1, 2, \cdots)$. If there exists a positive real number δ such that

$$\mathbb{E}[|W_0|^{2+\delta}] < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha_n^{\delta/(2+\delta)} < \infty,$$

then

$$\sigma^2 \triangleq \mathbb{E}[W_0^2] + 2\sum_{i=1}^{\infty} \mathbb{E}[W_0 W_i] < \infty,$$

and if $\sigma^2 > 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} W_i \Rightarrow \sigma N(0, 1)$$

as $n \to \infty$.

Lemma 4. Let $W = (W_i : i = 0, 1, \cdots)$ be a wide–sense stationary (i.e., covariance stationary) sequence of real–valued random variables satisfying $\mathbb{E}[W_0] = 0$. Suppose W is strongly mixing with the strong mixing coefficients $(\alpha_n : n = 1, 2, \cdots)$. If there exists a positive real number δ such that

$$\mathbb{E}[|W_0|^{2+\delta}] \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha_n^{\delta/(2+\delta)} < \infty,$$

then

$$\frac{1}{n} \sum_{i=0}^{n-1} W_i \to \mathbb{E}[W_0]$$

a.s. as $n \to \infty$.

Proof of Lemma 4. By Theorem 1.0 on p. 588 of Rio (1993),

$$|\mathbb{E}[W_0^2]| + 2\sum_{i=1}^{\infty} |\mathbb{E}[W_0 W_i]| < \infty$$

Since W is wide–sense stationary with absolutely summable autocovariances, by the strong law of large numbers for covariance stationary processes, we have

$$\frac{1}{n} \sum_{i=0}^{n-1} W_i \to \mathbb{E}[W_0]$$

a.s. as $n \to \infty$. $\square$

We are now ready to present the proofs of Theorems 1, 3, and 4.

Proof of Theorem 1. We begin by noticing that

$$r^{1/2}\left(\tilde{k}_r(\mathbf{y}) - k(\mathbf{y})\right) = \frac{r^{-1/2}\sum_{i=0}^{r-1}(\tilde{Z}_i - k(\mathbf{y}))I(\tilde{Y}_i = \mathbf{y})}{r^{-1}\sum_{i=0}^{r-1} I(\tilde{Y}_i = \mathbf{y})}$$

and that P1 (a) and P1 (b) imply Theorem 1.

It remains to show that A1 implies P1. We will use Lemma 3 to establish P1 from A1. First, consider the process $((\tilde{Z}_i - k(\mathbf{y}))I(\tilde{Y}_i = \mathbf{y}) : i = 0, 1, \cdots)$. Since $(\tilde{X}_i : i = 0, 1, \cdots)$ is strictly stationary and strongly mixing, so is $((\tilde{Z}_i - k(\mathbf{y}))I(\tilde{Y}_i = \mathbf{y}) : i = 0, 1, \cdots)$; see, for example, Lemmas 1 and 2. Note that A1 (a) implies

$$\mathbb{E}[|(\tilde{Z}_0 - k(\mathbf{y}))I(\tilde{Y}_0 = \mathbf{y})|^{2+\delta}] \leq \mathbb{E}[|\tilde{Z}_0 - k(\mathbf{y})|^{2+\delta}] < \infty.$$

By A1 (b) and Lemma 1, $((\tilde{Z}_i - k(\mathbf{y}))I(\tilde{Y}_i = \mathbf{y}) : i = 0, 1, \cdots)$ has strong mixing coefficients $(\beta_n : n = 1, 2, \cdots)$ satisfying $\sum_{n=1}^{\infty} \beta_n^{\delta/(2+\delta)} < \infty$. By Lemma 3, $\sigma^2 < \infty$. If $\sigma^2 > 0$, then P1 (a) holds with $\sigma_1 = \sigma$.

Next, consider the process $(I(\tilde{Y}_i = \mathbf{y}) : i = 0, 1, \cdots)$. $(I(\tilde{Y}_i = \mathbf{y}) : i = 0, 1, \cdots)$ has strong mixing coefficients $(\nu_n : n = 1, 2, \cdots)$ satisfying $\sum_{n=1}^{\infty} \nu_n^{\delta/(2+\delta)} < \infty$. By Lemma 4, P1 (b) follows. $\square$

Proof of Theorem 3. We will first establish Theorem 3 in the case where X is initialized from its stationary distribution. Once this is established, since X is a Harris ergodic Markov chain, Theorem 3 follows even when X is initialized from an arbitrary initial distribution; see Proposition 17.1.6 of Meyn and Tweedie (1993).

Suppose X is initialized from its stationary distribution. For $i = 0, 1, \cdots$, let

$$A_i \triangleq \left(\frac{1}{m}(Z_i - \mu(X_i)) + (\mu(X_i) - k(\mathbf{y}))\right) I(Y_i = \mathbf{y})$$

and

$$B_i = \frac{1}{m-1}\sum_{j=1}^{m-1}(Z_{ij} - \mu(X_i))I(Y_i = \mathbf{y}).$$

The proof of Theorem 3 consists of two steps.

Step 1: We will use Lemma 3 to establish P2 (a) from A3 (a) and (b). Let $(\gamma_n : n = 1, 2, \cdots)$ be the strong mixing coefficients of the process

$(A_i : i = 0, 1, \cdots)$. Since $(X_i : i = 0, 1, \cdots)$ is strictly stationary and strongly mixing, so is $(A_i : i = 0, 1, \cdots)$. By A3 (b), $(\gamma_n : n = 1, 2, \cdots)$ satisfies $\sum_{n=1}^{\infty} \gamma_n^{\delta/(2+\delta)} < \infty$. Also, by A3 (a), $\mathbb{E}[|A_i|^{2+\delta}] < \infty$. So, Lemma 3 guarantees that A3 (a) and (b) imply that $\sigma^2(m)$ is convergent and if $\sigma^2(m) > 0$, then P2 (a) holds with $\sigma_3 = \sigma(m)$.

Next, to show that A3 (b) implies P2 (c), we will use Lemma 4. Let $(\nu_n : n = 1, 2, \cdots)$ be the strong mixing coefficients of $(I(Y_i = \mathbf{y}) : i = 0, 1, \cdots)$. Since $(X_i : i = 0, 1, \cdots)$ is strictly stationary sequence with the strong mixing coefficients satisfying A3 (b), so is $(I(Y_i = \mathbf{y}) : i = 0, 1, \cdots)$ with $\sum_{n=1}^{\infty} \nu_n^{\delta/(2+\delta)} < \infty$. By Lemma 4, P2 (c) follows.

We next prove that A3 implies P2 (b) using the Lindeberg–Feller Theorem with the Lyapunov condition. Note that, conditional on $X_0, X_1, \cdots$, the B_i's are independent. So, it suffices to show that

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[B_i^2 | X_0, X_1, \cdots] \to \eta^2(\mathbf{y}) p(\mathbf{y})/(m-1) \tag{1}$$

and

$$\sup_{i=0,1,\cdots} \mathbb{E}[|B_i|^{2+\delta} | X_0, X_1, \cdots] < \infty. \tag{2}$$

One can see that (2) follows from A3 (c). To establish (1), note that

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[B_i^2 | X] = \frac{1}{n} \frac{1}{m-1} \sum_{i=0}^{n-1} \mathrm{var}[Z_i | X_i] I(Y_i = \mathbf{y}).$$

Let $f_2(X_i) \triangleq \frac{1}{m-1} \mathrm{var}[Z_i | X_i] I(Y_i = \mathbf{y})$, then f_2 is a measurable function, so $(f_2(X_i) : i = 0, 1, \cdots)$ is a strictly stationary sequence with the strong mixing coefficients $(\rho_n : n = 1, 2, \cdots)$ satisfying $\sum_{n=1}^{\infty} \rho_n^{\delta/(2+\delta)} < \infty$ (by A3 (b)) and $\mathbb{E}[|f(X_0)|^{2+\delta}] < \infty$ (by A3 (c)). By Lemma 4, it follows that

$$\frac{1}{n} \sum_{i=0}^{n-1} f_2(X_i) \to \mathbb{E}[f_2(X_0)] = \eta^2(\mathbf{y}) p(\mathbf{y})/(m-1)$$

a.s. as $n \to \infty$. So, (1) is established. Since (1) and (2) are established, the Lindeberg–Feller theorem with the Lyapunov condition implies P2 (b).

Step 2: Now that P2 is established, it remains to establish Theorem 3 from P2. Note that

$$
\begin{aligned}
&n^{1/2}(k_n(\mathbf{y})-k(\mathbf{y}))\\
&\qquad=\frac{1}{n^{-1}\sum_{i=0}^{n-1}I(Y_i=\mathbf{y})}\left\{n^{-1/2}\sum_{i=0}^{n-1}A_i+\frac{m-1}{m}n^{-1/2}\sum_{i=0}^{n-1}B_i\right\}. \qquad (3)
\end{aligned}
$$

Since

$$
\mathbb{E}\left[\sum_{i=0}^{n-1}A_i\sum_{i=0}^{n-1}B_i\right]=\mathbb{E}\left[\sum_{i=0}^{n-1}A_i\mathbb{E}\left[\sum_{i=0}^{n-1}B_i|X_0,X_1,\cdots\right]\right]=0,
$$

$\sum_{i=0}^{n-1}A_i$ and $\sum_{i=0}^{n-1}B_i$ are independent, and hence applying the continuous mapping theorem and Slutsky's theorem to (3) together with P2 yields Theorem 3. □

Proof of Theorem 4. Let $\overline{Z}_i=\frac{1}{m}Z_i+\frac{1}{m}\sum_{j=1}^{m-1}Z_{ij}$ for $i=0,1,\cdots$. Since $\mathbb{E}[\overline{Z}_i|Y_i=\mathbf{y}]=k(\mathbf{y})$, Corollary 2 of Masry (2005) guarantees that

$$
(na_n^d)^{1/2}(k_n'(\mathbf{y})-k(\mathbf{y}))\Rightarrow\nu'(m)N(0,1)
$$

for some positive constant $\nu'(m)$. It remains to show

$$
\begin{aligned}
\nu'^2(m)=&\left(\mathbb{E}[(\mu(X_i)-k(\mathbf{y}))^2|Y_i=\mathbf{y}]+\frac{1}{m}\mathbb{E}[(Z_i-\mu(X_i))^2|Y_i=\mathbf{y}]\right)\\
&\times\frac{\int_0^1\{f_\lambda(x)\}^2\,dx}{g(\mathbf{y})\left(\int_0^1 f_\lambda(x)dx\right)^2}. \qquad (4)
\end{aligned}
$$

(2.11) of Masry (2005) implies

$$
\nu'^2(m)=\mathrm{var}[\overline{Z}_i|Y_i=\mathbf{y}]\frac{\int_0^1\{f_\lambda(x)\}^2\,dx}{g(\mathbf{y})\left(\int_0^1 f_\lambda(x)dx\right)^2}. \qquad (5)
$$

Note that

$$
\begin{aligned}
&\mathrm{var}[\overline{Z}_i|Y_i=\mathbf{y}]\\
&\quad=\ \mathbb{E}\left[\left(\frac{1}{m}Z_i+\frac{1}{m}\sum_{j=1}^{m-1}Z_{ij}-k(\mathbf{y})\right)^2\middle|Y_i=\mathbf{y}\right]
\end{aligned}
$$

$$= \mathbb{E}\left[\left(\frac{1}{m}(Z_i - \mu(X_i)) + (\mu(X_i) - k(\mathbf{y})) + \frac{1}{m}\sum_{j=1}^{m-1}(Z_{ij} - \mu(X_i))\right)^2 \middle| Y_i = \mathbf{y}\right].$$

Set $C_i = \frac{1}{m}(Z_i - \mu(X_i)) + (\mu(X_i) - k(\mathbf{y}))$ and $D_i = \frac{1}{m}\sum_{j=1}^{m-1}(Z_{ij} - \mu(X_i))$ for $i = 0, 1, \cdots$. Then, since

$$\begin{aligned}
&\mathbb{E}[C_i D_i | Y_i = \mathbf{y}] = \mathbb{E}[\mathbb{E}[C_i D_i | X_i] | Y_i = \mathbf{y}] = 0, \\
&\mathbb{E}[D_i^2 | Y_i = \mathbf{y}] = \mathbb{E}[\mathbb{E}[D_i^2 | X_i] | Y_i = \mathbf{y}] = \frac{m-1}{m^2}\mathbb{E}[\mathbb{E}[(Z_{ij} - \mu(X_i))^2 | X_i] | Y_i = \mathbf{y}] \\
&\qquad = \frac{m-1}{m^2}\mathbb{E}[(Z_{ij} - \mu(X_i))^2 | Y_i = \mathbf{y}] \text{ by the tower property, and} \\
&\mathbb{E}[(Z_i - \mu(X_i))(\mu(X_i) - k(\mathbf{y})) | Y_i = \mathbf{y}], \\
&\qquad = \mathbb{E}[\mathbb{E}[(Z_i - \mu(X_i))(\mu(X_i) - k(\mathbf{y})) | X_i] | Y_i = \mathbf{y}] = 0,
\end{aligned}$$

we conclude

$$\begin{aligned}
\mathrm{var}[\overline{Z}_i | Y_i = \mathbf{y}] &= \mathbb{E}[C_i^2 | Y_i = \mathbf{y}] + \mathbb{E}[D_i^2 | Y_i = \mathbf{y}] \\
&= \frac{1}{m^2}\mathbb{E}[(Z_i - \mu(X_i))^2 | Y_i = \mathbf{y}] + \mathbb{E}[(\mu(X_i) - k(\mathbf{y}))^2 | Y_i = \mathbf{y}] \\
&\quad + \frac{2}{m}\mathbb{E}[(Z_i - \mu(X_i))(\mu(X_i) - k(\mathbf{y})) | Y_i = \mathbf{y}] \\
&\quad + \frac{m-1}{m^2}\mathbb{E}[(Z_{ij} - \mu(X_i))^2 | Y_i = \mathbf{y}] \\
&= \mathbb{E}[(\mu(X_i) - k(\mathbf{y}))^2 | Y_i = \mathbf{y}] + \frac{1}{m}\mathbb{E}[(Z_i - \mu(X_i))^2 | Y_i = \mathbf{y}]. \qquad (6)
\end{aligned}$$

Combining (5) and (6) yields (4), which completes the proof of Theorem 4.

□

EC.2. CLT for $\overline{k}_n'(\mathbf{y})$

To establish a CLT for $\overline{k}_n'(\mathbf{y})$, we note that

$$\overline{k}_n'(\mathbf{y}) - k(\mathbf{y}) = \frac{\sum_{i=0}^{n-1}\sum_{j=1}^{N_i(\theta\lambda_{in}(\mathbf{y}))}(Z_{ij} - k(\mathbf{y}))}{\sum_{i=0}^{n-1} N_i(\theta\lambda_{in}(\mathbf{y}))} = \frac{\overline{K}_n(\mathbf{y})}{\overline{N}_n(\mathbf{y})},$$

where

$$\overline{K}_n(\mathbf{y}) = \sum_{i=0}^{n-1} \left(\sum_{j=1}^{N_i(\theta\lambda_{in}(\mathbf{y}))} Z_{ij} - \theta\lambda_{in}(\mathbf{y})\mu(X_i) \right)$$

$$+ \theta \sum_{i=0}^{n-1} (\mu(X_i) - k(\mathbf{y}))\lambda_{in}(\mathbf{y}) + k(\mathbf{y}) \sum_{i=0}^{n-1} (\theta\lambda_{in}(\mathbf{y}) - N_i(\theta\lambda_{in}(\mathbf{y})))$$

and

$$\overline{N}_n(\mathbf{y}) = \sum_{i=0}^{n-1} (N_i(\theta\lambda_{in}(\mathbf{y})) - \theta\lambda_{in}(\mathbf{y})) + \theta \sum_{i=0}^{n-1} \lambda_{in}(\mathbf{y}).$$

So,

$$\begin{aligned}
&(na_n^d)^{1/2}(\overline{k}_n'(\mathbf{y}) - k(\mathbf{y})) \frac{\overline{N}_n(\mathbf{y})}{\sum_{i=0}^{n-1} \theta\lambda_{in}(\mathbf{y})} \\
&= \; a_n^{d/2} \frac{n^{-1/2} \sum_{i=0}^{n-1} \left(\sum_{j=1}^{N_i(\theta\lambda_{in}(\mathbf{y}))} Z_{ij} - \theta\lambda_{in}(\mathbf{y})\mu(X_i) \right)}{n^{-1} \sum_{i=0}^{n-1} \theta\lambda_{in}(\mathbf{y})} \\
&\quad + (na_n^d)^{1/2} \frac{\sum_{i=0}^{n-1} (\mu(X_i) - k(\mathbf{y}))\lambda_{in}(\mathbf{y})}{\sum_{i=0}^{n-1} \lambda_{in}(\mathbf{y})} \\
&\quad + a_n^{d/2} \frac{n^{-1/2} k(\mathbf{y}) \sum_{i=0}^{n-1} (\theta\lambda_{in}(\mathbf{y}) - N_i(\theta\lambda_{in}(\mathbf{y})))}{n^{-1} \sum_{i=0}^{n-1} \theta\lambda_{in}(\mathbf{y})}.
\end{aligned}$$

In order to analyze $\overline{k}'(\mathbf{y})$, we need the following properties.

P3. (a) There exists a positive real number σ_5 such that

$$(na_n^d)^{1/2} \frac{\sum_{i=0}^{n-1} (\mu(X_i) - k(\mathbf{y}))\lambda_{in}(\mathbf{y})}{\sum_{i=0}^{n-1} \lambda_{in}(\mathbf{y})} \Rightarrow \sigma_5 N(0,1)$$

as $n \to \infty$.

(b) As $n \to \infty$,

$$n^{-1} \sum_{i=0}^{n-1} \lambda_{in}(\mathbf{y}) \to g(\mathbf{y})$$

a.s.

(c)

$$a_n^{d/2} n^{-1/2} \sum_{i=0}^{n-1} \left(\sum_{j=1}^{N_i(\theta\lambda_{in}(\mathbf{y}))} Z_{ij} - \theta\lambda_{in}(\mathbf{y})\mu(X_i) \right) = o_p(1)$$

and

$$a_n^{d/2} n^{-1/2} \sum_{i=0}^{n-1} (\theta\lambda_{in}(\mathbf{y}) - N_i(\theta\lambda_{in}(\mathbf{y}))) = o_p(1)$$

as $n \to \infty$.

We now present a set of conditions under which P3 holds.

A5. (a) A2 holds with $\tilde{Z}_i$ replaced by $\mu(X_i)$, $\tilde{Y}_i$ replaced by Y_i, and $\tilde{X}_i$ replaced by X_i for $i = 0, 1, \cdots$.

(b) $\int_{\mathbb{R}^d} \lambda(\mathbf{z}) d\mathbf{z} = 1$ and $g(\cdot)$ is continuous at $\mathbf{y}$.

With A5 in place, we have a CLT for $\overline{k}_n'(\mathbf{y})$.

Lemma 5. Let $(X_i : i = 0, 1, \cdots)$ be a Markov chain with the same transition kernel as that of $(\tilde{X}_i : i = 0, 1, \cdots)$. Suppose $(X_i : i = 0, 1, \cdots)$ is initialized from its stationary distribution. Under A0 and A5, P3 holds with

$$\sigma_5^2 = \mathrm{var}[\mu(\tilde{X}_0)|\tilde{Y}_0 = \mathbf{y}] \frac{\int_0^1 \{f_\lambda(x)\}^2 dx}{g(\mathbf{y}) \left(\int_0^1 f_\lambda(x) dx\right)^2}$$

and

$$(na_n^d)^{1/2} (\overline{k}_n'(\mathbf{y}) - k(\mathbf{y})) \Rightarrow \sigma_5 N(0, 1)$$

as $n \to \infty$.

Proof of Lemma 5. Clearly, P3 implies Lemma 5. So, it remains to show that A5 implies P3.

We first note that A5 (a) implies P3 (a) with

$$\sigma_5^2 = \mathrm{var}[\mu(\tilde{X}_0)|\tilde{Y}_0 = \mathbf{y}] \frac{\int_0^1 \{f_\lambda(x)\}^2 dx}{g(\mathbf{y}) \left(\int_0^1 f_\lambda(x) dx\right)^2}$$

by (2.11) and Corollary 2 of Masry (2005).

We next note that A5 (b) implies P3 (b) by Theorem 4.1 (b) of Liebscher (1996).

Next, we establish P3 (c) from the assumption that $\lambda(\cdot)$ is bounded and $\mathbb{E}[|\tilde{Z}_0|] < \infty$. For $i = 0, 1, \cdots$ and $n = 1, 2, \cdots$, let

$$E_{in} = \sum_{j=1}^{N_i(\theta\lambda_{in}(\mathbf{y}))} Z_{ij} - \theta\lambda_{in}(\mathbf{y})\mu(X_i) \text{ and } F_{in} = \theta\lambda_{in}(\mathbf{y}) - N_i(\theta\lambda_{in}(\mathbf{y})),$$

and we will establish

$$n^{-1/2}\sum_{i=0}^{n-1} E_{in} = O_p(1) \text{ and } n^{-1/2}\sum_{i=0}^{n-1} F_{in} = O_p(1),$$

from which P3 (c) will follow.

For $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{\sqrt{n}}\left|\sum_{i=0}^{n-1} E_{in}\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2 n}\mathbb{E}\left[\left(\sum_{i=0}^{n-1} E_{in}\right)^2\right] \\
&= \frac{1}{\epsilon^2 n}\mathbb{E}\left[\mathbb{E}\left[\left(\sum_{i=0}^{n-1} E_{in}\right)^2 \middle| X_0, X_1, \cdots, X_{n-1}\right]\right] \\
&= \frac{1}{\epsilon^2 n}\mathbb{E}\left[\sum_{i=0}^{n-1}\mathbb{E}[E_{in}^2|X_0, X_1, \cdots, X_{n-1}]\right] \\
&= \frac{1}{\epsilon^2 n}\sum_{i=0}^{n-1}\mathbb{E}\left[\mathbb{E}[E_{in}^2|X_0, X_1, \cdots, X_{n-1}]\right] \\
&= \frac{1}{\epsilon^2 n}\sum_{i=0}^{n-1}\mathbb{E}[E_{in}^2] \\
&= \frac{1}{\epsilon^2}\mathbb{E}[E_{0n}^2] \\
&= \frac{1}{\epsilon^2}\mathbb{E}\left[\operatorname{var}\left[\sum_{j=1}^{N_0(\theta\lambda_{0n}(\mathbf{y}))} Z_{0j} \middle| X_0\right]\right] \\
&= \frac{1}{\epsilon^2}\mathbb{E}\left[\mathbb{E}[N_0(\theta\lambda_{0n}(\mathbf{y}))|X_0]\operatorname{var}[Z_{01}|X_0] + (\mathbb{E}[Z_{01}|X_0])^2 \operatorname{var}[N_0(\theta\lambda_{0n}(\mathbf{y}))|X_0]\right] \\
&= \frac{\theta}{\epsilon^2}\mathbb{E}[\lambda_{0n}(\mathbf{y})(\operatorname{var}[Z_0|X_0] + (\mathbb{E}[Z_0|X_0])^2)] \\
&\leq \frac{c\theta}{\epsilon^2}\mathbb{E}[\operatorname{var}[Z_0|X_0] + (\mathbb{E}[Z_0|X_0])^2] \\
&\qquad \text{for some positive constant } c \text{ since } \lambda(\cdot) \text{ is bounded} \\
&= \frac{c\theta}{\epsilon^2}(\operatorname{var}[Z_0] + \mathbb{E}[|Z_0|^2]) < \infty
\end{aligned}$$

by A2 (e). Therefore, $n^{-1/2}\sum_{i=0}^{n-1} E_{in} = O_p(1)$, and

$$a_n^{d/2} n^{-1/2} \sum_{i=0}^{n-1} \left(\sum_{j=1}^{N_i(\theta\lambda_{in}(\mathbf{y}))} Z_{ij} - \theta\lambda_{in}(\mathbf{y})\mu(X_i) \right) = o_p(1)$$

as $n \to \infty$.

On the other hand, for $\epsilon > 0$,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \left| \sum_{i=0}^{n-1} F_{in} \right| \geq \epsilon \right) \leq \frac{1}{\epsilon^2 n} \mathbb{E}\left[\left(\sum_{i=0}^{n-1} F_{in} \right)^2 \right]$$

$$= \frac{1}{\epsilon^2 n} \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{i=0}^{n-1} F_{in} \right)^2 \middle| X_0, X_1, \cdots, X_{n-1} \right] \right]$$

$$= \frac{1}{\epsilon^2} \mathbb{E}\left[F_{0n}^2 \right]$$

$$= \frac{1}{\epsilon^2} \mathbb{E}\left[\mathrm{var}\left[N_0(\theta\lambda_{0n}(\mathbf{y})) | X_0 \right] \right]$$

$$= \frac{1}{\epsilon^2} \mathbb{E}[\theta\lambda_{0n}(\mathbf{y})] < \infty$$

because $\lambda(\cdot)$ is bounded. Therefore, $n^{-1/2}\sum_{i=0}^{n-1} F_{in} = O_p(1)$ and

$$a_n^{d/2} n^{-1/2} \sum_{i=0}^{n-1} (\theta\lambda_{in}(\mathbf{y}) - N_i(\theta\lambda_{in}(\mathbf{y}))) = o_p(1)$$

as $n \to \infty$. Hence, P3 (c) is established. $\square$

Remark. When one compares the asymptotic variances of $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$ in Theorem 4 and Lemma 5, $\overline{k}_n'(\mathbf{y})$ achieves a lower asymptotic variance because

$$\sigma_5^2 = \mathrm{var}[\mu(\tilde{X}_0) | \tilde{Y}_0 = \mathbf{y}] \frac{\int_0^1 \{f_\lambda(x)\}^2 \, dx}{g(\mathbf{y}) \left(\int_0^1 f_\lambda(x) dx \right)^2}$$

$$\leq \left(\mathbb{E}[(\mu(\tilde{X}_i) - k(\mathbf{y}))^2 | \tilde{Y}_i = \mathbf{y}] + \frac{1}{m} \mathbb{E}[(\tilde{Z}_i - \mu(\tilde{X}_i))^2 | \tilde{Y}_i = \mathbf{y}] \right)$$

$$\times \frac{\int_0^1 \{f_\lambda(x)\}^2 \, dx}{g(\mathbf{y}) \left(\int_0^1 f_\lambda(x) dx \right)^2}$$

for any $m = 1, 2, \cdots$.

EC.3. CLT for $k_n^*(\mathbf{y})$

To analyze the asymptotic behavior of $k_n^*(\mathbf{y})$, we need to introduce the following set of assumptions:

A6. (a) $\lambda(0) = 1$, $a_n \to 0$ as $n \to \infty$, and $\sqrt{n} \sup_{\|\mathbf{z}\| > \mu/a_n} |\lambda(\mathbf{z})| \to 0$ as $n \to \infty$, where $\mu = \inf_{\mathbf{z} \in S', \mathbf{z} \neq \mathbf{y}} \|\mathbf{z} - \mathbf{y}\|$.

(b) A3 holds.

The following lemma establishes the CLT for $k_n^*(\mathbf{y})$.

Lemma 6. Suppose $(\tilde{X}_i : i = 0, 1, \cdots)$ is a Harris ergodic (i.e., aperiodic and positive recurrent) Markov chain. Let $(X_i : i = 0, 1, \cdots)$ be a Markov chain with the same transition kernel as that of $(\tilde{X}_i : i = 0, 1, \cdots)$ and an arbitrary initial distribution. Let $(Z_{ij} : i = 0, 1, \cdots, j = 1, 2, \cdots)$ be a family of real–valued rv's satisfying (4.1). Let $\tilde{A}_i \triangleq (\frac{1}{m}(\tilde{Z}_i - \mu(\tilde{X}_i)) + (\mu(\tilde{X}_i) - k(\mathbf{y})))I(\tilde{Y}_i - \mathbf{y})$ for $i = 0, 1, \cdots$. Under A0 and A6,

$$\sigma^2(m) \triangleq \mathrm{var}[\tilde{A}_0] + 2\sum_{i=1}^{\infty} \mathrm{cov}[\tilde{A}_0, \tilde{A}_i]$$

is convergent, and P2 (b) and (c) hold with $\sigma_3' = \eta(\mathbf{y})\sqrt{p(\mathbf{y})}$, where $\eta(\mathbf{y}) = \mathbb{E}[\mathrm{var}[\tilde{Z}_0|\tilde{X}_0]|\tilde{Y}_0 = \mathbf{y}]$. Furthermore, if $\sigma^2(m) > 0$, P2 (a) holds with $\sigma_3 = \sigma(m)$ and, as $n \to \infty$,

$$n^{1/2}(k_n^*(\mathbf{y}) - k(\mathbf{y})) \Rightarrow \frac{\nu(m)}{p(\mathbf{y})} N(0, 1),$$

where

$$\nu^2(m) = \sigma^2(m) + \frac{m-1}{m^2}\eta^2(\mathbf{y})p(\mathbf{y}).$$

Proof of Lemma 6. We will prove that $n^{1/2}(k_n^*(\mathbf{y}) - k_n(\mathbf{y})) \to 0$ a.s. as $n \to \infty$, regardless of how $(X_i : i = 0, 1, \cdots)$ is initialized. Then, Lemma 6 will follow directly from Theorem 3.

Let $\overline{Z}_i = \frac{1}{m}Z_i + \frac{1}{m}\sum_{j=1}^{m-1} Z_{ij}$ for $i = 0, 1, \cdots$ and note

$$\begin{aligned}
&n^{1/2}\left(k_n^*(\mathbf{y}) - k_n(\mathbf{y})\right) \\
&= \frac{n^{-1/2}\sum_{i=0}^{n-1} \overline{Z}_i \lambda_{in}(\mathbf{y})}{n^{-1}\sum_{i=0}^{n-1} \lambda_{in}(\mathbf{y})} - \frac{n^{-1/2}\sum_{i=0}^{n-1} \overline{Z}_i I(Y_i = \mathbf{y})}{n^{-1}\sum_{i=0}^{n-1} I(Y_i = \mathbf{y})}
\end{aligned}$$

$$
= \frac{n^{-1/2}\left[\sum_{i=0}^{n-1}\overline{Z}_i\lambda_{in}(\mathbf{y}) - \sum_{i=0}^{n-1}\overline{Z}_i I(Y_i=\mathbf{y})\right]}{n^{-1}\sum_{i=0}^{n-1}\lambda_{in}(\mathbf{y})}
+ \frac{n^{-1/2}\left[\sum_{i=0}^{n-1} I(Y_i=\mathbf{y}) - \sum_{i=0}^{n-1}\lambda_{in}(\mathbf{y})\right]\cdot n^{-1}\sum_{i=0}^{n-1}\overline{Z}_i I(Y_i=\mathbf{y})}{n^{-1}\sum_{i=0}^{n-1}\lambda_{in}(\mathbf{y})\cdot n^{-1}\sum_{i=0}^{n-1} I(Y_i=\mathbf{y})}. \tag{7}
$$

A6 (a) implies

$$
\begin{aligned}
& n^{-1/2}\left(\sum_{i=0}^{n-1}\overline{Z}_i\lambda_{in}(\mathbf{y}) - \sum_{i=0}^{n-1}\overline{Z}_i I(Y_i=\mathbf{y})\right) \\
&= n^{-1/2}\sum_{i=0}^{n-1}\overline{Z}_i\lambda_{in}(\mathbf{y}) I(Y_i\neq\mathbf{y}) \\
&= \frac{1}{m}n^{-1/2}\sum_{i=0}^{n-1} Z_i\lambda_{in}(\mathbf{y})I(Y_i\neq\mathbf{y}) + \frac{1}{m}n^{-1/2}\sum_{i=0}^{n-1}\left(\sum_{j=1}^{m-1} Z_{ij}\right)\lambda_{in}(\mathbf{y})I(Y_i\neq\mathbf{y}),
\end{aligned}
$$

so

$$
\begin{aligned}
& n^{-1/2}\left|\sum_{i=0}^{n-1}\overline{Z}_i\lambda_{in}(\mathbf{y}) - \sum_{i=0}^{n-1}\overline{Z}_i I(Y_i=\mathbf{y})\right| \\
&\leq \frac{1}{m\sqrt{n}}\sum_{i=0}^{n-1}|Z_i||\lambda_{in}(\mathbf{y})|I(Y_i\neq\mathbf{y}) \\
&\qquad + \frac{1}{m\sqrt{n}}\sum_{i=0}^{n-1}\sum_{j=1}^{m-1}|Z_{ij}||\lambda_{in}(\mathbf{y})|I(Y_i\neq\mathbf{y}) \\
&\leq \frac{1}{mn}\sqrt{n}\sup_{\|\mathbf{z}\|>\mu/a_n}|\lambda(\mathbf{z})|\sum_{i=0}^{n-1}|Z_i| \\
&\qquad + \frac{1}{mn}\sqrt{n}\sup_{\|\mathbf{z}\|>\mu/a_n}|\lambda(\mathbf{z})|\sum_{i=0}^{n-1}\sum_{j=1}^{m-1}|Z_{ij}|.
\end{aligned} \tag{8}
$$

On the other hand, by A0 and A3,

$$
\frac{1}{n}\sum_{i=0}^{n-1}|Z_i| \to \mathbb{E}[Z_0] \tag{9}
$$

a.s. as $n\to\infty$ if $(X_i : i = 0, 1, \cdots)$ is initialized from its stationary distribution. By Proposition 17.1.6 of Meyn and Tweedie (1993), (9) holds regardless of how $(X_i : i = 0, 1, \cdots)$ is initialized. (Similar arguments are applied in the rest of this proof.)

Also, conditional on $X_0, X_1, \cdots$,

$$\frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=1}^{m-1}|Z_{ij}| - \mathbb{E}\left[\frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=1}^{m-1}|Z_{ij}|\middle|X_0, X_1, \cdots\right] \to 0$$

a.s. as $n \to \infty$ by Kolmogorov's strong law (Theorem 2.3.10 of Sen and Singer (1993)) and

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=1}^{m-1}|Z_{ij}|\middle|X_0, X_1, \cdots\right] \\ &\qquad = \frac{m-1}{n}\left(\mathbb{E}[|Z_0||X_0] + \cdots + \mathbb{E}[|Z_{n-1}||X_{n-1}]\right) \to \frac{m-1}{n}\mathbb{E}[|Z_0|]\end{aligned}$$

a.s. as $n \to \infty$ by Theorem 4. Hence,

$$\frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=1}^{m-1}|Z_{ij}| = \frac{m-1}{n}\mathbb{E}[|Z_0|] \tag{10}$$

a.s. as $n \to \infty$.

From A6 (a), (8), (9), and (10),

$$n^{-1/2}\left(\sum_{i=0}^{n-1}\overline{Z}_i\lambda_{in}(\mathbf{y}) - \sum_{i=0}^{n-1}\overline{Z}_i I(Y_i = \mathbf{y})\right) \to 0 \tag{11}$$

a.s. as $n \to \infty$. Similarly,

$$n^{-1/2}\left[\sum_{i=0}^{n-1}\lambda_{in}(\mathbf{y}) - \sum_{i=0}^{n-1}I(Y_i = \mathbf{y}))\right] \to 0 \tag{12}$$

a.s. as $n \to \infty$.

Since $(Y_i : i = 0, 1, \cdots)$ has the strong mixing coefficients $(\eta_n : n = 1, 2, \cdots)$ satisfying $\sum_{n=1}^{\infty}\eta_n^{\delta/(2+\delta)} < \infty$, we have

$$\frac{1}{n}\sum_{i=0}^{n-1}I(Y_i = \mathbf{y}) \to p(\mathbf{y}) \tag{13}$$

a.s. as $n \to \infty$ by Theorem 4. By (12) and (13), we also have

$$\frac{1}{n}\sum_{i=0}^{n-1} \lambda_{in}(\mathbf{y}) \to p(\mathbf{y}) \tag{14}$$

a.s. as $n \to \infty$.

Since (7), (11), (12), (13), and (14) hold regardless of how $(X_i : i = 0, 1, \cdots)$ is initialized (Proposition 17.1.6 of Meyn and Tweedie (1993)), it follows that

$$n^{1/2}\left(k_n^*(\mathbf{y}) - k_n(\mathbf{y})\right) \to 0$$

a.s. as $n \to \infty$ regardless of how $(X_i : i = 0, 1, \cdots)$ is initialized. By Theorem 3, Lemma 6 follows. □

EC.4. Rationale behind $k_n^*(\mathbf{y})$

The rationale behind $k_n^*(\mathbf{y})$ lies in that it is often the case that the system under consideration can be suitably approximated by a continuous state space process (as rigorously supported through a limit theorem).

For example, suppose that there exists a parameter ϵ for which

$$\epsilon^{-s}(\epsilon^q \tilde{Y}_\epsilon(\epsilon^{-w}t) - \mathbf{y}(t)) \Rightarrow \tilde{Y}_0(t) \tag{15}$$

as $\epsilon \downarrow 0$ for $w \geq 0$, $0 \leq s < q$, and $t \geq 0$, where $\tilde{Y}_\epsilon = (\tilde{Y}_\epsilon(t) : t \geq 0)$ is an S'–valued stochastic process for each $\epsilon \geq 0$. Limit theorems such as (15) arise in the setting of multi–server and many–server queues in heavy traffic, as well as many other stochastic modeling contexts; see Whitt (2002). We take the view that $\tilde{Y}$ lives on the integer lattice in $\mathbb{R}^d$ and can be reasonably viewed as having the distribution of $\tilde{Y}_{\epsilon_0}$, for some small (suitably chosen) value ϵ_0.

In this setting,

$$\mathbb{P}\left(\epsilon^{-s}\left(\epsilon^q \tilde{Y}_\epsilon(0) - \mathbf{y}(0)\right) \in \times_{i=1}^{d}[y_i, y_i + \delta z_i]\right) \\ \to \int_{y_1}^{y_1+\delta z_1} \cdots \int_{y_d}^{y_d+\delta z_d} g(w_1, \cdots, w_d)dw_1 \cdots dw_d$$

as $\epsilon \downarrow 0$, provided that $\tilde{Y}_0(0)$ has a density g. It follows that if δ is small, we expect that

$$\mathbb{P}\left(\tilde{Y}_\epsilon(0) \in \times_{i=1}^d [\epsilon^{-q} y_i(0) + \epsilon^{s-q} y_i, \epsilon^{-q} y_i(0) + \epsilon^{s-q} y_i + \epsilon^{s-q}\delta z_i]\right) \approx g(y_1, \cdots, y_d)\delta^d \prod_{i=1}^d z_i. \quad (16)$$

In particular, if $\mathbf{y} = (y_1, \cdots, y_d)^T$ and we put $\delta = \epsilon^{q-s}$ and $z_i = 1$ for $1 \le i \le d$, then we expect that

$$\mathbb{P}\left(\tilde{Y}_\epsilon(0) = \lceil \epsilon^{-q}\mathbf{y}(0) + \epsilon^{s-q}\mathbf{y} \rceil\right) \approx g(\mathbf{y})\epsilon^{(q-s)d}$$

or, equivalently, we expect that

$$\mathbb{P}\left(\tilde{Y}_\epsilon(0) = \mathbf{y}\right) \approx g(\epsilon^{q-s}(\mathbf{y} - \epsilon^{-q}\mathbf{y}(0)))\epsilon^{(q-s)d} \quad (17)$$

when ϵ is small and $\mathbf{y}$ is within a distance of order ϵ^{s-q} from $\epsilon^{-q}\mathbf{y}(0)$.

Remark. Approximations such as (17) depend on "local limit theorems" that often accompany limit theorems of the form (15). Such local limit theorems, in the setting of the heavy traffic limit theorem for multi–server queues, can be found in, for example, Borovkov (1976). The approximation (15) makes clear that $p(\mathbf{y})$ will be (very) small when ϵ_0 is small and d is large.

The approximation (16) makes clear that $p(\cdot)$ can be replaced by the approximating density g, and that we can reasonably expect $p(\cdot)$ to inherit g's smoothness, provided that one is focused on a scale at which p's lattice behavior is not a dominant feature. (By "lattice behavior," we are referring to the fact that p is supported only on the integer lattice points in $\mathbb{R}^d$.) Thus, the estimators $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$ all can be applied in this discrete setting, provided that the bandwidth a_n is large enough that $\epsilon_0^{s-q} a_n$ is large (so that the smoothing enforced by $\lambda(\cdot)$ is enforced on a large number of integer lattice points near y).

EC.5. Simulation–based Prediction when the Underlying Distribution is Unknown

In our discussion, we have assumed full knowledge of the underlying distribution of $\tilde{X}$. As a consequence, we have the ability to simulate X from precisely the distribution underlying $\tilde{X}$. In reality, the distribution underlying $\tilde{X}$ must typically be estimated from observed real–world data. In particular, suppose that the distribution underlying $\tilde{X}$ is modeled parametrically through a family of probabilities $\mathbb{P}_\theta$, where θ^* is the "true" value of the parameter (so that $\mathbb{P}_{\theta^*}$ is the probability underlying $\tilde{X}$). Suppose that $\hat{\theta}$ is the estimator for θ^* obtained from the real–world data set. (The estimator $\hat{\theta}$ could be derived from the principle of maximum likelihood or the method of moments.) We then implement the simulations of Section 3 by simulating the stochastic process X using the distribution $\mathbb{P}_{\hat{\theta}}$. (In other words, we are implementing a "plug–in" estimator.)

Of course, we may be interested in confidence statements for $k(y)$ that incorporate the statistical uncertainty present in our plug–in estimator arising as a consequence of using the estimated distribution $\mathbb{P}_{\hat{\theta}}$ rather than $\mathbb{P}_{\theta^*}$. Let $k(\theta, \mathbf{y}) = \mathbb{E}_\theta(\tilde{Z}_0 | \tilde{Y}_0 = \mathbf{y})$, where $\mathbb{E}_\theta(\cdot)$ is the expectation operator associated with $\mathbb{P}_\theta$. We are interested in computing an approximation to the distribution of $k(\hat{\theta}, \mathbf{y}) - k(\theta^*, \mathbf{y})$.

Assuming our estimator $\hat{\theta}$ is consistent for θ^*, we can approximate

$$k(\hat{\theta}, \mathbf{y}) - k(\theta^*, \mathbf{y}) \approx \nabla k(\hat{\theta}, \mathbf{y})(\hat{\theta} - \theta^*),$$

where we write the gradient of k as a row vector. Assuming, as is usually the case, that $\hat{\theta}$ is asymptotically normal with a covariance matric Ω^* that can be estimated from the real–world data via $\hat{\Omega}$, it follows that $k(\hat{\theta}, \mathbf{y}) - k(\theta^*, \mathbf{y})$ is approximately normal with mean 0 and variance $\nabla k(\hat{\theta}, \mathbf{y}) \hat{\Omega} \nabla k(\hat{\theta}, \mathbf{y})^T$. Thus, the approximate distribution of $k(\hat{\theta}, \mathbf{y}) - k(\theta^*, \mathbf{y})$ is available, once we have an expression for $\nabla k(\hat{\theta}, \mathbf{y})$.

To illustrate the computation of $\nabla k(\hat{\theta}, \mathbf{y})$, we consider the setting in which $\mathbb{P}_\theta(Y = \mathbf{y}) > 0$. Then,

$$k(\hat{\theta}, \mathbf{y}) = \frac{\mathbb{E}_{\hat{\theta}} \tilde{Z}_1 I(\tilde{Y}_1 = \mathbf{y})}{\mathbb{P}_{\hat{\theta}}(\tilde{Y}_1 = \mathbf{y})}.$$

We assume the existence of a likelihood process $(L(\theta, [a,b]), 0 \le a \le b < \infty)$ such that

$$\begin{aligned} &\mathbb{E}[\xi(\tilde{X}(s) : a \le s \le b) \mid \tilde{X}(s) = \mathbf{x}] \\ &\qquad = \mathbb{E}_{\hat{\theta}}[\xi(\tilde{X}(s) : a \le s \le b) L(\theta, [a,b]) \mid \tilde{X}(s) = \mathbf{x}] \end{aligned} \tag{18}$$

for all suitably measurable non–negative path functionals ξ. In the presence of (18), we can write

$$k(\theta,\mathbf{y}) = \frac{\mathbb{E}_\theta \omega_1(\theta, \tilde{X}(1+w)) I(\tilde{Y}_1 = \mathbf{y})}{\mathbb{E}_\theta \omega_2(\theta, \tilde{X}(1+w)) I(\tilde{Y}_1 = \mathbf{y})},$$

where

$$\begin{aligned}
\omega_1(\theta,\mathbf{x}) &= \mathbb{E}_{\hat{\theta}}[\tilde{Z}_1 L(\theta, [1+w, 2+w+t]) \mid \tilde{X}(1+w) = \mathbf{x})], \\
\omega_2(\theta,\mathbf{x}) &= \mathbb{E}_{\hat{\theta}}[L(\theta, [1+w, 2+w+t]) \mid \tilde{X}(1+w) = \mathbf{x})].
\end{aligned}$$

But (18) implies that $\omega_2(\theta,\mathbf{x}) = 1$ for all $\mathbf{x} \in S$ and all θ. So, it follows that

$$k(\theta,\mathbf{y}) = \frac{\int_S \tilde{\omega}_1(\theta,\mathbf{x}) \pi(\theta, d\mathbf{x})}{\int_S \tilde{\omega}_2(\mathbf{x}) \pi(\theta, d\mathbf{x})},$$

where

$$\begin{aligned}
\tilde{\omega}_1(\theta,\mathbf{x}) &= \omega_1(\theta,\mathbf{x}) I(\tau(\mathbf{x}) = \mathbf{y}), \\
\tilde{\omega}_2(\mathbf{x}) &= I(\tau(\mathbf{x}) = \mathbf{y}),
\end{aligned}$$

and $(\pi(\theta, d\mathbf{x}) : \mathbf{x} \in S)$ is the distribution of $\tilde{X}(1+w)$ under $\mathbb{P}_\theta$. We can therefore formally write

$$\begin{aligned}
\nabla k(\hat{\theta},\mathbf{y}) &= \frac{\int_S \nabla \tilde{\omega}_1(\hat{\theta},\mathbf{x}) \pi(\hat{\theta}, d\mathbf{x})}{\int_S \tilde{\omega}_2(\mathbf{x}) \pi(\hat{\theta}, d\mathbf{x})} + \frac{\int_S [\tilde{\omega}_1(\hat{\theta},\mathbf{x}) - k(\hat{\theta},\mathbf{y}) \tilde{\omega}_2(\mathbf{x})] \nabla \pi(\hat{\theta}, d\mathbf{x})}{\int_S \tilde{\omega}_2(\mathbf{x}) \pi(\hat{\theta}, d\mathbf{x})} \\
&= \frac{\mathbb{E}_{\hat{\theta}} \tilde{Z}_1 \nabla L(\hat{\theta}, [1+w, 1+w+t]) I(\tilde{Y}_1 = \mathbf{y})}{\mathbb{P}_{\hat{\theta}}(\tilde{Y}_1 = \mathbf{y})} \\
&\quad + \frac{\int_S \mathbb{E}_{\hat{\theta}}[(\tilde{Z}_1 - k(\hat{\theta},\mathbf{y})) I(\tilde{Y}_1 = \mathbf{y}) \mid \tilde{X}(1+w) = \mathbf{x}] \nabla \pi(\hat{\theta}, d\mathbf{x})}{\mathbb{P}_{\hat{\theta}}(\tilde{Y}_1 = \mathbf{y})}.
\end{aligned}$$

Noting that $\pi(\hat{\theta}, d\mathbf{x})$ is the stationary distribution of the Markov chain $(\tilde{X}(n+w) : n \geq 0)$, we can invoke representation results for $\nabla \pi(\hat{\theta}, d\mathbf{x})$ (see, for example, Glynn and Olvera-Cravioto (2019)), yielding

$$\nabla k(\hat{\theta},\mathbf{y}) = \frac{\mathbb{E}_{\hat{\theta}} \tilde{Z}_1 \nabla L(\hat{\theta}, [1+w, 1+w+t]) I(\tilde{Y}_1 = \mathbf{y})}{\mathbb{P}_{\hat{\theta}}(\tilde{Y}_1 = \mathbf{y})}$$

$$+ \frac{\sum_{n=1}^{\infty} \mathbb{E}_{\hat{\theta}} \nabla L(\hat{\theta}, [w, 1+w])(\tilde{Z}_n - k(\hat{\theta}, \mathbf{y}))I(\tilde{Y}_n = \mathbf{y})}{\mathbb{P}_{\hat{\theta}}(\tilde{Y}_1 = \mathbf{y})}. \quad (19)$$

Each term on the right–hand side of (19) can be estimated via simulation, yielding a potential numerical vehicle towards computing $\nabla k(\hat{\theta}, \mathbf{y})$. A similar approach can be followed in the setting where $\tilde{Y}_1$ is a continuous rv.

EC.6. Closed-form Computation of $k(\mathbf{y})$ in Section 6.1

In this section, we consider the numerical example in Section 6.1 and derive that $k(2) = -1.10$.

We first note that

$$\begin{aligned} \tilde{X}_1(u) &= e^{-u}\tilde{X}_1(0) + \int_0^u e^{-(u-v)} d\tilde{B}_1(v) \text{ and} \\ \tilde{X}_2(u) &= e^{-u}\tilde{X}_2(0) - \int_0^u e^{-(u-v)} \tilde{X}_1(v) dv + \int_0^u e^{-(u-v)} d\tilde{B}_2(v). \end{aligned} \quad (20)$$

Note that $\tilde{X}_1(0) \overset{D}{=} \int_0^\infty e^{-u} dB(-u)$ so

$$\text{var } \tilde{X}_1(\infty) = \int_0^\infty e^{-2u} du = 1/2 \quad \text{and} \quad \mathbb{E}\tilde{X}_1(\infty) = 0.$$

Also,

$$\tilde{X}_2(0) \overset{D}{=} -\int_0^\infty e^{-u} \tilde{X}_1(-u) du + \int_0^\infty e^{-u} dB_2(-u)$$

so

$$\begin{aligned} \text{cov}(\tilde{X}_1(0), \tilde{X}_2(0)) &= \text{cov } \left(\tilde{X}_1(0), \int_0^\infty e^{-u} X_1(-u) du + \int_0^\infty e^{-u} dB_2(-u) \right) \\ &= -\int_0^\infty e^{-u} \text{cov } (\tilde{X}_1(0), \tilde{X}_1(-u)) du. \end{aligned}$$

Furthermore,

$$\text{cov}(\tilde{X}_1(u), \tilde{X}_1(0)) = \text{cov } \left(e^{-u}\tilde{X}_1(0) + \int_0^\infty e^{-u} dB_1(-u),\ \tilde{X}_1(0) \right)$$

$$= e^{-u}\text{var}\tilde{X}_1(0) = e^{-u}/2.$$

So, cov $(\tilde{X}_1(0), \tilde{X}_2(0)) = -\int_0^\infty e^{-2u}/2du = -1/4$. On the other hand,

$$\begin{aligned}
\text{var}\tilde{X}_2(0) &= \text{var}\left(\int_0^\infty e^{-u}X_1(-u)du\right) + \text{var}\left(\int_0^\infty e^{-u}dB_2(-u)\right)\\
&= 2\int_0^\infty e^{-v}\int_v^\infty e^{-z}\text{cov}\,(X_1(v), X_1(z))dzdv + 1/2\\
&= 2\int_0^\infty e^{-v}\int_v^\infty e^{-z}e^{-(z-v)}/2dzdv + 1/2 = 3/4
\end{aligned}$$

Suppose that $\tilde{Y}(u) = \tilde{X}_1(u)$ and $\tilde{Z}(u) = \tilde{X}_2(u)$. Then

$$\begin{aligned}
&\mathbb{E}[\tilde{Z}(s+t)|\tilde{Y}(s)] = \mathbb{E}[\tilde{X}_2(t)|\tilde{X}_1(0)]\\
&= e^{-t}\mathbb{E}[\tilde{X}_2(0)|\tilde{X}_1(0)] - \int_0^t e^{-(t-s)}\mathbb{E}[\tilde{X}_1(s)|\tilde{X}_1(0)]dt\\
&= e^{-t}\frac{\text{cov}(\tilde{X}_2(0), \tilde{X}_1(0))}{\text{var}\tilde{X}_1(0)}(\tilde{X}_1(0) - \mathbb{E}\tilde{X}_1(0)) - \int_0^t e^{-(t-s)}e^{-s}\tilde{X}_1(0)ds\\
&= -\frac{1}{2}e^{-t}\tilde{X}_1(0) - te^{-t}\tilde{X}_1(0).
\end{aligned}$$

So,

$$\mathbb{E}[\tilde{Z}(s+t)|\tilde{Y}(s)] = \mathbb{E}[\tilde{X}_2(1)|\tilde{X}_1(0) = 2] = (-1-2)e^{-1} = -1.104.$$

EC.7. Selecting the Kernel for $k_n'(\mathbf{y})$

In this section, we observe the numerical behavior of $k_n'(\mathbf{y})$ for different types of kernel $\lambda(\cdot)$. We consider the two–dimensional Ornstein–Uhlenbeck process $\tilde{X} = ((\tilde{X}_1\ (u), \tilde{X}_2(u)) : u \geq 0)$ described in Section 6.1. The goal is to estimate $k(2) = \mathbb{E}[\tilde{X}_2(s+1)|\tilde{X}_1(s) = 2]$.

We tested three types of kernels: (1) the uniform kernel over $[-1, 1]$, i.e.,

$$\lambda(z) = \begin{cases} 1/2, & \text{if } -1 \leq z \leq 1\\ 0, & \text{otherwise,}\end{cases}$$

(2) the Epanechnikov kernel, i.e.,

$$\lambda(z) = \begin{cases} 1 - z^2, & \text{if } -1 \leq z \leq 1\\ 0, & \text{otherwise,}\end{cases}$$

and (3) the triangular kernel with the lower limit at -1, the upper limit at 1, and the mode at 0, i.e.,

$$\lambda(z) = \begin{cases} z+1, & \text{if } -1 \le z \le 0 \\ -z+1, & \text{if } 0 \le z \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

For each n, we generated (X_i, Y_i, Z_i) along with the Z_{ij}'s for $0 \le i \le n-1$ with X initialized at $(0,0)$. Using the (X_i, Y_i, Z_i)'s and Z_{ij}'s, we computed $k_n'(2)$ using the uniform kernel, the Epanechnikov kernel, and the triangular kernel, respectively. It should be noted that the three kernels have the same support, so the same (X_i, Y_i, Z_i)'s and Z_{ij}'s can be used to compute $k_n'(2)$ associated with each type of kernel. We then computed $(k_n'(2) - k(2))^2$ as an estimate of the mean square error (MSE) of $k_n'(2)$. It should be noted that $k(2)$ can be computed analytically and is given by $k(2) = -1.10$; see EC.6. for the detailed derivation of $k(2) = -1.10$. This procedure was repeated 1000 times independently, generating 1000 copies of $(k_n'(2) - k(2))^2$ for each type of kernel. Using the 1000 copies, we computed the 95% confidence intervals of the MSE of $k_n'(2)$. Tables 1 and 2 report the 95% confidence intervals for various values of m, $n = 5,000, n = 10,000$, and $a_n = 0.5n^{-1/5}$.

The results show that, of the three types of kernels we tested, the uniform kernel produces the lowest mean square error and the triangular kernel the highest mean square error. This suggests that applying a uniform weight, rather than a weight that decays as Y_i gets farther away from $\mathbf{y}$, yields better performance.

We conducted all simulations in this section using a 64-bit computer with an Intel(R) Core(TM) i5-6200U CPU at 2.3 GHz and a RAM of 8 GB, and programmed all simulations in MATLAB R2019a.

EC.8. Comparisons between $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$

In this section, we compare the performance of $k_n'(\mathbf{y})$ to that of $\overline{k}_n'(\mathbf{y})$ using the two–dimensional Ornstein–Uhlenbeck process $\tilde{X} = ((\tilde{X}_1(u), \tilde{X}_2(u)) : u \ge 0)$ described in Section 6.1. The goal is to estimate $k(2) = \mathbb{E}[\tilde{X}_2(s+1)|\tilde{X}_1(s) = 2]$.

For λ, we used the triangular distribution with the lower limit at -1, the

Table 1: The 95% confidence intervals of the MSE of $k_n'(2)$ when $n = 5000$.

m	Uniform kernel	Epanechnikov kernel	Triangular kernel
1	0.0711 ± 0.0082	0.0887 ± 0.0101	0.0975 ± 0.0110
10	0.0172 ± 0.0019	0.0200 ± 0.0022	0.0225 ± 0.0025
100	0.0111 ± 0.0011	0.0134 ± 0.0014	0.0146 ± 0.0015

Table 2: The 95% confidence intervals of the MSE of $k_n'(2)$ when $n = 10000$.

m	Uniform kernel	Epanechnikov kernel	Triangular kernel
1	0.0386 ± 0.0037	0.0471 ± 0.0044	0.0526 ± 0.0048
10	0.0089 ± 0.0008	0.0109 ± 0.0010	0.0121 ± 0.0011
100	0.0054 ± 0.0005	0.0062 ± 0.0006	0.0068 ± 0.0007

upper limit at 1, and the mode at 0, i.e.,

$$\lambda(z) = \begin{cases} z+1, & \text{if } -1 \le z \le 0 \\ -z+1, & \text{if } 0 \le z \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

For each n, we generated (X_i, Y_i, Z_i) for $0 \le i \le n-1$ with X initialized at $(0,0)$. Using the (X_i, Y_i, Z_i)'s, we generated the Z_{ij}'s and used them to compute $k_n'(\mathbf{y})$. Using the same (X_i, Y_i, Z_i)'s, we generated another set of the Z_{ij}'s and used them to compute $\overline{k}_n'(\mathbf{y})$. We then computed $(k_n'(\mathbf{y}) - k(2))^2$ and $(\overline{k}_n'(\mathbf{y}) - k(2))^2$ as estimates of the MSEs of $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$, respectively. It should be noted that $k(2)$ can be computed analytically and is given by $k(2) = -1.10$; see EC.6. for the detailed derivation of $k(2) = -1.10$. This procedure was repeated 1000 times independently, generating 1000 copies of each of $(k_n'(\mathbf{y}) - k(2))^2$ and $(\overline{k}_n'(\mathbf{y}) - k(2))^2$. Using these copies, we computed the 95% confidence intervals of the MSEs of $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$. Tables 3, 4, and 5 report the computer time (T) required to compute $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$, and the 95% confidence intervals of $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$ (CI) for the cases when $a_n = 0.1n^{-1/5}$, $a_n = 2n^{-1/5}$, and $a_n = 5n^{-1/5}$, respectively. For each case, we report the results for the cases when $m = \theta = 50$ and $m = \theta = 200$.

The results show that when m, θ, and n are small, $k_n'(\mathbf{y})$ produces a lower mean square error. However, when m, θ, and n get larger, the difference

Table 3: The computer time (T) required to compute $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$, and the 95% confidence intervals of the MSEs of $k_n'(\mathbf{y})$ and $\overline{k}_n'(\mathbf{y})$ when $a_n = 0.1n^{-1/5}$.

	$k_n'(2)$ with $m = 50$		$\overline{k}_n'(2)$ with $\theta = 50$	
n	T	CI	T	CI
5000	0.48	0.2431 ± 0.0267	0.47	0.2649 ± 0.0275
40000	3.80	0.0143 ± 0.0014	3.78	0.0155 ± 0.0015
80000	7.60	0.0082 ± 0.0008	7.56	0.0086 ± 0.0008
	$k_n'(2)$ with $m = 200$		$\overline{k}_n'(2)$ with $\theta = 200$	
n	T	CI	T	CI
5000	0.50	0.2242 ± 0.0261	0.48	0.2309 ± 0.0262
40000	3.95	0.0143 ± 0.0014	3.85	0.0143 ± 0.0014
80000	7.85	0.0072 ± 0.0007	7.68	0.0075 ± 0.0007

between the mean square error of $k_n'(\mathbf{y})$ and that of $\overline{k}_n'(\mathbf{y})$ becomes negligible. In all cases, it requires less computer time to compute $\overline{k}_n'(\mathbf{y})$ than to compute $k_n'(\mathbf{y})$. This is intuitively acceptable because when computing $\overline{k}_n'(\mathbf{y})$, we generate less Z_{ij}'s for Y_i that is far away from $\mathbf{y}$ (since $\lambda((Y_i - \mathbf{y})/a_n)$ decays as Y_i gets farther away from $\mathbf{y}$). On the other hand, when computing $k_n'(\mathbf{y})$, we generate a fixed number of Z_{ij}'s regardless of how far away Y_i is from $\mathbf{y}$. We can thus conclude that, when m, θ, and n are sufficiently large, $\overline{k}_n'(\mathbf{y})$ produces a similar mean square error to that of $k_n'(\mathbf{y})$ with less computer time.

We conducted all simulations in this section using a 64-bit computer with an Intel(R) Core(TM) i5-6200U CPU at 2.3 GHz and a RAM of 8 GB, and programmed all simulations in MATLAB R2019a.

Table 4: The computer time (T) required to compute $k'_n(\mathbf{y})$ and $\overline{k}'_n(\mathbf{y})$, and the 95% confidence intervals of the MSEs of $k'_n(\mathbf{y})$ and $\overline{k}'_n(\mathbf{y})$ when $a_n = 2n^{-1/5}$.

	$k'_n(2)$ with $m = 50$		$\overline{k}'_n(2)$ with $\theta = 50$	
n	T	CI	T	CI
5000	0.74	0.0050 ± 0.0004	0.60	0.0057 ± 0.0005
40000	5.30	0.0011 ± 0.0001	4.67	0.0011 ± 0.0001
80000	9.63	0.0006 ± 0.0001	8.65	0.0006 ± 0.0001
	$k'_n(2)$ with $m = 200$		$\overline{k}'_n(2)$ with $\theta = 200$	
n	T	CI	T	CI
5000	1.51	0.0048 ± 0.0004	0.94	0.0049 ± 0.0004
40000	8.38	0.0010 ± 0.0001	5.97	0.0010 ± 0.0001
80000	15.37	0.0006 ± 0.0001	11.38	0.0006 ± 0.0001

Table 5: The computer time (T) required to compute $k'_n(\mathbf{y})$ and $\overline{k}'_n(\mathbf{y})$, and the 95% confidence intervals of the MSEs of $k'_n(\mathbf{y})$ and $\overline{k}'_n(\mathbf{y})$ when $a_n = 5n^{-1/5}$.

	$k'_n(2)$ with $m = 50$		$\overline{k}'_n(2)$ with $\theta = 50$	
n	T	CI	T	CI
5000	2.04	0.0518 ± 0.0008	0.99	0.0519 ± 0.0008
40000	9.23	0.0132 ± 0.0002	6.09	0.0132 ± 0.0002
80000	16.01	0.0081 ± 0.0001	11.35	0.0081 ± 0.0001
	$k'_n(2)$ with $m = 200$		$\overline{k}'_n(2)$ with $\theta = 200$	
n	T	CI	T	CI
5000	6.74	0.0527 ± 0.0008	2.49	0.0527 ± 0.0008
40000	25.43	0.0131 ± 0.0002	12.31	0.0131 ± 0.0002
80000	38.96	0.0081 ± 0.0001	20.41	0.0081 ± 0.0001

EC.9. Selecting the Bandwidth $(a_n : n \geq 1)$ via Cross Validation

In this section, we demonstrate how to select the bandwidth $(a_n : n \geq 1)$ for $k'_n(\mathbf{y})$ using leave-one-out cross validation. We consider the two–dimensional Ornstein–Uhlenbeck process $\tilde{X} = ((\tilde{X}_1(u), \tilde{X}_2(u)) : u \geq 0)$ described in Section 6.1. The goal is to estimate $k(2) = \mathbb{E}[\tilde{X}_2(s+1)|\tilde{X}_1(s) = 2]$.

We set $\lambda(z) = 0.1$ if $-5 \leq z \leq 5$, and $\lambda(z) = 0$ otherwise. We initialized X at $(0,0)$ and generated (X_i, Y_i, Z_i) along with the Z_{ij}'s for $i = 0, \cdots, n-1$ and $j = 1, \cdots, m-1$. It should be noted that the Z_{ij}'s should be generated for each $i = 0, \cdots, n-1$.

For the bandwidth sequence $(a_n : n \geq 1)$, we tested five different sequences $(a_n^{(1)} : n \geq 1)$, $(a_n^{(2)} : n \geq 1)$, $(a_n^{(3)} : n \geq 1)$, $(a_n^{(4)} : n \geq 1)$, and $(a_n^{(5)} : n \geq 1)$. $a_n^{(1)}$ and $a_n^{(2)}$ were computed based on cross validation while $a_n^{(3)}$, $a_n^{(4)}$, and $a_n^{(5)}$ were given by $a_n^{(3)} = 0.5n^{-1/5}, a_n^{(4)} = 0.1n^{-1/5}$, and $a_n^{(5)} = 0.08n^{-1/5}$ for $n = 1, 2, \cdots$.

More specifically, $a_n^{(1)}$ is the minimizer of

$$\mathrm{CV}(\zeta) \triangleq \frac{1}{n}\sum_{i=0}^{n-1}(k'_{-i,\zeta}(Y_i) - \overline{Z}_i)^2$$

over $\zeta \in \{0.01, 0.02, \cdots, 0.10\}$, where $\overline{Z}_i = \frac{1}{m}Z_i + \frac{1}{m}\sum_{j=1}^{m-1} Z_{ij}$ for $i = 0, 1, \cdots, n-1$ and $k'_{-i,\zeta}(Y_i)$ is calculated from

$$k'_{-i,\zeta}(Y_i) = \frac{\sum_{k=0,k\neq i}^{n-1} \overline{Z}_k \lambda((Y_k - Y_i)/\zeta)}{\sum_{k=0,k\neq i}^{n-1} \lambda((Y_k - Y_i)/\zeta)}.$$

When $\sum_{k=0,k\neq i}^{n-1} \lambda((Y_k - Y_i)/\zeta) = 0$, $k'_{-i,\zeta}(Y_i)$ is set to be 0.

On the other hand, $a_n^{(2)}$ is the minimizer of

$$\mathrm{CV}_*(\zeta) \triangleq \frac{1}{n}\sum_{i=0}^{n-1}(k'_{-i,\zeta,*}(Y_i) - Z_i)^2$$

over $\zeta \in \{0.01, 0.02, \cdots, 0.10\}$, where $k'_{-i,\zeta,*}(Y_i)$ is calculated from

$$k'_{-i,\zeta,*}(Y_i) = \frac{\sum_{k=0,k\neq i}^{n-1} Z_k \lambda((Y_k - Y_i)/\zeta)}{\sum_{k=0,k\neq i}^{n-1} \lambda((Y_k - Y_i)/\zeta)}.$$

Table 6: The 95% confidence intervals of the MSE of $k'_n(2)$ when the bandwidth is chosen from $a_n^{(1)}, a_n^{(2)}, a_n^{(3)}, a_n^{(4)}, a_n^{(5)}$ with $m = 10$.

n	$a_n = a_n^{(1)}$	$a_n = a_n^{(2)}$
5000	0.0077 ± 0.0011	0.0076 ± 0.0009
10000	0.0049 ± 0.0006	0.0050 ± 0.0006
20000	0.0026 ± 0.0003	0.0029 ± 0.0004

n	$a_n = a_n^{(3)}$	$a_n = a_n^{(4)}$	$a_n = a_n^{(5)}$
5000	0.0176 ± 0.0012	0.0152 ± 0.0023	0.0221 ± 0.0042
10000	0.0107 ± 0.0007	0.0093 ± 0.0013	0.0121 ± 0.0018
20000	0.0066 ± 0.0004	0.0051 ± 0.0007	0.0070 ± 0.0010

When $\sum_{k=0,k\neq i}^{n-1} \lambda((Y_k - Y_i)/\zeta) = 0$, $k'_{-i,\zeta,*}(Y_i)$ is set to be 0.

In each case of $a_n = a_n^{(1)}, a_n^{(2)}, a_n^{(3)}, a_n^{(4)}$ and $a_n^{(5)}$, we computed $k'_n(2)$ using the same (X_i, Y_i, Z_i) and the Z_{ij}'s generated, and computed $(k'_n(2) - k(2))^2$ as an estimate of the MSE of $k'_n(2)$. It should be noted that $k(2)$ can be computed analytically and is given by $k(2) = -1.10$; see EC.6. for the detailed derivation of $k(2) = -1.10$. This procedure was repeated 400 times independently, generating 400 independent copies of $(k'_n(2) - k(2))^2$ for each choice of the bandwidth. Using these 400 values in each case, we computed the 95% confidence interval of the MSE of $k'_n(2)$.

Tables 6 and 7 report the 95% confidence intervals of the MSE of $k'_n(2)$ for a variety of n values when $m = 10$ and $m = 50$, respectively. The bandwidths $a_n^{(1)}$ and $a_n^{(2)}$ selected from cross validation show superior performance compared to fixed sequences $a_n^{(3)}, a_n^{(4)}$, and $a_n^{(5)}$.

We conducted all simulations in this section using a 64-bit computer with an Intel(R) Core(TM) i5-6200U CPU at 2.3 GHz and a RAM of 8 GB, and programmed all simulations in MATLAB R2019a.

References

Borovkov, A. A. (1976). *Stochastic Processes in Queueing Theory.* Springer–Verlag, New York.

Table 7: The 95% confidence intervals of the MSE of $k_n'(2)$ when the bandwidth is chosen from $a_n^{(1)}, a_n^{(2)}, a_n^{(3)}, a_n^{(4)}, a_n^{(5)}$ with $m = 50$.

n	$a_n = a_n^{(1)}$	$a_n = a_n^{(2)}$
5000	0.0055 ± 0.0005	0.0063 ± 0.0008
10000	0.0037 ± 0.0005	0.0039 ± 0.0004
20000	0.0021 ± 0.0003	0.0023 ± 0.0003

n	$a_n = a_n^{(3)}$	$a_n = a_n^{(4)}$	$a_n = a_n^{(5)}$
5000	0.0167 ± 0.0009	0.0102 ± 0.0017	0.0133 ± 0.0021
10000	0.0105 ± 0.0006	0.0068 ± 0.0010	0.0084 ± 0.0012
20000	0.0068 ± 0.0004	0.0037 ± 0.0005	0.0046 ± 0.0006

Glynn, P. W. and Olvera-Cravioto, M. (2019). Likelihood ratio gradient estimation for steady–state parameters. *Stoch. Syst.*, 9:83–100.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.*, 7:349–382.

Liebscher, E. (1996). Strong convergence of sums of α–mixing random variables with applications to density estimation. *Stochastic Process. Appl.*, 65:69–80.

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Process. Appl.*, 115:155–177.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer–Verlag, London.

Rio, E. (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. Henri. Poincaré*, 29:587–597.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman & Hall, London.

Whitt, W. (2002). *Stochastic–process Limits: An Introduction to Stochastic-process Limits and their Application to Queues*. Springer, New York.