Change-point Detection for Sparse and Dense Functional Data in General Dimensions (Supplementary Materials)

Carlos Misael Madrid Padilla
Department of Mathematics
University of Notre Dame
cmadridp@nd.edu

Daren Wang
Department of Statistics
University of Notre Dame
dwang24@nd.edu

Zifeng Zhao
Mendoza College of Business
University of Notre Dame
zzhao2@nd.edu

Yi Yu
Department of Statistics
University of Warwick
yi.yu.2@warwick.ac.uk

Additional numerical results and all technical details are included in the supplementary materials.

A Additional numerical results

A.1 Detailed simulation results

We present the tables containing the results of the simulation study in Section 4.1 of the main text. On each table, the mean over 100 repetitions is reported, and the numbers in parenthesis denote the standard errors. For the purpose of identifying underestimation and overestimation, we also include the proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive.

Table 1: **Scenario 1** ($n = 1, d = 1$ changes from $6\cos$-$6\sin$-$6\cos$)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\|\hat{K} - K\|$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0.05	0.86	0.09	0.17 (0.05)	16.15 (4.09)

Changes occur at the times 30 and 130.

Table 2: **Scenario 2** ($n = 10, d = 1$, changes from $2\cos$-$2\sin$-$2\cos$)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\|\hat{K} - K\|$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0.05	0.95	0	0.05 (0.02)	3.32 (1)
BGHK	0.58	0.42	0	1.12 (0.14)	20.11 (1.82)
HK	0.16	0.47	0.37	0.78 (0.08)	66.45 (7.87)
SN	0.04	0.03	0.93	1.83 (0.04)	181.11 (5.05)

Changes occur at the times 30 and 130.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Table 3: **Scenario 3** ($n = 50, d = 1$, changes from cos-sin-cos)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\|\hat{K} - K\|$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0	0.93	0.07	0.07 (0.03)	7.35 (0.54)
BGHK	0.85	0.15	0	2.97 (0.22)	32.88 (1.82)
HK	0	0.08	0.92	1.71 (0.06)	172.52 (5.61)
SN	0.02	0.04	0.94	1.85 (0.05)	183.63 (4.57)

Changes occur at the times 30 and 130.

Table 4: **Scenario 4** ($n = 10, d = 2$, changes from 0-3$x^{(1)}x^{(2)}$-0)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\|\hat{K} - K\|$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0	0.92	0.08	0.08 (0.021)	5.02 (1.25)

Changes occur at the times 100 and 150.

Table 5: **Scenario 5** ($n = 50, d = 1$, changes from 0-sin-2 sin)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\|\hat{K} - K\|$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0.02	0.98	0	0.02 (0.01)	16.9 (0.93)
BGHK	0.48	0.30	0.22	1.09 (0.11)	34.36 (1.78)
HK	0	0.19	0.81	0.81 (0.04)	48.24 (1.71)
SN	0.08	0.33	0.59	0.85 (0.07)	65.15 (6.38)

Changes occur at the times 68 and 134.

A.2 Details of Figure 2

We zoom in the top-left and top-right corners of each panel in Figure 2 and present in Figures A.2 and A.2, respectively. The top-left and top-right corners correspond to the northwest and northeast coasts of Australia, where the changes occur.

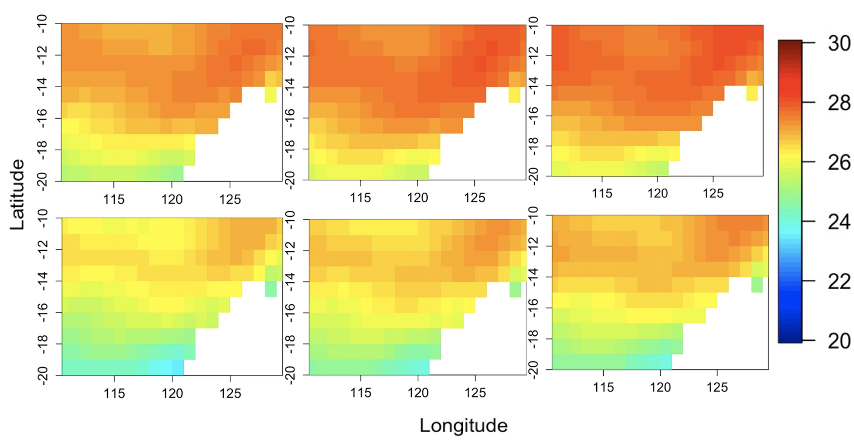

Figure 1: Average SST of northwest coast of Australia. From left to right: average SST from 1940 to 1981, average SST from 1982 to 1996, and average SST from 1997 to 2019. The top and bottom rows correspond to the June and July data respectively.

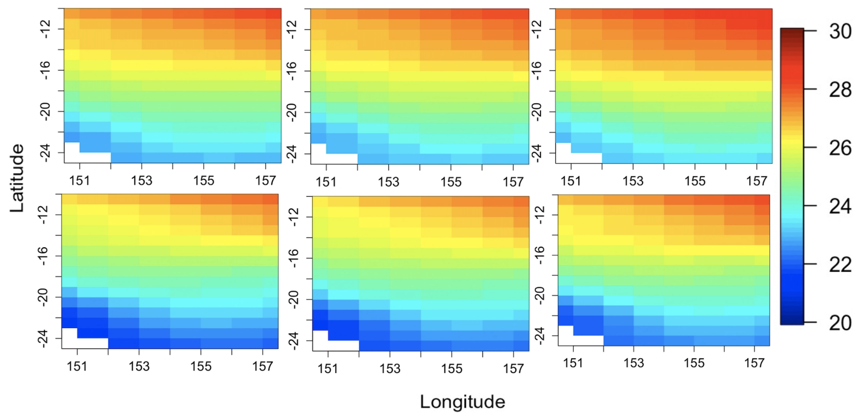

Figure 2: Average SST of northeast coast of Australia. From left to right: average SST from 1940 to 1981, average SST from 1982 to 1996, and average SST from 1997 to 2019. The top and bottom rows correspond to the June and July data respectively.

A.3 Sea surface temperature on Caribbean sea

We consider an additional real data example, also from the COBE-SSTE dataset [7], using data from June and July. FSBS is applied to estimate potential change points on a 1 degree latitude by 1 degree longitude grid (10×6), located at the Caribbean sea. In both months, FSBS identified the year 2004 as a change-point. This might be associated with the development of a Modoki El Niño – a rare type of El Niño in which unfavourable conditions are produced over the eastern Pacific instead of the Atlantic basin due to warmer sea surface temperatures farther west along the equatorial Pacific [13]. Variability in the climate of northeastern Caribbean is connected with this phenomenon, see for example [6].

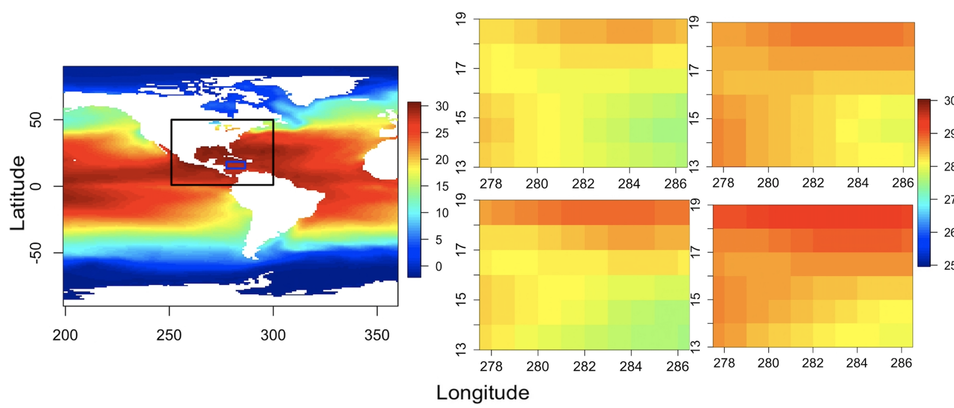

Figure 3: Average SST of Caribbean sea. From left to right: The first image shows the region chosen, the small blue rectangle into the black rectangle. The second image contains four different sub-images. Here, from left to right, the average SST from 1940 to 2003 and average SST from 2004 to 2019 is presented. The top and bottom rows correspond to the June and July data respectively.

A.4 On the dimension d

Recall that the localisation error rate of change-point estimation in Theorem 1 is

$$C_{\mathrm{FSBS}} \log^{\max\{1,10/q\}}(T) \left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right) \kappa_k^{-2},$$

which is an increasing function of d, i.e. a larger d will lead to a worse localization error rate.

In addition, Assumption 3 requires that the signal-to-noise ratio to be lower bounded by

$$C_{\mathrm{SNR}} \log^{\max\{1/2,5/q\}}(T)\Big(1+T^{\frac{d}{2r+d}}n^{\frac{-2r}{2r+d}}\Big)^{1/2},$$

which implies that a larger d will also require a stronger signal.

We conducted additional numerical results to further show the influence of d. Using the same setting as that in Scenario 4 in Section 4, we vary the dimension $d \in \{2, 3, 5, 10\}$. Results are collected in Table 6 and Appendix A.4, supporting our theoretical findings.

Table 6: **FSBS on Scenario 4** ($n = 10$, changes from 0-3$x^{(1)}x^{(2)}$-0)

Dimension	$K-\hat{K}<0$	$K-\hat{K}=0$	$K-\hat{K}>0$	$\lvert\hat{K}-K\rvert$	$d(\hat{\mathcal{C}},\mathcal{C})$
$d=2$	0	0.92	0.08	0.08 (0.02)	5.02 (1.25)
$d=3$	0.02	0.89	0.09	0.11 (0.03)	5.73 (1.22)
$d=5$	0.18	0.82	0	0.18 (0.05)	5.92 (1.23)
$d=10$	0.21	0.79	0	0.22 (0.08)	6.58 (1.24)

Performance of FSBS with different choices of dimension d is studied for **S4**. The mean over 100 repetitions is reported, and the numbers in parenthesis denote standard errors. It includes the proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive.

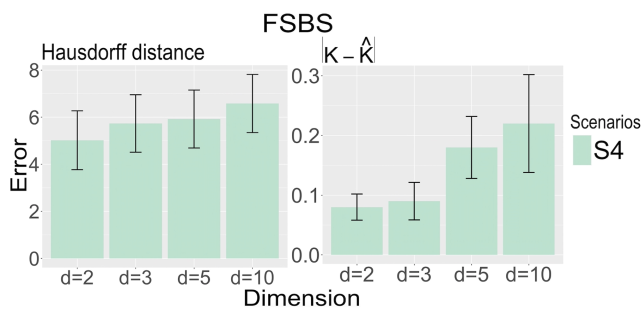

Figure 4: Bar plots for simulation results of FSBS performance on **S4** with respect to the dimension d. Each bar reports the mean and standard error computed based on 100 experiments.

A.5 Choice of kernels

The choice of kernels may affect the performance of kernel based methods. We choose Gaussian kernel in Section 4 and demonstrate the robustness against the choice of kernels of FSBS in this section, by choosing different kernels. Tables 7, 8 and Appendix A.5 collect results of the performance of the FSBS with different choices of kernels, based on the settings detailed in Scenarios 1 and 2 in Section 4, with Gaussian, Uniform, Epanechnikov and Quartic kernels.

Table 7: **FSBS in Scenario 1 (different kernels comparison)**

Kernel	$K-\hat{K}<0$	$K-\hat{K}=0$	$K-\hat{K}>0$	$\lvert\hat{K}-K\rvert$	$d(\hat{\mathcal{C}},\mathcal{C})$
Gaussian	0.05	0.86	0.09	0.17 (0.05)	16.15 (4.09)
Uniform	0.01	0.99	0	0.01 (0.01)	13.32 (0.42)
Epanechnikov	0.06	0.87	0.17	0.13 (0.03)	15.14 (2.40)
Quartic	0.07	0.84	0.09	0.20 (0.04)	18.28 (1.63)

The mean over 100 repetitions is reported together with the standard errors into parenthesis. The proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive are included.

Table 8: **FSBS in Scenario 2 (different kernels comparison)**

Kernel	$K-\hat{K}<0$	$K-\hat{K}=0$	$K-\hat{K}>0$	$\lvert\hat{K}-K\rvert$	$d(\hat{\mathcal{C}},\mathcal{C})$
Gaussian	0.05	0.95	0	0.05 (0.02)	3.32 (1)
Uniform	0	0.99	0.01	0.01 (0.01)	2.93 (1.03)
Epanechnikov	0	100	0	0 (0)	1.24 (0.28)
Quartic	0.01	0.99	0	0.01 (0.01)	2.3 (0.55)

The mean over 100 repetitions is reported together with the standard errors into parenthesis. The proportions of estimations for which the $\hat{K}-K$ distance is negative, zero, or positive are included.

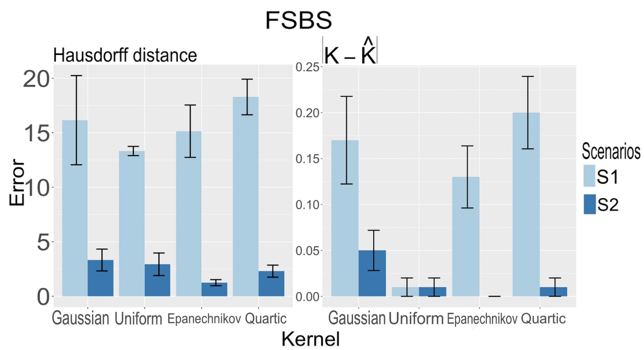

Figure 5: Bar plots for simulation results of FSBS performance on **S1** and **S2** with respect to different choices of kernels. Each bar reports the mean and standard error computed based on 100 experiments.

A.6 Computational costs

Our method is computationally efficient and its computational complexity is $O(nT\log T + T(\log T)^2)$. Specifically, as can be seen from Algorithm 1, we need to conduct kernel smoothing of the sampling distribution and mean function at $\log T$ measurement locations, which costs $O(nT\log T)$ operations. Once this is done, we conduct seeded binary segmentation (SBS) at the $\log T$ measurement locations/grids. It is known that SBS has a computational cost of $O(T\log T)$. Thus, this step costs $O(T(\log T)^2)$ computational complexity. In total, the computational complexity of our method is $O(nT\log T + T(\log T)^2)$.

As for existing methods in the literature, in terms of implementation, they all rely on the two-stage procedure. Specifically, the first stage is to register/estimate the discretely observed points into a functional curve on each time t. Taking the B-spline smoothing with p basis functions for example, this costs $O(n^2p + p^3)$ computational complexity for each time t due to a least square estimation. Thus this step costs $O(T(n^2p + p^3))$ computational complexity. Once the functional curves are registered, in the second stage, the existing methods conduct functional PCA to extract p' principle component scores from each function and then conduct mean change-point detection on the p'-dimension time series of principle component scores. Ignoring the computational cost of functional PCA, the change-point detection procedure costs at least $O(T\log T)$ computational complexity if a standard binary segmentation is used and could be more expensive if other segmentation algorithms are used to conduct change-point estimation. Thus, in total, the computational complexity of existing methods is at least $O(T(n^2p + p^3) + T\log T)$, which is more expensive unless $n \preceq \log T$.

B Proof of Theorem 1

In this section, we present the proofs of theorem Theorem 1. To this end, we will invoke the following well-known l_∞ bounds for kernel density estimation.

Lemma 1. *Let $\{x_{t,i}\}_{i=1,t=1}^{n,T}$ be random grid points independently sampled from a common density function $u : [0,1]^d \to \mathbb{R}$. Under Assumption 2-**b**, the density estimator of the sampling distribution μ,*

$$\hat{p}(x) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n_t} K_{\bar{h}}(x - x_{i,t}), \quad x \in [0,1]^d,$$

satisfies,

$$||\hat{p} - \mathbb{E}(\hat{p})||_\infty \le C \sqrt{\frac{\log(nT) + \log(1/\bar{h})}{nT\bar{h}^d}} \tag{1}$$

*with probability at least $1 - \frac{1}{nT}$. Moreover, under Assumption 2-**a**, the bias term satisfies*

$$||\mathbb{E}(\hat{p}) - u||_\infty \le C_2 \bar{h}^r. \tag{2}$$

Therefore,

$$||\hat{p} - u||_\infty = O\Big(\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \tag{3}$$

with probability at least $1 - \frac{1}{nT}$.

The verification of these bounds can be found in many places in the literature. For equation (1) see for example [1], [8], [9] and [2]. For equation (2), [10] is a common reference.

Proof of Theorem 1. For any $(s, e] \subseteq (0, T]$, let

$$\widetilde{f}_t^{(s,e]}(x) = \sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{l=s+1}^{t} f_l^*(x) - \sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{l=t+1}^{e} f_l^*(x), \ x \in [0,1]^d.$$

For any $\tilde{r} \in (\rho, T-\rho]$ and $x \in [0,1]$, we consider

$$\mathcal{A}_x((s,e], \rho, \lambda) = \Big\{ \max_{t=s+\rho+1}^{e-\rho} |\widetilde{F}_{t,h}^{s,e}(x) - \widetilde{f}_t^{s,e}(x)| \le \lambda \Big\};$$

$$\mathcal{B}_x(\tilde{r}, \rho, \lambda) = \Big\{ \max_{N=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}+1}^{\tilde{r}+N} F_{t,h}(x) - \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}+1}^{\tilde{r}+N} f_t(x) \Big| \le \lambda \Big\} \bigcup$$
$$\Big\{ \max_{N=\rho}^{\tilde{r}} \Big| \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}-N+1}^{\tilde{r}} F_{t,h}(x) - \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}-N+1}^{\tilde{r}} f_t(x) \Big| \le \lambda \Big\}.$$

From Algorithm 1, we have that

$$\rho = \frac{\log(T)}{nh^d}.$$

We observe that, $\rho n h^d = \log(T)$ and for $T \ge 3$, we have that

$$\rho^{1/2-1/q} \ge (nh^d)^{1/2-(q-1)/q}.$$

Therefore, Proposition 1 and Corollary 1 imply that with

$$\lambda = C_\lambda \Big(\log^{5/q}(T) \sqrt{\frac{1}{nh^d} + 1} + \sqrt{\frac{\log(T)}{nh^d}} + \sqrt{T} h^r + \sqrt{T} \Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}} \Big), \tag{4}$$

for some diverging sequence C_λ, it holds that

$$P\Big\{ \mathcal{A}_x^c((s,e], \rho, \lambda) \Big\} \le 4C_1 \frac{\log(T)}{(\log^{5/q}(T))^q} + \frac{2}{T^5} + \frac{10}{Tn}$$

and

$$P\Big\{ \mathcal{B}_x^c(r, \rho, \lambda) \Big\} \le 2C_1 \frac{\log(T)}{(\log^{5/q}(T))^q} + \frac{1}{T^5} + \frac{5}{Tn}.$$

Then, using that $\log^4(T) = O(T)$, from above

$$P\Big\{\mathcal{A}_x^c((s,e],\rho,\lambda)\Big\} = O(\log^{-4}(T)) \quad \text{and} \quad P\Big\{\mathcal{B}_x^c(r,\rho,\lambda)\Big\} = O(\log^{-4}(T)).$$

Now, we notice that,

$$\sum_{k=1}^{\mathcal{K}} \tilde{n}_k = \sum_{k=1}^{\mathcal{K}} (2^k - 1) \leq \sum_{k=1}^{\mathcal{K}} 2^k \leq 2(2^{\lceil \log(2) C_{\mathcal{K}} (\log(\log(T)))/\log 2 \rceil} - 1)$$
$$\leq 4(2^{(\log(\log(T)))/\log 2})^{\log(2) C_{\mathcal{K}}} = O(\log^{\log(2)C_{\mathcal{K}}}((T))).$$

In addition, there are $K = O(1)$ number of change-points. In consequence, it follows that

$$P\Big\{\mathcal{A}_u(\mathcal{I},\rho,\lambda) \text{ for all } \mathcal{I} \in \mathcal{J} \text{ and all } u \in \{u_m\}_{m=1}^{\log(T)}\Big\} \geq 1 - \frac{1}{\log^2(T)}, \tag{5}$$

$$P\Big\{\mathcal{B}_u(s,\rho,\lambda) \cup \mathcal{B}_u(e,\rho,\lambda) \text{ for all } (s,e] = \mathcal{I} \in \mathcal{J} \text{ and all } u \in \{u_m\}_{m=1}^{\log(T)}\Big\} \geq 1 - \frac{1}{\log(T)}, \tag{6}$$

$$P\Big\{\mathcal{B}_u(\eta_k,\rho,\lambda) \text{ for all } 1 \leq k \leq K \text{ and all } u \in \{u_m\}_{m=1}^{\log(T)}\Big\} \geq 1 - \frac{1}{\log^3(T)}. \tag{7}$$

The rest of the argument is made by assuming the events in equations (5), (6) and (7) hold.

Denote

$$\Upsilon_k = C \log^{\max\{1,10/q\}}(T)\Big(1+T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\Big)\kappa_k^{-2} \quad \text{and} \quad \Upsilon_{\max} = C \log^{\max\{1,10/q\}}(T)\Big(1+T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\Big)\kappa^{-2},$$

where $\kappa = \min\{\kappa_1, \ldots, \kappa_K\}$. Since Υ_k is the desired localisation rate, by induction, it suffices to consider any generic interval $(s,e] \subseteq (0,T]$ that satisfies the following three conditions:

$$\eta_{m-1} \leq s \leq \eta_m \leq \ldots \leq \eta_{m+q} \leq e \leq \eta_{m+q+1}, \quad q \geq -1;$$
$$\text{either } \eta_m - s \leq \Upsilon_m \quad \text{or} \quad s - \eta_{m-1} \leq \Upsilon_{m-1};$$
$$\text{either } \eta_{m+q+1} - e \leq \Upsilon_{m+q+1} \quad \text{or} \quad e - \eta_{m+q} \leq \Upsilon_{m+q}.$$

Here $q = -1$ indicates that there is no change-point contained in $(s,e]$.

Denote

$$\Delta_k = \eta_{k-1} - \eta_k \text{ for } k = 1, \ldots, K+1 \quad \text{and} \quad \Delta = \min\{\Delta_1, \ldots, \Delta_{K+1}\}.$$

Observe that since $\kappa_k > 0$ for all $1 \leq k \leq K$ and that $\Delta_k = \Theta(T)$, it holds that $\Upsilon_{\max} = o(\Delta)$. Therefore, it has to be the case that for any true change-point $\eta_m \in (0,T]$, either $|\eta_m - s| \leq \Upsilon_m$ or $|\eta_m - s| \geq \Delta - \Upsilon_{\max} \geq \Theta(T)$. This means that $\min\{|\eta_m - e|, |\eta_m - s|\} \leq \Upsilon_m$ indicates that η_m is a detected change-point in the previous induction step, even if $\eta_m \in (s,e]$. We refer to $\eta_m \in (s,e]$ as an undetected change-point if $\min\{\eta_m - s, \eta_m - e\} = \Theta(T)$. To complete the induction step, it suffices to show that FSBS $((s,e],h,\tau)$

(i) will not detect any new change point in $(s,e]$ if all the change-points in that interval have been previously detected, and

(ii) will find a point $D_{m*}^{\mathcal{I}^*}$ in $(s,e]$ such that $|\eta_m - D_{m*}^{\mathcal{I}^*}| \leq \Upsilon_m$ if there exists at least one undetected change-point in $(s,e]$.

In order to accomplish this, we need the following series of steps.

Step 1. We first observe that if $\eta_k \in \{\eta_k\}_{k=1}^K$ is any change-point in the functional time series, by Lemma 8, there exists a seeded interval $\mathcal{I}_k = (s_k, e_k]$ containing exactly one change-point η_k such that

$$\min\{\eta_k - s_k, e_k - \eta_k\} \geq \frac{1}{16}\zeta_k, \quad \text{and} \quad \max\{\eta_k - s_k, e_k - \eta_k\} \leq \zeta_k$$

where,

$$\zeta_k = \frac{9}{10}\min\{\eta_{k+1} - \eta_k, \eta_k - \eta_{k-1}\}.$$

Even more, we notice that if $\eta_k \in (s, e]$ is any undetected change-point in $(s, e]$. Then it must hold that

$$s - \eta_{k-1} \le \Upsilon_{\max}.$$

Since $\Upsilon_{\max} = O(\log^{\max\{1,10/q\}}(T) T^{\frac{d}{2r+d}})$ and $O(\log^a(T)) = o(T^b)$ for any positive numbers a and b, we have that $\Upsilon_{\max} = o(T)$. Moreover, $\eta_k - s_k \le \zeta_k \le \frac{9}{10}(\eta_k - \eta_{k-1})$, so that it holds that

$$s_k - \eta_{k-1} \ge \frac{1}{10}(\eta_k - \eta_{k-1}) > \Upsilon_{\max} \ge s - \eta_{k-1}$$

and in consequence $s_k \ge s$. Similarly $e_k \le e$. Therefore

$$\mathcal{I}_k = (s_k, e_k] \subseteq (s, e].$$

Step 2. Consider the collection of intervals $\{\mathcal{I}_k = (s_k, e_k]\}_{k=1}^K$ in **Step 1.** In this step, it is shown that for each $k \in \{1, \dots, K\}$, it holds that

$$\max_{t=s_k+\rho}^{t=e_k-\rho} \max_{m=1}^{m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \ge c_1 \sqrt{T} \kappa_k, \tag{8}$$

for some sufficient small constant c_1.

Let $k \in \{1, \dots, K\}$. By **Step 1**, $\mathcal{I}_k$ contains exactly one change-point η_k. Since for every u_m, $f_t^*(u_m)$ is a one dimensional population time series and there is only one change-point in $\mathcal{I}_k = (s_k, e_k]$, it holds that

$$f_{s_k+1}^*(u_m) = \dots = f_{\eta_k}^*(u_m) \ne f_{\eta_k+1}^*(u_m) = \dots = f_{e_k}^*(u_m)$$

which implies, for $s_k < t < \eta_k$

$$\begin{aligned}
\widetilde{f}_t^{(s_k,e_k]}(u_m) =& \sqrt{\frac{e_k - t}{(e_k - s_k)(t - s_k)}} \sum_{l=s_k+1}^{t} f_{\eta_k}^*(u_m) - \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} \sum_{l=t+1}^{\eta_k} f_{\eta_k}^*(u_m) \\
&- \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} \sum_{l=\eta_k+1}^{e_k} f_{\eta_k+1}^*(u_m) \\
=& (t - s_k) \sqrt{\frac{e_k - t}{(e_k - s_k)(t - s_k)}} f_{\eta_k}^*(u_m) - (\eta_k - t) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k}^*(u_m) \\
&- (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& \sqrt{\frac{(t - s_k)(e_k - t)}{(e_k - s_k)}} f_{\eta_k}^*(u_m) - (\eta_k - t) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k}^*(u_m) \\
&- (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& (e_k - t) \sqrt{\frac{t - s_k}{(e_k - t)(e_k - s_k)}} f_{\eta_k}^*(u_m) - (\eta_k - t) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k}^*(u_m) \\
&- (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - t)(e_k - s_k)}} f_{\eta_k}^*(u_m) - (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - t)(e_k - s_k)}} (f_{\eta_k}^*(u_m) - f_{\eta_k+1}^*(u_m)).
\end{aligned}$$

Similarly, for $\eta_k \le t \le e_k$

$$f_t^{(s_k,e_k]}(u_m) = \sqrt{\frac{e_k - t}{(e_k - s_k)(t - s_k)}}(\eta_k - s_k)(f^*_{\eta_k}(u_m) - f^*_{\eta_k+1}(u_m)).$$

Therefore,

$$\widetilde{f}_t^{(s_k,e_k]}(u_m) = \begin{cases} \sqrt{\frac{t-s_k}{(e_k-s_k)(e_k-t)}}(e_k - \eta_k)(f^*_{\eta_k}(u_m) - f^*_{\eta_k+1}(u_m)), & s_k < t < \eta_k; \\ \sqrt{\frac{e_k-t}{(e_k-s_k)(t-s_k)}}(\eta_k - s_k)(f^*_{\eta_k}(u_m) - f^*_{\eta_k+1}(u_m)), & \eta_k \le t \le e_k. \end{cases} \tag{9}$$

By Lemma 7, with probability at least $1 - o(1)$, there exists $u_{\tilde{k}} \in \{u_m\}_{m=1}^{\log(T)}$ such that

$$|f^*_{\eta_k}(u_{\tilde{k}}) - f^*_{\eta_k+1}(u_{\tilde{k}})| \ge \frac{3}{4}\kappa_k.$$

Since $\Delta = \Theta(T)$, $\rho = O(\log(T)T^{\frac{d}{2r+d}})$ and $\log^a(T) = o(T^b)$ for any positive numbers a and b, we have that

$$\min\{\eta_k - s_k, e_k - \eta_k\} \ge \frac{1}{16}\zeta_k \ge c_2 T > \rho, \tag{10}$$

so that $\eta_k \in [s_k + \rho, e_k - \rho]$. Then, from (9), (10) and the fact that $|e_k - s_k| < T$ and $|\eta_k - s_k| < T$,

$$|\widetilde{f}_{\eta_k}^{(s_k,e_k]}(u_{\tilde{k}})| = \sqrt{\frac{e_k - \eta_k}{(e_k - s_k)(\eta_k - s_k)}}(\eta_k - s_k)|f^*_{\eta_k}(u_{\tilde{k}}) - f^*_{\eta_k+1}(u_{\tilde{k}})| \ge c_2\sqrt{T}\frac{3}{4}\kappa_k. \tag{11}$$

Therefore, it holds that

$$\begin{aligned}\max_{t=s_k+\rho}^{t=e_k-\rho} \max_{m=1}^{m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \ge & |\widetilde{F}_{\eta_k,h}^{(s_k,e_k]}(u_{\tilde{k}})| \\ \ge & |\widetilde{f}_{\eta_k}^{(s_k,e_k]}(u_{\tilde{k}})| - \lambda \\ \ge & c_2\frac{3}{4}\sqrt{T}\kappa_k - \lambda,\end{aligned}$$

where the first inequality follows from the fact that $\eta_k \in [s_k + \rho, e_k - \rho]$, the second inequality follows from the good event in (5), and the last inequality follows from (11).

Next, we observe that $\log^{\frac{5}{q}}(T)\sqrt{\frac{1}{nh^d} + 1} = o(\sqrt{T^{\frac{2r+d}{d}}})O(\sqrt{T^{\frac{d}{2r+d}}}) = o(\sqrt{T})$, $\rho < c_2 T$, $h^r = o(1)$ and $\left(\frac{\log nT}{nT}\right)^{\frac{2r}{2r+d}} = o(1)$. In consequence, since κ_k is a positive constant, by the upper bound of λ on Equation (4), for sufficiently large T, it holds that

$$\frac{c_2}{4}\sqrt{T}\kappa_k \ge \lambda.$$

Therefore,

$$\max_{t=s_k+\rho}^{t=e_k-\rho} \max_{m=1}^{m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \ge \frac{c_2}{2}\sqrt{T}\kappa_k.$$

Therefore Equation (8) holds with $c_1 = \frac{c_2}{2}$.

Step 3. In this step, it is shown that FSBS$((s,e],h,\tau)$ can consistently detect or reject the existence of undetected change-points within $(s,e]$.

Suppose $\eta_k \in (s,e]$ is any undetected change-point. Then by the second half of **Step 1**, $\mathcal{I}_k \subseteq (s,e]$. Therefore

$$A_{m^*}^{\mathcal{I}^*} \ge \max_{t=s_k+\rho}^{t=e_k-\rho} \max_{m=1}^{m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \ge c_1\sqrt{T}\kappa_k > \tau,$$

where the second inequality follows from Equation (8), and the last inequality follows from the fact that, $\log^a(T) = o(T^b)$ for any positive numbers a and b implies $\tau = C_\tau\left(\log^{\max\{1,10/q\}}(T)\sqrt{\frac{1}{nh^d} + 1}\right) = o(\sqrt{T})$.

Suppose there does not exist any undetected change-point in $(s,e]$. Then for any $\mathcal{I} = (\alpha,\beta] \subseteq (s,e]$, one of the following situations must hold,

(a) There is no change-point within $(\alpha, \beta]$;

(b) there exists only one change-point η_k within $(\alpha, \beta]$ and $\min\{\eta_k - \alpha, \beta - \eta_k\} \leq \Upsilon_k$;

(c) there exist two change-points η_k, η_{k+1} within $(\alpha, \beta]$ and

$$\eta_k - \alpha \leq \Upsilon_k \quad \text{and} \quad \beta - \eta_{k+1} \leq \Upsilon_{k+1}.$$

The calculations of (c) are provided as the other two cases are similar and simpler. Note that for any $x \in [0,1]^d$, it holds that

$$|f^*_{\eta_{k+1}}(x) - f^*_{\eta_{k+1}+1}(x)| \leq \|f^*_{\eta_{k+1}} - f^*_{\eta_{k+1}+1}\|_\infty = \kappa_{k+1}$$

and similarly

$$|f^*_{\eta_k}(x) - f^*_{\eta_k+1}(x)| \leq \kappa_k.$$

By Lemma 10 and the assumption that $(\alpha, \beta]$ contains only two change-points, it holds that for all $x \in [0,1]^d$,

$$\begin{aligned}
\max_{t=\alpha}^{\beta} |\widetilde{f}_t^{(\alpha,\beta]}(x)| \leq & \sqrt{\beta - \eta_{r+1}}|f^*_{\eta_{r+1}}(x) - f^*_{\eta_{r+1}+1}(x)| + \sqrt{\eta_r - \alpha}|f^*_{\eta_r}(x) - f_{\eta_r^*+1}(x)| \\
\leq & \sqrt{\Upsilon_{k+1}}\kappa_{k+1} + \sqrt{\Upsilon_k}\kappa_k \leq 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}}.
\end{aligned}$$

Thus

$$\max_{t=\alpha}^{\beta} \|\widetilde{f}_t^{(\alpha,\beta]}\|_\infty \leq 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}}. \tag{12}$$

Therefore in the good event in Equation (5), for any $1 \leq m \leq \log(T)$ and any $\mathcal{I} = (\alpha, \beta] \subseteq (s, e]$, it holds that

$$\begin{aligned}
A_m^{\mathcal{I}} = & \max_{t=\alpha+\rho}^{\beta-\rho} |\widetilde{F}_{t,h}^{(\alpha,\beta]}(u_m)| \\
\leq & \max_{t=\alpha+\rho}^{\beta-\rho} \|\widetilde{f}_t^{(\alpha,\beta]}\|_\infty + \lambda \\
\leq & 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}} + \lambda,
\end{aligned}$$

where the first inequality follows from Equation (5), and the last inequality follows from Equation (12). Then,

$$\begin{aligned}
& 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}} + \lambda \\
= & 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{\frac{1}{nh^d} + 1} \\
& + C_\lambda \log^{5/q}(T)\sqrt{\frac{1}{nh^d} + 1} + C_\lambda\sqrt{\frac{\log(T)}{nh^d}} + C_\lambda\sqrt{T}h^r + C_\lambda\sqrt{T}\Big(\frac{\log nT}{nT}\Big)^{\frac{2r}{2r+d}}.
\end{aligned}$$

We observe that $\sqrt{\frac{\log(T)}{nh^d}} = O(\log(T)^{1/2}\sqrt{\frac{1}{nh^d} + 1})$. Moreover,

$$\sqrt{T}h^r = \sqrt{T}\Big(\frac{1}{nT}\Big)^{\frac{r}{2r+d}} \leq (T^{\frac{1}{2} - \frac{r}{2r+d}})\frac{1}{n^{\frac{r}{2r+d}}},$$

and given that,

$$\frac{1}{2} - \frac{r}{2r+d} = \frac{d}{2(2r+d)},$$

we get,

$$\sqrt{T}h^r = o\Big(\log^{\max 1/2,5/q}(T)\sqrt{\frac{1}{nh^d} + 1}\Big).$$

Following the same line of arguments, we have that

$$\sqrt{T}\Big(\frac{\log nT}{nT}\Big)^{\frac{2r}{2r+d}} = T^{\frac{1}{2} - \frac{2r}{2r+d}} \log^{\frac{2r}{2r+d}}(T) = o\Big(\log T\sqrt{\frac{1}{nh^d} + 1}\Big).$$

Thus, by the choice of τ, it holds that with sufficiently large constant C_τ,

$$A_m^{\mathcal{I}} \le \tau \quad \text{for all } 1 \le m \le \log(T) \quad \text{and all} \quad \mathcal{I} \subseteq (s,e]. \tag{13}$$

As a result, FSBS $((s,e],h,\tau)$ will correctly reject if $(s,e]$ contains no undetected change-points.

Step 4. Assume that there exists an undetected change-point $\eta_{\tilde{k}} \in (s,e]$ such that

$$\min\{\eta_{\tilde{k}} - s, \eta_{\tilde{k}} - e\} = \Theta(T).$$

Let m^* and $\mathcal{I}^*$ be defined as in FSBS $((s,e],h,\tau)$ with

$$\mathcal{I}^* = (\alpha^*, \beta^*].$$

To complete the induction, it suffices to show that, there exists a change-point $\eta_k \in (s,e]$ such that $\min\{\eta_k - s, \eta_k - e\} = \Theta(T)$ and $|D_{m^*}^{\mathcal{I}^*} - \eta_k| \le \Upsilon_k$.

Consider the uni-variate time series

$$F_{t,h}(u_{m*}) = \frac{1}{n}\sum_{i=1}^n y_{t,i} K_h(u_{m*} - x_{t,i}) \quad \text{and} \quad f_t^*(u_{m*}) \quad \text{for all } 1 \le t \le T.$$

Since the collection of the change-points of the time series $\{f_t^*(u_{m*})\}_{t\in\mathcal{I}^*}$ is a subset of that of $\{\eta_k\}_{k=0}^{K+1} \cap (s,e]$, we may apply Lemma 9 to by setting

$$\mu_t = F_{t,h}(u_{m*}) \quad \text{and} \quad \omega_t = f_t^*(u_{m*})$$

on the interval $\mathcal{I}^*$. Therefore, it suffices to justify that all the assumptions of Lemma 9 hold.

In the following, λ is used in Lemma 9. Then Equation (33) and Equation (34) are directly consequence of Equation (5), Equation (6), Equation (7).
We observe that, for any $\mathcal{I} = (\alpha,\beta] \subseteq (s,e]$,

$$\max_{t=\alpha^*+\rho}^{\beta^*-\rho} |\widetilde{F}_{t,h}^{(\alpha^*,\beta^*]}(u_{m^*})| = A_{m^*}^{\mathcal{I}^*} \ge A_m^{\mathcal{I}} = \max_{t=\alpha+\rho}^{\beta-\rho} |\widetilde{F}_{t,h}^{(\alpha,\beta]}(u_m)|$$

for all m. By **Step 1** with $\mathcal{I}_k = (s_k, e_k]$, it holds that

$$\min\{\eta_k - s_k, e_k - \eta_k\} \ge \frac{1}{16}\zeta_k \ge c_2 T,$$

Therefore for all $k \in \{\tilde{k} : \min\{\eta_{\tilde{k}} - s, e - \eta_{\tilde{k}}\} \ge c_2 T\}$,

$$\max_{t=\alpha^*+\rho}^{\beta^*-\rho} |\widetilde{F}_{t,h}^{(\alpha^*,\beta^*]}(u_{m^*})| \ge \max_{t=s_k+\rho, m=1}^{t=e_k-\rho, m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \ge c_1\sqrt{T}\kappa_k,$$

where the last inequality follows from Equation (8). Therefore Equation (35) holds in Lemma 9. Finally, Equation (36) is a direct consequence of the choices that

$$h = C_h(Tn)^{\frac{-1}{2r+d}} \quad \text{and} \quad \rho = \frac{\log(T)}{nh^d}.$$

Thus, all the conditions in Lemma 9 are met. So that, there exists a change-point η_k of $\{f_t^*(u_{m*})\}_{t\in\mathcal{I}^*}$, satisfying

$$\min\{\beta^* - \eta_k, \eta_k - \alpha^*\} > cT, \tag{14}$$

and

$$\begin{aligned} |D_{m*}^{\mathcal{I}^*} - \eta_k| \le \max\{C_3\lambda^2\kappa_k^{-2}, \rho\} &\le C_4 \log^{\max\{10/q,1\}}(T)\left(1 + \frac{1}{nh^d} + Th^{2r} + T\left(\frac{\log(nT)}{nT}\right)^{\frac{4r}{2r+d}}\right)\kappa_k^{-2} \\ &\le C\log^{\max\{10/q,1\}}(T)\left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)\kappa_k^{-2} \end{aligned}$$

for sufficiently large constant C, where we have followed the same line of arguments than for the conclusion of (13). Observe that
i) The change-points of $\{f_t^*(u_{m^*})\}_{t\in\mathcal{I}^*}$ belong to $(s,e] \cap \{\eta_k\}_{k=1}^K$; and
ii) Equation (14) and $(\alpha^*,\beta^*] \subseteq (s,e]$ imply that

$$\min\{e - \eta_k, \eta_k - s\} > cT \ge \Upsilon_{\max}.$$

As discussed in the argument before **Step 1**, this implies that η_k must be an undetected change-point of $\{f_t^*(u_{m^*})\}_{t\in\mathcal{I}^*}$. $\square$

C Deviation bounds related to kernels

In this section, we deal with all the large probability events occurred in the proof of Theorem 1. Recall that $F_{t,h}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} y_{t,i} K_h(x - x_{t,i})}{\hat{p}(x)}$, and

$$\widetilde{F}_{t,h}^{(s,e]}(x) = \sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{l=s+1}^{t} F_{l,h}(x) - \sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{l=t+1}^{e} F_{l,h}(x).$$

By assumption 2, we have $\max_{l=1}^{q} \|K^l\|_\infty = \max_{l=1}^{q} \|K\|_\infty^l < C_K$, where $C_K > 0$ is an absolute constant. Moreover, assumption 1b implies $|f_t^*(x)| < C_f$ for any $x \in [0,1]^d, t \in 1, ..., T$.

Proposition 1. *Suppose that Assumption 1 and 2 hold, that $\rho n h^d \geq \log(T)$ and that $T \geq 3$. Then for any $x \in [0,1]^d$*

$$\mathbb{P}\Big(\max_{k=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \big(F_{t,h}(x) - f_t^*(x)\big)\Big| \geq \frac{2}{\tilde{c}} z \sqrt{\frac{1}{nh^d} + 1} + \frac{\tilde{C}_1}{\tilde{c}} \Big(\sqrt{\frac{\log(T)}{nh^d}}\Big) + \frac{\tilde{C}}{\tilde{c}} \sqrt{T} h^r + \frac{\bar{C} C_f}{\tilde{c}} \sqrt{T} \Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big)$$
$$\leq 2C_1 \frac{\log(T)}{z^q} + T^{-5} + \frac{5}{Tn}; \tag{15}$$

$$\mathbb{P}\Big(\max_{k=\rho}^{\tilde{r}} \Big| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}-k+1}^{\tilde{r}} \big(F_{t,h}(x) - f_t^*(x)\big)\Big| \geq \frac{2}{\tilde{c}} z \sqrt{\frac{1}{nh^d} + 1} + \frac{\tilde{C}_1}{\tilde{c}} \Big(\sqrt{\frac{\log(T)}{nh^d}}\Big) + \frac{\tilde{C}}{\tilde{c}} \sqrt{T} h^r + \frac{\bar{C} C_f}{\tilde{c}} \sqrt{T} \Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big)$$
$$\leq 2C_1 \frac{\log(T)}{z^q} + T^{-5} + \frac{5}{Tn}. \tag{16}$$

Proof. The proofs of Equation (15) and Equation (16) are the same. So only the proof of Equation (15) is presented. We define the events $E_1 = \Big\{||\hat{p} - u||_\infty \leq \bar{C}\Big(\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}\Big)\Big\}$ and $E_2 = \Big\{\hat{p} \geq \tilde{c},\ \tilde{c} = \inf_x u(x) - \bar{C}\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}\Big\}$. Using Lemma 1, especifically by equation (3), we have that $P(E_1) \geq 1 - \frac{1}{nT}$. Then, we observe that in event E_1, for $x \in [0,1]^d$

$$\inf_s u(s) - \hat{p}(x) \leq u(x) - \hat{p}(x) \leq |u(x) - \hat{p}(x)| \leq \bar{C}\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}$$

which implies $E1 \subseteq E_2$. Therefore, $P(E_2^c) \leq \frac{1}{nT}$.
Now, for any x, observe that, by definition of $F_{t,h}$ and triangle inequality

$$\begin{aligned}
I &= \max_{k=\rho}^{T-\tilde{r}} \frac{1}{\sqrt{k}} \Big| \sum_{t=\tilde{r}+1}^{\tilde{r}+k} F_{t,h}(x) - \sum_{t=\tilde{r}+1}^{\tilde{r}+k} f_t^*(x) \Big| \\
&\leq \max_{k=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \Big(\frac{f_t^*(x_{t,i}) K_h(x - x_{t,i})}{\hat{p}(x)} - f_t^*(x) \Big) \Big| \\
&+ \max_{k=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \frac{\xi_t(x_{t,i}) K_h(x - x_{t,i})}{\hat{p}(x)} \Big| \\
&+ \max_{k=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{t,i} K_h(x - x_{t,i})}{\hat{p}(x)} \Big| \\
&= I_1 + I_2 + I_3.
\end{aligned} \tag{17}$$

In the following, we will show that $I_1 \leq I_{1,1} + I_{1,2} + I_{1,3}$, and that

1. $\mathbb{P}\Big(I_{1,1} \geq \frac{\tilde{C}_1}{\tilde{c}}\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big)\Big) \leq \frac{1}{T^5} + \frac{1}{Tn}$,
2. $\mathbb{P}\Big(I_{1,2} \geq \frac{\tilde{C}}{\tilde{c}} \sqrt{T} h^r\Big) \leq \frac{1}{Tn}$,

3. $\mathbb{P}\Big(I_{1,3} \geq \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \leq \frac{1}{Tn}$,

4. $\mathbb{P}\Big(I_2 \geq \frac{1}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}\Big) \leq \frac{C_1 \log T}{z^q} + \frac{1}{Tn}$,

5. $\mathbb{P}\Big(I_3 \geq \frac{1}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}\Big) \leq \frac{C_1 \log T}{z^q} + \frac{1}{Tn}$,

in order to conclude that,

$$
\begin{aligned}
&\mathbb{P}\Big(I \geq 2z\sqrt{\frac{1}{nh^d}+1} + \tilde{C}_1\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big) + \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r + \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \\
\leq&\mathbb{P}\Big(I_{1,1} \geq \tilde{C}_1\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big)\Big) + \mathbb{P}\Big(I_{1,2} \geq \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r\Big) + \mathbb{P}\Big(I_{1,3} \geq \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \\
+&\mathbb{P}\Big(I_2 \geq z\sqrt{\frac{1}{nh^d}+1}\Big) + \mathbb{P}\Big(I_3 \geq z\sqrt{\frac{1}{nh^d}+1}\Big) \\
\leq&2C_1\frac{\log(T)}{z^q} + T^{-5} + \frac{5}{Tn}.
\end{aligned}
$$

Step 1. The analysis for I_1 is done. We observe that,

$$
\begin{aligned}
&\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - f_t^*(x)\Big)\Big| \\
\leq&\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - \frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)}\Big)\Big| \\
+&\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)} - f_t^*(x)\Big)\Big| = I_{1,1} + \tilde{I}_1.
\end{aligned}
$$

Step 1.1 The analysis for $I_{1,1}$ is done. We note that the random variables $\{f_t^*(x_{t,i})K_h(x - x_{t,i})\}_{1\leq i\leq n_t, 1\leq t\leq N}$ are independent distributed with mean $\int f_t^*(z)K_h(x-z)d\mu(z)$ and

$$
\begin{aligned}
\mathrm{Var}\big(f_t^*(x_{t,i})K_h(x-x_{t,i})\big) \leq& E\big\{(f_t^*)^2(x_{t,i})K_h^2(x-x_{t,i})\big\} \\
=&\int_{[0,1]^d}(f_t^*)^2(z)\frac{1}{h^{2d}}K^2\big(\frac{x-z}{h}\big)d\mu(z) \\
\leq&\frac{C_f^2}{h^d}\int_{[0,1]^d}\frac{1}{h^d}K^2\big(\frac{x-z}{h}\big)d\mu(z) \\
=&\frac{C_f^2}{h^d}\int_{[0,1]^d}K^2\big(u\big)d\mu(u) < \frac{C_f^2C_K^2}{h^d}.
\end{aligned}
$$

Since $|f_t^*(x_{t,i})K_h(x-x_{t,i})| \leq C_fC_Kh^{-d}$, by Bernstein inequality [11], we have that

$$
\mathbb{P}\Big(\Big|\frac{1}{kn}\sum_{t=r+1}^{r+k}\sum_{i=1}^{n}f_t^*(x_{t,i})K_h(x-x_{t,i}) - \int f_t^*(z)K_h(x-z)d\mu(z)\Big| \geq \tilde{C}_1\Big\{\sqrt{\frac{\log(T)}{knh^d}} + \frac{\log(T)}{knh^d}\Big\}\Big) \leq T^{-6}.
$$

Since $knh^d \geq \log(T)$ if $k \geq \rho$, with probability at most T^{-5}, it holds that

$$
\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}n}\sum_{t=r+1}^{r+k}\sum_{i=1}^{n}\Big(f_t^*(x_{t,i})K_h(x-x_{t,i}) - \int f_t^*(z)K_h(x-z)d\mu(z)\Big)\Big| \geq \tilde{C}_1\sqrt{\frac{\log(T)}{nh^d}}.
$$

Therefore, using that $P(E_2^c) \leq \frac{1}{Tn}$, we conclude

$$
\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - \frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)}\Big)\Big| \geq \frac{\tilde{C}_1}{\tilde{c}}\sqrt{\frac{\log(T)}{nh^d}}
$$

with probability at most $T^{-5}+\frac{1}{nT}$.

Step 1.2 The analysis for $I_{1,2}$ and $I_{1,3}$ is done. We observe that

$$\tilde{I}_1 = \max_{k=\rho}^{T-\tilde{r}} \left| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\int f_t^*(z) K_h(x-z) d\mu(z)}{\hat{p}(x)} - f_t^*(x) \right) \right|$$

$$\leq \max_{k=\rho}^{T-\tilde{r}} \left| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\int f_t^*(z) K_h(x-z) d\mu(z)}{\hat{p}(x)} - \frac{f_t^*(x) u(x)}{\hat{p}(x)} \right) \right| \tag{18}$$

$$+ \max_{k=\rho}^{T-\tilde{r}} \left| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{f_t^*(x) u(x)}{\hat{p}(x)} - f_t^*(x) \right) \right| = I_{1,2} + I_{1,3}. \tag{19}$$

Then, we observe that

$$\begin{aligned}
I_{1,2} =& \left| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \left(\int f_t^*(z) K_h(x-z) d\mu(z) - f_t^*(x) u(x) \right) \right| \\
\leq& \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \left| \int f_t^*(z) K_h(x-z) d\mu(z) - f_t^*(x) u(x) \right| \\
\leq& \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \tilde{C} h^r \\
=& \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \tilde{C} h^r \\
=& \sqrt{k} \tilde{C} h^r
\end{aligned}$$

where the second inequality follows from assumption 2. Therefore, using event E_2, we can bound (18) by $\frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r$ with probability at least $1-\frac{1}{nT}$. Meanwhile, for (19) we have that,

$$I_{1,3} = \max_{k=\rho}^{T-\tilde{r}} \left| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{f_t^*(x) u(x)}{\hat{p}(x)} - f_t^*(x) \right) \right|$$

$$\leq \max_{k=\rho}^{T-\tilde{r}} \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} |f_t^*(x)| \left| \frac{u(x) - \hat{p}(x)}{\hat{p}(x)} \right|. \tag{20}$$

Then, since in the event E_1, it is satisfies that

$$||\hat{p} - u||_\infty \leq \bar{C}\left(\left(\frac{\log(Tn)}{Tn} \right)^{\frac{2r}{2r+d}} \right), \text{ and } \hat{p} \geq \bar{c};$$

we have that equation (20), is bounded by

$$\max_{k=\rho}^{T-\tilde{r}} \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{1}{n} \sum_{i=1}^{n} \frac{\bar{C} C_f}{\tilde{c}} \left(\frac{\log(nT)}{nT} \right)^{\frac{2r}{2r+d}} \leq \frac{\bar{C} C_f}{\tilde{c}} \sqrt{T} \left(\frac{\log(nT)}{nT} \right)^{\frac{2r}{2r+d}}$$

with probability at least $1-\frac{1}{nT}$.

Step 2. The analysis for I_2 and I_3 is done. For $1 \leq t \leq T$, let

$$Z_t = \frac{1}{n} \sum_{i=1}^{n} \xi_t(x_{t,i}) K_h(x - x_{t,i}) \quad \text{and} \quad W_t = \frac{1}{n} \sum_{i=1}^{n} \delta_{t,i} K_h(x - x_{t,i}).$$

By Lemma 2 and event E_2, it holds that

$$\mathbb{P}\left\{ \max_{k=\rho}^{T-\tilde{r}} \left| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{Z_t}{\hat{p}(x)} \right| \geq \frac{1}{\tilde{c}} z \sqrt{\frac{1}{nh^d} + 1} \right\} \leq \frac{C_1 \log(T)}{z^q} + \frac{1}{nT}$$

and

$$\mathbb{P}\Big\{ \max_{k=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{k}} \sum_{t=\tilde{r}+1}^{\tilde{r}+k} \frac{W_t}{\hat{p}(x)} \Big| \geq \frac{1}{\tilde{c}} z \sqrt{\frac{1}{nh^d} + 1} \Big\} \leq \frac{C_1 \log(T)}{z^q} + \frac{1}{nT}.$$

The desired result follows from putting the previous steps together. □

Corollary 1. *Suppose that $\rho n h^d \geq \log(T)$ and that $T \geq 3$. Then for $z > 0$*

$$\mathbb{P}\Big\{ \max_{t=s+\rho+1}^{e-\rho} \Big| \widetilde{F}_{t,h}^{(s,e]}(x) - \widetilde{f}_t^{(s,e]}(x) \Big| \geq \frac{4}{\tilde{c}} z \sqrt{\frac{1}{nh^d} + 1} + \frac{2\tilde{C}_1}{\tilde{c}} \Big(\sqrt{\frac{\log(T)}{nh^d}} \Big) + \frac{2\tilde{C}}{\tilde{c}} \sqrt{T} h^r + \frac{2\bar{C} C_f}{\tilde{c}} \sqrt{T} \Big(\frac{\log(nT)}{nT} \Big)^{\frac{2r}{2r+d}} \Big\}$$
$$\leq 2T^{-5} + \frac{4C_1 \log(T)}{z^q} + 10 \frac{1}{Tn}.$$

Proof. By definition of $\widetilde{F}_{t,h}^{(s,e]}$ and $\widetilde{f}_t^{(s,e]}$, we have that

$$\begin{aligned} \Big| \widetilde{F}_{t,h}^{(s,e]}(x) - \widetilde{f}_t^{(s,e]}(x) \Big| \leq & \Big| \sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{l=s+1}^{t} (F_{l,h}(x) - f_l^*(x)) \Big| \\ & + \Big| \sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{l=t+1}^{e} (F_{l,h}(x) - f_l^*(x)) \Big|. \end{aligned}$$

Then, we observe that,

$$\sqrt{\frac{e-t}{(e-s)(t-s)}} \leq \sqrt{\frac{1}{t-s}} \text{ if } s \leq t, \text{ and } \sqrt{\frac{t-s}{(e-s)(e-t)}} \leq \sqrt{\frac{1}{e-t}} \text{ if } t \leq e.$$

Therefore,

$$\begin{aligned} X = \max_{t=s+\rho+1}^{e-\rho} \Big| \widetilde{F}_{t,h}^{(s,e]}(x) - \widetilde{f}_t^{(s,e]}(x) \Big| \leq & \max_{t=s+\rho+1}^{e-\rho} \Big| \sqrt{\frac{1}{t-s}} \sum_{l=s+1}^{t} \Big(F_{l,h}(x) - \{f_l^*(x) \Big) \Big| \\ & + \max_{t=s+\rho+1}^{e-\rho} \Big| \sqrt{\frac{1}{e-t}} \sum_{l=t+1}^{e} \Big(F_{l,h}(x) - f_l^*(x) \Big) \Big| = X_1 + X_2. \end{aligned}$$

Finally, letting $\lambda = \frac{4}{\tilde{c}} z \sqrt{\frac{1}{nh^d} + 1} + \frac{2\tilde{C}_1}{\tilde{c}} \Big(\sqrt{\frac{\log(T)}{nh^d}} \Big) + \frac{2\tilde{C}}{\tilde{c}} \sqrt{T} h^r + \frac{2\bar{C} C_f}{\tilde{c}} \sqrt{T} \Big(\frac{\log(nT)}{nT} \Big)^{\frac{2r}{2r+d}}$, we get that

$$\begin{aligned} \mathbb{P}(X \geq \lambda) \leq & \mathbb{P}(X_1 + X_2 \geq \frac{\lambda}{2} + \frac{\lambda}{2}) \\ \leq & \mathbb{P}(X_1 \geq \frac{\lambda}{2}) + \mathbb{P}(X_2 \geq \frac{\lambda}{2}) \\ \leq & 2T^{-5} + \frac{4C_1 \log(T)}{z^q} + 10 \frac{1}{Tn}, \end{aligned}$$

where the last inequality follows from Proposition 1. □

C.1 Additional Technical Results

The following lemmas provide lower bounds for

$$Z_t = \frac{1}{n}\sum_{i=1}^{n} \xi_t(x_{t,i}) K_h(x - x_{t,i}) \quad \text{and} \quad W_t = \frac{1}{n}\sum_{i=1}^{n} \delta_{t,i} K_h(x - x_{t,i}).$$

They are a direct consequence of the temporal dependence and heavy-tailedness of the data considered in Assumption 1.

Lemma 2. *Let $\rho \le T$ be such that $\rho n h^d \ge \log(T)$ and $T \ge 3$. Let $N \in \mathbb{Z}^+$ be such that $N \ge \rho$.*
a. *Suppose that for any $q \ge 3$ it holds that*

$$\sum_{t=1}^{\infty} t^{1/2-1/q} \mathbb{E}\big\{\|\xi_t - \xi_t^*\|_\infty^q\big\}^{1/q} = \mathrm{O}(1). \tag{21}$$

Then for any $z > 0$,

$$\mathbb{P}\bigg\{ \max_{k=\rho}^{N} \bigg| \bigg\{ \frac{1}{nh^d} + 1 \bigg\}^{-1/2} \frac{1}{\sqrt{k}} \sum_{t=1}^{k} Z_t \bigg| \ge z \bigg\} \le \frac{C_1 \log(T)}{z^q}.$$

b. *Suppose that for some $q \ge 3$,*

$$\sum_{t=1}^{\infty} t^{1/2-1/q} \max_{i=1}^{n} \big\{\mathbb{E}|\delta_{t,i} - \delta_{t,i}^*|^q\big\}^{1/q} < \mathrm{O}(1). \tag{22}$$

Then for any $w > 0$,

$$\mathbb{P}\bigg\{ \max_{k=\rho}^{N} \bigg| \bigg\{ \frac{1}{nh^d} + 1 \bigg\}^{-1/2} \frac{1}{\sqrt{k}} \sum_{t=1}^{k} W_t \bigg| \ge w \bigg\} \le \frac{C_1 \log(T)}{w^q}.$$

Proof. The proof of part **b** is similar and simpler than that of part **a**. For conciseness, only the proof of **a** is presented.

By Lemma 4 and Equation (21), for all $J \in \mathbb{Z}^+$, it holds that

$$\mathbb{E}\bigg\{ \max_{k=1}^{J} |\sum_{t=1}^{k} Z_t|^q \bigg\}^{1/q} \le J^{1/2} C\bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{1/2} + 1 \bigg\} + J^{1/q} C''\bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{(q-1)/q} + 1 \bigg\}.$$

As a result there exists a constant C_1 such that

$$\mathbb{E}\bigg\{ \max_{k=1}^{J} |\sum_{t=1}^{k} Z_t|^q \bigg\} \le C_1 J^{q/2} \bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{1/2} + 1 \bigg\}^q + C_1 J \bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{(q-1)/q} + 1 \bigg\}^q.$$

We observe that

$$\begin{aligned}
J^{q/2} &= \frac{q}{2}\int_0^J x^{q/2-1} dx && (23)\\
&= \frac{q}{2}\Big(\int_0^1 x^{q/2-1} dx + \int_1^J x^{q/2-1} dx \Big) && (24)\\
&\le \frac{q}{2}\Big(1 + \int_1^J x^{q/2-1} dx \Big) && (25)\\
&= \frac{q}{2}\Big(1 + \int_1^2 x^{q/2-1} dx + \dots + \int_{J-1}^J x^{q/2-1} dx \Big) && (26)\\
&\le \frac{q}{2}\Big(1 + \int_1^2 2^{q/2-1} dx + \dots + \int_{J-1}^J J^{q/2-1} dx \Big) && (27)\\
&= \frac{q}{2}\sum_{k=1}^{J} k^{q/2-1}, && (28)
\end{aligned}$$

which implies, there is a constant C_2 such that

$$C_1 J^{q/2}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q + C_1 J\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}^q \le C_2\sum_{k=1}^{J}\alpha_k,$$

where

$$\alpha_k = k^{q/2-1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q + \Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}^q.$$

By theorem B.2 of Kirch (2006),

$$\begin{aligned}
\mathbb{E}\Big\{\max_{k=1}^{N}\Big|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}Z_t\Big|\Big\}^q \le& 4C_2\sum_{l=1}^{N} l^{-q/2}\alpha_l\\
=&4C_2\sum_{l=1}^{N}\Big(l^{-1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q + l^{-q/2}\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}^q\Big)\\
\le& C_3\log(N)\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q + C_3 N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}^q
\end{aligned}$$

where the last inequality follows from the fact that $\int_1^N \frac{1}{x} = log(N)$ and that $\int_1^N x^{\frac{-q}{2}} = O(N^{-q/2+1})$. Since

$$N^{1/2-1/q} \ge \rho^{1/2-1/q} \ge (nh^d)^{1/2-(q-1)/q},$$

it holds that, $\frac{1}{nh^d} \le N$. Moreover,

$$\begin{aligned}
N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}^q =& N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q+1/2-1/2}+1\Big\}^q\\
=& N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q-1/2}\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\\
\le& N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q-1/2}+1\Big\}^q\\
=& N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2-1/q}+1\Big\}^q\\
=& N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-2)/(2q)}+1\Big\}^q\\
\le& C_3' N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-2)/(2q)}\Big\}^q\\
=& C_3' N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-2)/(2)}\Big\}\\
\le& C_3' N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q\Big\{\Big(\frac{1}{nh^d}\Big)^{q/2-1}\Big\}\\
\le& C_3' N^{-q/2+1}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q N^{q/2-1}\\
=& C_3'\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q.
\end{aligned}$$

It follows that,

$$\mathbb{E}\Big\{\max_{k=1}^{N}\Big|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}Z_t\Big|\Big\}^q \le C_4\log(N)\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q.$$

By Markov's inequality, for any $z > 0$ and the assumption that $T \geq N$,

$$\mathbb{P}\bigg\{ \max_{k=1}^{N} \bigg\{ \frac{1}{nh^d} + 1 \bigg\}^{-1/2} \bigg| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} Z_t \bigg| \geq z \bigg\} \leq \frac{C_1 \log(T)}{z^q}.$$

Since $N \geq \rho$, this directly implies that

$$\mathbb{P}\bigg\{ \max_{k=\rho}^{N} \bigg\{ \frac{1}{nh^d} + 1 \bigg\}^{-1/2} \bigg| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} Z_t \bigg| \geq z \bigg\} \leq \frac{C_1 \log(T)}{z^q}.$$

□

Lemma 3. *Suppose Assumption 1* **c** *holds and $q \geq 2$. Then there exists absolute constants $C > 0$ so that*

$$\mathbb{E}|Z_t - Z_t^*|^q \leq C\mathbb{E}\big\{\|\xi_t - \xi_t^*\|_\infty^q\big\}\bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{q-1} + 1 \bigg\}. \tag{29}$$

If in addition $\mathbb{E}\big\{\|\xi_t\|_\infty^q\big\} = O(1)$, then there exists absolute constants C' such that

$$\mathbb{E}|Z_t|^q \leq C'\bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{q-1} + 1 \bigg\}. \tag{30}$$

Proof. The proof of the Equation (30) is simpler and simpler than Equation (29). So only the proof of Equation (29) is presented. Note that since $\{x_t\}_{t=1}^T$ and $\{\xi_t\}_{t=1}^T$ are independent, and that $\{x_t\}_{t=1}^T$ are independent identically distributed,

$$Z_t^* = \frac{1}{n} \sum_{i=1}^{n} \xi_t^*(x_{t,i}) K_h(x - x_{t,i}).$$

Step 1. Note that, by the Newton's binomial

$$\begin{aligned}
\mathbb{E}|Z_t - Z_t^*|^q =& \mathbb{E}\bigg\{ \bigg| \frac{1}{n} \sum_{i=1}^{n} \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \bigg|^q \bigg\} \\
\leq& \frac{1}{n^q} \mathbb{E}\bigg\{ \sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta_1 \geq 0, \ldots, \beta_n \geq 0}} \binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \prod_{j=1}^{n} \big| \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\} \\
=& \frac{1}{n^q} \mathbb{E}\bigg\{ \sum_{k=1}^{q} \sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta=(\beta_1,\ldots,\beta_n), \|\beta\|_0=k, \beta\geq 0}} \binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \prod_{j=1}^{n} \big| \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\}.
\end{aligned}$$

Step 2. For a fixed $\beta = (\beta_1, \ldots, \beta_n)$ such that $\beta_1 + \ldots + \beta_n = q$ and that $\|\beta\|_0 = k$, consider

$$\mathbb{E}\bigg\{ \prod_{j=1}^{n} \big| \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\}.$$

Without loss of generality, assume that $\beta_1, \ldots, \beta_k$ are non-zero. Then it holds that

$$\begin{aligned}
& \mathbb{E}\bigg\{ \big|(\xi_t^* - \xi_t)(x_{t,1})\big|^{\beta_1} \big|K_h(x - x_{t,1})\big|^{\beta_1} \cdots \big|(\xi_t^* - \xi_t)(x_{t,k})\big|^{\beta_k} \big|K_h(x - x_{t,k})\big|^{\beta_k} \bigg\} \\
=& \mathbb{E}_\xi\bigg\{ \int \big|(\xi_t^* - \xi_t)(r)\big|^{\beta_1} \big|K_h(x - r)\big|^{\beta_1} d\mu(r) \cdots \int \big|(\xi_t^* - \xi_t)(r)\big|^{\beta_k} \big|K_h(x - r)\big|^{\beta_k} d\mu(r) \bigg\} \\
=& \mathbb{E}_\xi\bigg\{ \int \big|(\xi_t^* - \xi_t)(x - sh)\big|^{\beta_1} \frac{\big|K(s)\big|^{\beta_1}}{h^{d(\beta_1 - 1)}} d\mu(s) \cdots \int \big|(\xi_t^* - \xi_t)(x - sh)\big|^{\beta_k} \frac{\big|K(s)\big|^{\beta_k}}{h^{d(\beta_k - 1)}} d\mu(s) \bigg\} \\
\leq& h^{-d\sum_{j=1}^{k}(\beta_j - 1)} \mathbb{E}_\xi\bigg\{ \|\xi_t^* - \xi_t\|_\infty^{\beta_1} C_K^{\beta_1} \cdots \|\xi_t^* - \xi_t\|_\infty^{\beta_k} C_K^{\beta_k} \bigg\} \\
\leq& h^{-d(q-k)} C_K^q \mathbb{E}_\xi\bigg\{ \|\xi_t^* - \xi_t\|_\infty^{\sum_{j=1}^{k} \beta_k} \bigg\} \\
\leq& h^{-d(q-k)} C_K^q \mathbb{E}_\xi\big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\}
\end{aligned}$$

where the third equality follows by using the change of variable $s = \frac{x-r}{h}$, the first inequality by assumption 2.

Step 3. Let $k \in \{1, \ldots, q\}$ be fixed. Note that $\binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \le q!$. Consider set

$$\mathcal{B}_k = \bigg\{ \beta \in \mathbb{N}^n : \beta \ge 0, \beta_1 + \ldots + \beta_n = q, |\beta|_0 = k \bigg\}.$$

To bound the cardinality of the set $\mathcal{B}_k$, first note that since $|\beta|_0 = k$, there are $\binom{n}{k}$ number of ways to choose the index of non-zero entries of β.
Suppose $\{i_1, \ldots i_k\}$ are the chosen index such that $\beta_{i_1} \ne 0, \ldots, \beta_{i_k} \ne 0$. Then the constrains $\beta_{i_1} > 0, \ldots, \beta_{i_1} > 0$ and $\beta_{i_1} + \ldots + \beta_{i_k} = q$ are equivalent to that of diving q balls into k groups (without distinguishing each ball). As a result there are $\binom{q-1}{k-1}$ number of ways to choose the $\{\beta_{i_1}, \ldots, \beta_{i_k}\}$ once the index $\{i_1, \ldots i_k\}$ are chosen.

Step 4. Combining the previous three steps, it follows that for some constants $C_q, C_1 > 0$ only depending on q,

$$\begin{aligned}
\mathbb{E}|Z_t - Z_t^*|^q \le & \frac{1}{n^q} \mathbb{E} \bigg\{ \sum_{k=1}^q \sum_{\substack{\beta_1 + \beta_2 + \ldots + \beta_n = q \\ \beta = (\beta_1, \ldots, \beta_n), |\beta|_0 = k, \beta \ge 0}} \binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \prod_{j=1}^n \big| (\xi_t^* - \xi_t)(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\} \\
\le & \frac{1}{n^q} \sum_{k=1}^q \binom{n}{k} \binom{q-1}{k-1} q! h^{-d(q-k)} C_K^q \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} \\
\le & \frac{1}{n^q} \sum_{k=1}^q n^k C_q C_K^q h^{-d(q-k)} \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} \\
\le & C_1 \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} \bigg\{ \bigg(\frac{1}{nh^d} \bigg)^{q-1} + \bigg(\frac{1}{nh^d} \bigg)^{q-2} + \ldots + \bigg(\frac{1}{nh^d} \bigg) + 1 \bigg\} \\
\le & C_1 \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} q \bigg\{ \bigg(\frac{1}{nh^d} \bigg)^{q-1} + 1 \bigg\},
\end{aligned}$$

where the second inequality is satisfied by step 3 and that $\binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \le q!$, while the third inequality is achieved by using that $\binom{n}{k}\binom{q-1}{k-1} q! \le \binom{n}{k} C_q \le n^k C_q$. Moreover, given that $\frac{1}{n^q} n^k h^{-d(q-k)} = \left(\frac{1}{nh^d}\right)^{q-k}$ the fourth inequality is obtained. The last inequality holds because if $\frac{1}{nh^d} \le 1$, then $\left\{ \left(\frac{1}{nh^d}\right)^{q-1} + \ldots + \left(\frac{1}{nh^d}\right) + 1 \right\} \le q$, and if $\frac{1}{nh^d} \ge 1$, then $\left\{ \left(\frac{1}{nh^d}\right)^{q-1} + \ldots + \left(\frac{1}{nh^d}\right) + 1 \right\} \le q \left(\frac{1}{nh^d}\right)^{q-1}$. □

Lemma 4. *Suppose Assumption 1* **c** *holds. Let $\rho \le T$ be such that $\rho n h^d \ge \log(T)$ and $T \ge 3$. Let $N \in \mathbb{Z}^+$ be such that $N \ge \rho$. Then, it holds that*

$$\bigg\{ \mathbb{E} \max_{k=1}^N \Big| \sum_{t=1}^k Z_t \Big|^q \bigg\}^{1/q} \le N^{1/2} C \bigg\{ \bigg(\frac{1}{nh^d} \bigg)^{1/2} + 1 \bigg\} + N^{1/q} C' \bigg\{ \bigg(\frac{1}{nh^d} \bigg)^{(q-1)/q} + 1 \bigg\}.$$

Proof. We have that $q > 2$ and $\mathbb{E}|Z_1| < \infty$ by the use of Lemma 3. Then, making use of Theorem 1 of Liu et al. (2013), we obtain that

$$\begin{aligned}
\bigg\{ \mathbb{E} \max_{k=1}^N \Big| \sum_{t=1}^k Z_t \Big|^q \bigg\}^{1/q} \le & N^{1/2} C_1 \bigg\{ \sum_{j=1}^N \Theta_{j,2} + \sum_{j=N+1}^\infty \Theta_{j,q} + \{\mathbb{E}|Z_1|^2\}^{1/2} \bigg\} \\
& + N^{1/q} C_2 \bigg\{ \sum_{j=1}^N j^{1/2 - 1/q} \Theta_{j,q} + \{\mathbb{E}|Z_1|^q\}^{1/q} \bigg\},
\end{aligned}$$

where $\Theta_{j,q} = \{\mathbb{E}(|Z_j^* - Z_j|^q)\}^{1/q}$. Moreover, we observe that since $\Theta_{j,2} \leq \Theta_{j,q}$ for any $q \geq 2$, it follows

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1\Big\{\sum_{j=1}^{\infty}\Theta_{j,q} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}$$
$$+N^{1/q}C_2\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}\Theta_{j,q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\},$$

Next, by the first part of Lemma 3,

$$\Theta_{j,q}^q \leq C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}.$$

even more, we have that $N \geq \frac{1}{nh^d}$, implies that

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{'}\Big\{\sum_{j=1}^{\infty} C\mathbb{E}\{\|\xi_j - \xi_j^*\|_\infty^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}\Big\}^{1/q} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}$$
$$+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}C\mathbb{E}\{\|\xi_j - \xi_j^*\|_\infty^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}$$
$$\leq N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty} C\mathbb{E}\{\|\xi_j - \xi_j^*\|_\infty^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}$$
$$+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}C\mathbb{E}\{\|\xi_j - \xi_j^*\|_\infty^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}$$
$$\leq N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty} C\mathbb{E}\{\|\xi_j - \xi_j^*\|_\infty^q\}\Big\{\Big(N\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}$$
$$+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}C\mathbb{E}\{\|\xi_j - \xi_j^*\|_\infty^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}.$$

From Assumption 1 **c**,

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{'''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}$$
$$+N^{1/q}C_2^{''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}.$$

By the second part of Lemma 3, it holds that

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{''''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big) + 1\Big\}^{1/2}\Big\} + N^{1/q}C_2^{'''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q}\Big\}.$$

This immediately implies the desired result. $\square$

Lemma 5. *Suppose Assumption 1 holds. Then there exists absolute constants C_1 such that*

$$\mathbb{E}|W_t - W_t^*|^q \leq C_1 \max_{i=1}^{n}\mathbb{E}\big\{|\delta_{t,i} - \delta_{t,i}^*|^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}. \tag{31}$$

If in addition $\mathbb{E}\big\{|\delta_{t,i}|_q^q\big\} = O(1)$ for all $1 \leq i \leq n$, then there exists absolute constants C' such that

$$\mathbb{E}(|W_t|^q)^{1/q} \leq C'\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q}. \tag{32}$$

Proof. The proof is similar to that of Lemma 3. The proof of the Equation (32) is simpler and simpler than Equation (31). So only the proof of Equation (31) is presented. Note that since $\{x_t\}_{t=1}^T$ and $\{\delta_t\}_{t=1}^T$ are independent, and that $\{x_t\}_{t=1}^T$ are independent identically distributed,

$$\delta_t^* = \frac{1}{n}\sum_{i=1}^n \delta_{t,i}^* K_h(x - x_{t,i}).$$

Step 1. Note that, by the Newton's binomial

$$\begin{aligned}
\mathbb{E}|\delta_t - \delta_t^*|^q =&\mathbb{E}\Big\{\Big|\frac{1}{n}\sum_{i=1}^n(\delta_{t,i}^* - \delta_{t,i})K_h(x - x_{t,i})\Big|^q\Big\} \\
\le&\frac{1}{n^q}\mathbb{E}\Big\{\sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q\\ \beta_1\ge 0,\ldots,\beta_n\ge 0}}\binom{q}{\beta_1,\beta_2,\ldots,\beta_n}\prod_{j=1}^n\big|(\delta_{t,i}^* - \delta_{t,i})K_h(x-x_{t,i})\big|^{\beta_j}\Big\} \\
=&\frac{1}{n^q}\mathbb{E}\Big\{\sum_{k=1}^q\sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q\\ \\ \beta=(\beta_1,\ldots,\beta_n),|\beta|_0=k,\beta\ge 0}}\binom{q}{\beta_1,\beta_2,\ldots,\beta_n}\prod_{j=1}^n\big|(\delta_{t,i}^* - \delta_{t,i})K_h(x-x_{t,i})\big|^{\beta_j}\Big\}.
\end{aligned}$$

Step 2. For a fixed $\beta = (\beta_1,\ldots,\beta_n)$ such that $\beta_1+\ldots+\beta_n = q$ and that $|\beta|_0 = k$, consider

$$\mathbb{E}\Big\{\prod_{j=1}^n\big|(\delta_{t,i}^* - \delta_{t,i})K_h(x-x_{t,i})\big|^{\beta_j}\Big\}.$$

Without loss of generality, assume that $\beta_1,\ldots,\beta_k$ are non-zero. Then it holds that

$$\begin{aligned}
&\mathbb{E}\Big\{\big|(\delta_{t,1}^* - \delta_{t,1})\big|^{\beta_1}\big|K_h(x-x_{t,1})\big|^{\beta_1}\cdots\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}\big|K_h(x-x_{t,k})\big|^{\beta_k}\Big\} \\
=&\mathbb{E}_\delta\Big\{\int\big|(\delta_{t,1}^* - \delta_{t,1}\big|^{\beta_1}\big|K_h(x-r)\big|^{\beta_1}d\mu(r)\cdots\int\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}\big|K_h(x-r)\big|^{\beta_k}d\mu(r)\Big\} \\
=&\mathbb{E}_\delta\Big\{\int\big|(\delta_{t,1}^* - \delta_{t,1}\big|^{\beta_1}\frac{\big|K(s)\big|^{\beta_1}}{h^{d(\beta_1-1)}}d\mu(s)\cdots\int\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}\frac{\big|K(s)\big|^{\beta_k}}{h^{d(\beta_k-1)}}d\mu(s)\Big\} \\
\le&h^{-d\sum_{j=1}^k(\beta_j-1)}\mathbb{E}_\delta\Big\{\big|(\delta_{t,1}^* - \delta_{t,1})\big|^{\beta_1}C_K^{\beta_1}\cdots\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}C_K^{\beta_k}\Big\} \\
\le&h^{-d(q-k)}C_K^q\mathbb{E}_\delta\Big\{\max_{i=1}^n|\delta_{t,i} - \delta_{t,i}^*|^{\sum_{j=1}^k\beta_k}\Big\} \\
\le&h^{-d(q-k)}C_K^q\mathbb{E}_\delta\{\max_{i=1}^n|\delta_{t,i} - \delta_{t,i}^*|^q\}
\end{aligned}$$

where the third equality follows by using the change of variable $s = \frac{x-r}{h}$, the first inequality by assumption 2.

Step 3. Let $k \in \{1,\ldots,q\}$ be fixed. Note that $\binom{q}{\beta_1,\beta_2,\ldots,\beta_n} \le q!$. Consider set

$$\mathcal{B}_k = \Big\{\beta\in\mathbb{N}^n : \beta\ge 0, \beta_1+\ldots+\beta_n = q, |\beta|_0 = k\Big\}.$$

To bound the cardinality of the set $\mathcal{B}_k$, first note that since $|\beta|_0 = k$, there are $\binom{n}{k}$ number of ways to choose the index of non-zero entries of β.

Suppose $\{i_1,\ldots i_k\}$ are the chosen index such that $\beta_{i_1}\neq 0,\ldots,\beta_{i_k}\neq 0$. Then the constrains $\beta_{i_1} > 0,\ldots,\beta_{i_1} > 0$ and $\beta_{i_1}+\ldots+\beta_{i_k} = q$ are equivalent to that of diving q balls into k groups (without distinguishing each ball). As a result there are $\binom{q-1}{k-1}$ number of ways to choose the $\{\beta_{i_1},\ldots,\beta_{i_k}\}$ once the index $\{i_1,\ldots i_k\}$ are chosen.

Step 4. Combining the previous three steps, it follows that for some constants $C_q, C_1 > 0$

only depending on q,

$$\begin{aligned}
\mathbb{E}|W_t - W_t^*|^q \leq & \frac{1}{n^q}\mathbb{E}\Big\{\sum_{k=1}^{q} \sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta=(\beta_1,\ldots,\beta_n),|\beta|_0=k,\beta\geq 0}} \binom{q}{\beta_1,\beta_2,\ldots,\beta_n} \prod_{j=1}^{n} \big|(\delta_{t,i}^* - \delta_{t,i})K_h(x-x_{t,i})\big|^{\beta_j}\Big\} \\
\leq & \frac{1}{n^q}\sum_{k=1}^{q}\binom{n}{k}\binom{q-1}{k-1} q! h^{-d(q-k)} C_K^q \mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\} \\
\leq & \frac{1}{n^q}\sum_{k=1}^{q} n^k C_q C_K^q h^{-d(q-k)} \mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\} \\
\leq & C_1 \mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + \Big(\frac{1}{nh^d}\Big)^{q-2} + \ldots + \Big(\frac{1}{nh^d}\Big) + 1\Big\} \\
\leq & C_1 \mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\} q \Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\},
\end{aligned}$$

where the second inequality is satisfied by step 3 and that $\binom{q}{\beta_1,\beta_2,\ldots,\beta_n} \leq q!$, while the third inequality is achieved by using that $\binom{n}{k}\binom{q-1}{k-1}q! \leq \binom{n}{k}C_q \leq n^k C_q$. Moreover, given that $\frac{1}{n^q}n^k h^{-d(q-k)} = \Big(\frac{1}{nh^d}\Big)^{q-k}$ the fourth inequality is obtained. The last inequality holds because if $\frac{1}{nh^d} \leq 1$, then $\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + \ldots + \Big(\frac{1}{nh^d}\Big) + 1\Big\} \leq q$, and if $\frac{1}{nh^d} \geq 1$, then $\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + \ldots + \Big(\frac{1}{nh^d}\Big) + 1\Big\} \leq q\Big(\frac{1}{nh^d}\Big)^{q-1}$. □

Lemma 6. *Suppose Assumption 1* **d** *holds. Let* $\rho \leq T$ *be such that* $\rho n h^d \geq \log(T)$ *and* $T \geq 3$. *Let* $N \in \mathbb{Z}^+$ *be such that* $N \geq \rho$. *Then, it holds that*

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} W_t|^q\Big\}^{1/q} \leq N^{1/2}C\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + N^{1/q}C'\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q} + 1\Big\}.$$

Proof. We have that $q > 2$ and $E|W_1| < \infty$ by the use of Lemma 5. Then, making use of Theorem 1 of Liu et al. (2013), we obtain that

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1\Big\{\sum_{j=1}^{N}\Theta_{j,2} + \sum_{j=N+1}^{\infty}\Theta_{j,q} + \{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& + N^{1/q}C_2\Big\{\sum_{j=1}^{N} j^{1/2-1/q}\Theta_{j,q} + \{\mathbb{E}|W_1|^q\}^{1/q}\Big\},
\end{aligned}$$

where $\Theta_{j,q} = \{\mathbb{E}(|W_j^* - W_j|^q)\}^{1/q}$. Moreover, we observe that since $\Theta_{j,2} \leq \Theta_{j,q}$ for any $q \geq 2$, it follows

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} W_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1\Big\{\sum_{j=1}^{\infty}\Theta_{j,q} + \{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& + N^{1/q}C_2\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}\Theta_{j,q} + \{\mathbb{E}|W_1|^q\}^{1/q}\Big\}.
\end{aligned}$$

Next, by the first part of Lemma 3,

$$\Theta_{j,q}^q \leq C\mathbb{E}\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}.$$

Since we have that $N \geq \frac{1}{nh^d}$, the above inequality further implies that

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k}W_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1^{'}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^{*}|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}\Big\}^{1/q}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& +N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^{*}|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\} \\
\leq & N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^{*}|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& +N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^{*}|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\} \\
\leq & N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^{*}|^q\}\Big\{\Big(N\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& +N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^{*}|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\}.
\end{aligned}$$

From Assumption 1 **d**, the above inequality further implies that

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k}W_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1^{'''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& +N^{1/q}C_2^{''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\}.
\end{aligned}$$

By the second part of Lemma 3, it holds that

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k}Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{''''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)+1\Big\}^{1/2}\Big\}+N^{1/q}C_2^{'''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}\Big\}.$$

This immediately implies the desired result. □

D Additional Technical Results

Lemma 7. *Suppose that $f, g : [0,1]^d \to \mathbb{R}$ such that $f, g \in \mathcal{H}^r(L)$ for some $r \geq 1$ $L > 0$. Suppose in addition that $\{x_m\}_{m=1}^M$ is a collection of grid points randomly sampled from a density $u : [0,1]^d \to \mathbb{R}$ such that $\inf_{x\in[0,1]^d} u(x) \geq c_u > 0$. If $\|f-g\|_\infty \geq \kappa$ for some parameter $\kappa > 0$, then*

$$\mathbb{P}\Big\{\max_{m=1}^{M}|f(x_m)-g(x_m)| \geq \frac{3}{4}\kappa\Big\} \geq 1-\exp\big(-cM\kappa^d\big),$$

where c is a constant only depending on d.

Proof. Let $h = f - g$. Since $f, g \in \mathcal{H}^r(L)$, $h \in \mathcal{H}^r(L)$. Since $r \geq 1$, we have that

$$|h(x)-h(x')| \leq L|x-x'| \quad \text{for all} \quad x, x' \in [0,1]^d.$$

for some absolute constant $L > 0$. Let $x_0 \in [0,1]^d$ be such that

$$|h(x_0)| = \|h\|_\infty.$$

Then for all $x' \in B(x_0, \frac{\kappa}{4L}) \cap [0,1]^d$,

$$|h(x')| \geq |h(x_0)| - L|x_0-x'| \geq \frac{3}{4}\kappa.$$

Therefore

$$\mathbb{P}\Big\{\max_{m=1}^{M}|f(x_m)-g(x_m)| < \frac{3}{4}\kappa\Big\} \leq P\Big(\{x_m\}_{m=1}^M \notin B\big(x_0, \frac{\kappa}{4L}\big)\Big).$$

Since

$$P\Big(\{x_m\}_{m=1}^M \notin B(x_0, \frac{\kappa}{4L})\Big) = \Big\{1 - P\Big(x_1 \in B(x_0, \frac{\kappa}{4L})\Big)\Big\}^M \leq \Big(1 - \{\frac{c_u \kappa}{4L}\}^d\Big)^M \leq \exp\big(-Mc\kappa^d\big),$$

the desired result follows. □

Lemma 8. *Let $\mathcal{J}$ be defined as in Definition 1 and suppose Assumption 1* **e** *holds. Denote*

$$\zeta_k = \frac{9}{10}\min\{\eta_{k+1} - \eta_k, \eta_k - \eta_{k-1}\}\ k \in \{1, ..., K\}.$$

Then for each change-point η_k there exists a seeded interval $\mathcal{I}_k = (s_k, e_k]$ such that
a. *$\mathcal{I}_k$ contains exactly one change-point η_k;*
b. $\min\{\eta_k - s_k, e_k - \eta_k\} \geq \frac{1}{16}\zeta_k$*; and*
c. $\max\{\eta_k - s_k, e_k - \eta_k\} \leq \zeta_k$*;*

Proof. These are the desired properties of seeded intervals by construction. The proof is the same as theorem 3 of Kovács et al. (2020) and is provided here for completeness.

Since $\zeta_k = \Theta(T)$, by construction of seeded intervals, one can find a seeded interval $(s_k, e_k] = (c_k - r_k, c_k + r_k]$ such that $(c_k - r_k, c_k + r_k] \subseteq (\eta_k - \zeta_k, \eta_k + \zeta_k]$, $r_k \geq \frac{\zeta_k}{4}$ and $|c_k - \eta_k| \leq \frac{5r_k}{8}$. So $(c_k - r_k, c_k + r_k]$ contains only one change-point η_k. In addition,

$$e_k - \eta_k = c_k + r_k - \eta_k \geq r_k - |c_k - \eta_k| \geq \frac{3r_k}{8} \geq \frac{3\zeta_k}{32},$$

and similarly $\eta_k - s_k \geq \frac{3\zeta_k}{32}$, so **b** holds. Finally, since $(c_k - r_k, c_k + r_k] \subseteq (\eta_k - \zeta_k, \eta_k + \zeta_k]$, it holds that $c_k + r_k \leq \eta_k + \zeta_k$ and so

$$e_k - \eta_k = c_k + r_k - \eta_k \leq \zeta_k.$$

□

D.1 Univariate CUSUM

We introduce some notation for one-dimensional change-point detection and the corresponding CUSUM statistics. Let $\{\mu_i\}_{i=1}^n, \{\omega_i\}_{i=1}^n \subseteq \mathbb{R}$ be two univariate sequences. We will make the following assumptions.

Assumption 1 (Univariate mean change-points)**.** *Let $\{\eta_k\}_{k=0}^{K+1} \subseteq \{0, \dots, n\}$, where $\eta_0 = 0$ and $\eta_{K+1} = T$, and*

$$\omega_t \neq \omega_{t+1} \ \text{ if and only if } \ t \in \{\eta_1, ...\eta_K\},$$

Assume

$$\min_{k=1}^{K+1}(\eta_k - \eta_{k-1}) \geq \Delta > 0,$$
$$0 < |\omega_{\eta_{k+1}} - \omega_{\eta_k}| = \kappa_k \text{ for all } k = 1, \dots, K.$$

We also have the corresponding CUSUM statistics over any generic interval $[s, e] \subseteq [1, T]$ defined as

$$\widetilde{\mu}_t^{s,e} = \sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{i=s+1}^{t}\mu_i - \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{i=t+1}^{e}\mu_i,$$
$$\widetilde{\omega}_t^{s,e} = \sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{i=s+1}^{t}\omega_i - \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{i=t+1}^{e}\omega_i.$$

Throughout this section, all of our results are proven by regarding $\{\mu_i\}_{i=1}^T$ and $\{\omega_i\}_{i=1}^T$ as two deterministic sequences. We will frequently assume that $\widetilde{\mu}_t^{s,e}$ is a good approximation of $\widetilde{\omega}_t^{s,e}$ in ways that we will specify through appropriate assumptions.

Consider the following events

$$\mathcal{A}((s,e],\rho,\gamma)=\left\{\max_{t=s+\rho+1}^{e-\rho}|\widetilde{\mu}_t^{s,e}-\widetilde{\omega}_t^{s,e}|\le\gamma\right\};$$

$$\mathcal{B}(r,\rho,\gamma)=\left\{\max_{N=\rho}^{T-r}\left|\frac{1}{\sqrt{N}}\sum_{t=r+1}^{r+N}(\mu_t-\omega_t)\right|\le\gamma\right\}\bigcup\left\{\max_{N=\rho}^{r}\left|\frac{1}{\sqrt{N}}\sum_{t=r-N+1}^{r}(\mu_t-\omega_t)\right|\le\gamma\right\}.$$

Lemma 9. *Suppose Assumption 1 holds. Let $[s,e]$ be an subinterval of $[1,T]$ and contain at least one change-point η_r with $\min\{\eta_r-s,e-\eta_r\}\ge cT$ for some constant $c>0$. Let $\kappa_{\max}^{s,e}=\max\{\kappa_p:\min\{\eta_p-s,e-\eta_p\}\ge cT\}$. Let*

$$b\in\arg\max_{t=s+\rho}^{e-\rho}|\widetilde{\mu}_t^{s,e}|.$$

For some $c_1>0$, $\lambda>0$ and $\delta>0$, suppose that the following events hold

$$\mathcal{A}((s,e],\rho,\gamma), \tag{33}$$

$$\mathcal{B}(s,\rho,\gamma)\cup\mathcal{B}(e,\rho,\gamma)\cup\bigcup_{\eta\in\{\eta_k\}_{k=1}^{K}}\mathcal{B}(\eta,\rho,\gamma) \tag{34}$$

and that

$$\max_{t=s+\rho}^{e-\rho}|\widetilde{\mu}_t^{s,e}|=|\widetilde{\mu}_b^{s,e}|\ge c_1\kappa_{\max}^{s,e}\sqrt{T} \tag{35}$$

If there exists a sufficiently small $c_2>0$ such that

$$\gamma\le c_2\kappa_{\max}^{s,e}\sqrt{T}\quad\textit{and that}\quad\rho\le c_2T, \tag{36}$$

then there exists a change-point $\eta_k\in(s,e)$ such that

$$\min\{e-\eta_k,\eta_k-s\}>c_3T\quad\textit{and}\quad|\eta_k-b|\le C_3\max\{\gamma^2\kappa_k^{-2},\rho\},$$

where c_3 is some sufficiently small constant independent of T.

Proof. The proof is the same as that for Lemma 22 in Wang et al. (2020). □

Lemma 10. *If $[s,e]$ contain two and only two change-points η_r and η_{r+1}, then*

$$\max_{t=s}^{e}|\widetilde{\omega}_t^{s,e}|\le\sqrt{e-\eta_{r+1}}\kappa_{r+1}+\sqrt{\eta_r-s}\kappa_r.$$

Proof. This is Lemma 15 in Wang et al. (2020). □

E Common Stationary Processes

Basic time series models which are widely used in practice, can be incorporated by Assumption 1b and c. Functional autoregressive model (FAR) and functional moving average model (FMA) are presented in examples 1 below. The vector autoregressive (VAR) model and vector moving average (VMA) model can be defined in similar and simpler fashions.

Example 1 (FMA and FAR)**.** *Let $\mathcal{L}=\mathcal{L}(H,H)$ be the set of bounded linear operators from H to H, where $H=\mathcal{L}_\infty$. For $A\in\mathcal{L}$, we define the norm operator $||A||_{\mathcal{L}}=\sup_{||\varepsilon||_H\le1}||A\varepsilon||_H$. Suppose $\theta_1,\Psi\in\mathcal{L}$ with $\|\Psi\|_{\mathcal{L}}<1$ and $\|\theta_1\|_{\mathcal{L}}<\infty$.*

a) *For FMA model, let $(\varepsilon_t:t\in\mathbb{Z})$ be a sequence of independent and identically distributed random $\mathcal{L}_\infty$ functions with mean zero. Then the FMA time series $(\xi_j:j\in\mathbb{Z})$ of order 1 is given by the equation*

$$\xi_t=\theta_1(\varepsilon_{t-1})+\varepsilon_t=g(\dots,\varepsilon_{-1},\varepsilon_0,\varepsilon_1,\dots,\varepsilon_{t-1},\varepsilon_t). \tag{37}$$

For any $t\ge2$, by (37) *we have that*

$$\xi_t-\xi_t^*=0$$

and $\xi_1 - \xi_1^* = \theta_1(\varepsilon_0) - \theta_1(\varepsilon_0^{'})$. *As a result*

$$\sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}(||\xi_t - \xi_t^*||_\infty^q)^{1/q} = \mathbb{E}(||\xi_1 - \xi_1^*||_\infty^q)^{1/q} = \mathbb{E}(\|\theta_1(\varepsilon_0) - \theta_1(\varepsilon_0^{'})\|_\infty^q)^{1/q} < \infty.$$

Therefore Assumption 1b is satisfied by FMA models.

b) *We can define a FAR time series as*

$$\xi_t = \Psi(\xi_{t-1}) + \varepsilon_t. \tag{38}$$

It admits the expansion,

$$\begin{aligned}
\xi_t =& \sum_{j=0}^{\infty} \Psi^j(\varepsilon_{t-j}) \\
=&\Psi(\varepsilon_t) + \Psi^1(\varepsilon_{t-1}) + ... + \Psi^t(\varepsilon_0) + \Psi^{t+1}(\varepsilon_{-1}) + ... \\
=&g(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t).
\end{aligned}$$

Then for any $t \geq 1$, *we have that* $\xi_t - \xi_t^* = \Psi^t(\varepsilon_0) - \Psi^t(\varepsilon_0^{'})$. *Thus,*

$$\begin{aligned}
\sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}(||\xi_t - \xi_t^*||_\infty^q)^{1/q} =& \sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}(||\Psi^t(\varepsilon_0) - \Psi^t(\varepsilon_0')||_\infty^q)^{1/q} \\
\leq& \sum_{t=1}^{\infty} t^{1/2-1/q}||\Psi||_{\mathcal{L}}^t\mathbb{E}(||\varepsilon_0 - \varepsilon_0'||_\infty^q)^{1/q} < \infty.
\end{aligned}$$

Assumption 1b incorporates FAR time series.

References

[1] Giné, E. and Guillou, A. [2002], 'Rates of strong uniform consistency for multivariate kernel density estimators', *Annales de l'Institut Henri Poincare (B) Probability and Statistics* **38**(6), 907–921.

[2] Jiang, H. [2017], Uniform convergence rates for kernel density estimation, *in* 'International Conference on Machine Learning', PMLR, pp. 1694–1703.

[3] Kirch, C. [2006], Resampling methods for the change analysis of dependent data, PhD thesis, Universität zu Köln.

[4] Kovács, S., Li, H., Bühlmann, P. and Munk, A. [2020], 'Seeded binary segmentation: A general methodology for fast and optimal change point detection', *arXiv preprint arXiv:2002.06633* .

[5] Liu, W., Xiao, H. and Wu, W. B. [2013], 'Probability and moment inequalities under dependence', *Statistica Sinica* pp. 1257–1272.

[6] National Oceanic and Atmospheric Administration [n.d.], 'ENSO effects across the northeastern Caribbean', `https://www.weather.gov/sju/climo_enso`.

[7] Physical Sciences Laboratory [2020], 'COBE SST2 and Sea-Ice', `https://psl.noaa.gov/data/gridded/data.cobe2.html`.

[8] Rinaldo, A. and Wasserman, L. [2010], 'Generalized density clustering', *The Annals of Statistics* **38**(5), 2678–2722.

[9] Sriperumbudur, B. and Steinwart, I. [2012], Consistency and rates for clustering with dbscan, *in* 'Artificial Intelligence and Statistics', PMLR, pp. 1090–1098.

[10] Tsybakov, A. B. [2009], *Introduction to Nonparametric Estimation*, Springer series in statistics, Springer, Dordrecht.

[11] Vershynin, R. [2018], *High-dimensional probability: An introduction with applications in data science*, Vol. 47, Cambridge university press.

[12] Wang, D., Yu, Y. and Rinaldo, A. [2020], 'Univariate mean change point detection: Penalization, cusum and optimality', *Electronic Journal of Statistics* **14**(1), 1917–1961.

[13] Wikipedia, 2004 Atlantic hurricane season [2022], '2004 Atlantic hurricane season', `https://en.wikipedia.org/wiki/2004_Atlantic_hurricane_season#:~:text=The`.