Testing for a monotone density using L_k-distances between the empirical distribution function and its concave majorant

Vladimir N. Kulikov[†]
Hendrik P. Lopuhaä[‡]

01.06.2004

Abstract: We prove asymptotic normality for L_k-functionals $\int |\hat{F}_n - F_n|^k g(t)\, dt$, where F_n is the empirical distribution function of a sample from a decreasing density and $\hat{F}_n$ is the least concave majorant of F_n. From this we derive two test statistics for the null hypothesis that a probability density is monotone. These tests are compared with existing proposals such as the supremum distance between $\hat{F}_n$ and F_n.

[†]EURANDOM, P.O. Box 513 - 5600 MB Eindhoven, the Netherlands, kulikov@eurandom.nl
[‡]Delft University of Technology, Mekelweg 4 2628 CD Delft, the Netherlands, h.p.lopuhaa@its.tudelft.nl

1 Introduction

In many nonparametric statistical models the parameter of interest is some unknown function on the real line. This function may, for instance, be a distribution function, a probability density, a regression function or a failure rate. Often qualitative assumptions such as concavity, monotonicity or unimodality are plausible, and a natural question is how to test such assumptions.

For testing unimodality of a probability density there exist several proposals. Silverman (1981) proposed a test based on the critical bandwidth of a kernel density estimate. Hartigan and Hartigan (1985) and Müller and Sawitzki (1991) based a test on the DIP or excess mass functional. Further results concerning these functionals can be found in Mammen, Maron and Fisher (1992) and Cheng and Hall (1997). In the context of regression there is an extensive literature on testing monotonicity. Schlee (1982) proposed a test based on an estimate of the derivative of the regression function. Bowman, Jones and Gijbels (1998) extended Silverman's (1981) test. Hall and Heckman (2000) proposed a test based on running gradients, which is related to the DIP/excess mass method in the density context. Ghosal, Sen and van der Vaart (2000) discuss a locally weighted version of Kendall's tau. Durot (2003) investigates a test statistic based on the supremum distance between the empirical distribution function and its concave majorant. For testing a constant failure rate against a monotone increasing failure rate, Robertson, Wright and Dijkstra (1988) (Chapter 5) mention several references. Cohen and Sackrowitz (1993) compare several methods for testing an increasing intensity of a Poisson process and Dümbgen and Spokoiny (2001) study several qualitative hypotheses in the white noise model.

The literature on testing monotonicity in the context of probability densities seems to be limited. Woodroofe and Sun (1999) propose two test statistics for testing the simple null hypothesis of f being uniform on $[0,1]$ against the alternative that it is nonincreasing on $[0,1]$. The first one is a likelihood ratio test statistic based on the penalized maximum likelihood estimator for f. The second one compares the corresponding cumulative distribution function with the uniform distribution function by means of the supremum distance.

In this paper we address the problem of testing the composite hypothesis that f is non-increasing against the alternative it is not. We will discuss a number of test statistics based on the difference between the empirical distribution function F_n and its least concave majorant $\hat{F}_n$, by which we mean the smallest concave function that lies above F_n. This difference has been of interest to several authors. Kiefer and Wolfowitz (1976) showed that $(\log n)^{-1}n^{2/3}\sup|\hat{F}_n - F_n|$ converges to zero with probability one, but the precise rate of convergence or limiting distribution was not given. Wang (1994) obtained the limit distribution of $n^{2/3}(\hat{F}_n(t) - F_n(t))$, for t being a fixed point in $(0,1)$. This was extended to process convergence by Kulikov and Lopuhaä (2003). Pointwise convergence was also obtained in a regression setting by Durot and Toquet (2002).

The test statistic based on $\sup|\hat{F}_n - F_n|$ would be a direct extension of Durot (2003) and is similar in spirit to the second proposal of Woodroofe and Sun (1999). Similar to their results we show that for this test statistic the uniform density is least favorable among all non-increasing densities on $[0,1]$, and we determine its limit distribution under uniformity. This enables us to determine critical values and compute probabilities of committing a type I error. However, if f satisfies additional smoothness conditions, then the results of Kiefer and Wolfowitz (1976) imply that $\sup|\hat{F}_n - F_n|$ is of smaller order, which causes the probability of committing a type I error tend to zero. This calls for a more detailed description of the limit distribution under smooth f, but this is still unknown. Computer simulations demonstrate that $\sup|\hat{F}_n - F_n|$ is not so powerful at increasing alternatives that are close to the uniform density.

We propose to construct a test statistic based on the L_k-distance between $\hat{F}_n$ and F_n. We investigate two possibilities, for which we determine the limiting distribution under uniformity. To investigate their behavior under additional smoothness assumptions, we show that for smooth f, the L_k-distance between $\hat{F}_n$ and F_n is asymptotically normal. This result is established in Section 2 and is similar to the one in Durot and Toquet (2002) who, independently of our efforts, obtained a similar result in the regression setting. One of the main differences between the regression setting and our setup is the embedding of the empirical process. In the regression setting the empirical process can be embedded directly into Brownian motion itself, whereas in our setup it can only be embedded in the process

$$s \mapsto W\Big(n^{1/3}\big(F(t+n^{-1/3}s) - F(t)\big)\Big). \tag{1.1}$$

This introduces an additional difficulty of approximating the value of the concave majorant of the process at zero by the corresponding value of the process $s \mapsto W\big(f(t)s\big)$. Although the maximum difference between the two processes is too large, the key observation that makes things work is that the values of the concave majorants at zero are sufficiently close.

By estimating the normalizing constants in the limit theorem for the L_k distance we derive two possible test statistics for the null hypothesis that f is strictly increasing. In Section 3 we investigate their limit behavior, as well as that of the supremum distance between $\hat{F}_n$ and F_n. In Section 4 we discuss the results of a small simulation study in which we compare our test statistic with other proposals from Woodroofe and Sun (1999) and Durot (2003) at several underlying densities including the uniform. The tests based on L_k distances appear to be superior against convex alternatives. All proofs are put in an Appendix at the end of the paper.

2 Convergence of L_k-functionals

Let $X_1, X_2, \ldots, X_n$ be a sample from a decreasing density f and denote F as the corresponding distribution function. Suppose that f has bounded support, which then without loss of generality may be taken to be the interval $[0,1]$. Let $\hat{F}_n$ be the least concave majorant of the empirical distribution function F_n on $[0,1]$. Consider the process

$$A_n(t) = n^{2/3}\Big(\hat{F}_n(t) - F_n(t)\Big), \quad t \in [0,1]. \tag{2.1}$$

The limiting distribution of A_n, can be described in terms of the operator CM_I that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the least concave majorant of h on the interval $I \subset \mathbb{R}$. If we define the process

$$Z(t) = W(t) - t^2, \tag{2.2}$$

where W denotes standard two-sided Brownian motion originating from zero, then it is shown in Wang (1994) that, for $t \in (0,1)$ fixed, $A_n(t)$ converges in distribution to $c_1(t)\zeta(0)$, where $c_1(t)$ is defined in (2.5), and

$$\zeta(t) = [\mathrm{CM}_{\mathbb{R}} Z](t) - Z(t). \tag{2.3}$$

This result was extended to process convergence by Kulikov and Lopuhaä (2003), where it is proved that, for $t \in (0,1)$ fixed, the process

$$\zeta_{nt}(s) = \begin{cases} c_1(t)A_n(t + c_2(t)sn^{-1/3}) & \text{for } t + c_2(t)sn^{-1/3} \in (0,1), \\ 0 & \text{otherwise,} \end{cases} \tag{2.4}$$

converges in distribution to the process ζ in the space $D(\mathbb{R})$ of cadlag functions on $\mathbb{R}$, where

$$c_1(t) = \left(\frac{|f'(t)|}{2f^2(t)}\right)^{1/3} \quad \text{and} \quad c_2(t) = \left(\frac{4f(t)}{|f'(t)|^2}\right)^{1/3}. \tag{2.5}$$

In the remainder of this section we will prove asymptotic normality of L_k-functionals of the type $\int_0^1 A_n(t)^k g(t)\,dt$, where g is continuous. Let us very briefly outline the line of reasoning how we establish this result.

First observe that, up to constants, A_n is the image of F_n under the operator D_I, that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the difference between the least concave majorant of h on the interval I and h itself:

$$\mathrm{D}_I h = \mathrm{CM}_I h - h.$$

We can therefore write $A_n = n^{2/3}[\mathrm{D}_{[0,1]} F_n]$. We will approximate F_n by means of a Brownian motion version and use its image under $\mathrm{D}_{[0,1]}$ to approximate A_n. To this end, let E_n denote the empirical process $\sqrt{n}(F_n - F)$ and let B_n be a Brownian bridge constructed on the same probability space as the uniform empirical process $E_n \circ F^{-1}$ via the Hungarian embedding of Kómlos, Major and Tusnády (1975). Let ξ_n be a $N(0,1)$ distributed random variable independent of B_n and define versions W_n of Brownian motion by

$$W_n(t) = B_n(t) + \xi_n t, \quad t \in [0,1]. \tag{2.6}$$

Write $F_n^E = F_n$ and let F_n^W be its Brownian approximation defined by

$$F_n^W(t) = F(t) + n^{-1/2} W_n\big(F(t)\big), \quad t \in [0,1], \tag{2.7}$$

where W_n is defined in (2.6), and let

$$A_n^W(t) = n^{2/3}[\mathrm{D}_{[0,1]} F_n^W](t). \tag{2.8}$$

We impose the following conditions on f:

- (A1) f is a twice continuous differentiable decreasing density with support on $[0,1]$;
- (A2) $0 < f(1) \le f(t) \le f(s) \le f(0) < \infty$ for $0 \le s \le t \le 1$;
- (A3) $\inf_{t \in [0,1]} |f'(t)| > 0$.

Under these conditions, for $J = E, W$, we first approximate the process $s \mapsto F_n^J(t + n^{-1/3}s)$ by the process $Y_{nt}(s) + L_{nt}^J(s)$, where

$$Y_{nt}(s) = W_n\left(n^{1/3}\big(F(t + n^{-1/3}s) - F(t)\big)\right) + \frac{1}{2}f'(t)s^2,$$

and L_{nt}^J denotes a linear drift (see Lemma 5.1). Since the operator D_I is invariant under addition of linear terms, the moments of $A_n^J(t)$ can be approximated by the moments of

$[\mathrm{D}_I Y_{nt}](0)$ (see Lemma 5.5). By uniform continuity of Brownian motion on compacta, the process $Y_{nt}(s)$ is close to the process $W_n(f(t)s) + \frac{1}{2}f'(t)s^2$. This leads to the following key result:

$$EA_n^J(t)^k = \left(\frac{2f(t)^2}{|f'(t)|}\right)^{k/3} E\zeta(0)^k + o(n^{-1/6}),$$

where ζ is defined in (2.3) (see Lemma 5.7). A direct consequence is that the difference between the processes $A_n(t)^k$ and $A_n^W(t)^k$ is of smaller order than $n^{-1/6}$ (see Lemma 5.8). In view of Theorem 2.1, this means that it suffices to prove asymptotic normality for L_k-functionals

$$n^{1/6} \int_0^1 \left(A_n^W(t)^k - EA_n^W(t)^k\right) g(t)\, dt.$$

The fact that Brownian motion has independent increments will ensure that the process A_n^W is mixing (see Lemma 5.9). This allows us to approximate the integral by a sum of independent integral terms, which then leads to the following theorem.

Theorem 2.1 *Suppose that f satisfies conditions (A1)-(A3). Let g be a continuous function on $[0,1]$ and let A_n be defined by (2.1). Then for all $k \geq 1$, with*

$$\mu = E\big[\zeta(0)^k\big] \int_0^1 \frac{2^{k/3} f(t)^{2k/3}}{|f'(t)|^{k/3}} g(t)\, dt,$$

$n^{1/6}\left(\int_0^1 A_n(t)^k g(t)\, dt - \mu\right)$ *converges in distribution to a normal random variable with mean zero and variance*

$$\sigma^2 = \int_0^1 \frac{2^{(2k+5)/3} f(t)^{(4k+1)/3}}{|f'(t)|^{(2k+2)/3}} g(t)^2\, dt \int_0^\infty \mathrm{cov}(\zeta(0)^k, \zeta(s)^k)\, ds,$$

where ζ is defined in (2.3).

Remark 2.1 The condition (A1) on f in Theorem 2.1 can be relaxed somewhat. At the cost of additional technicalities, the theorem remains true if we require $|f'(x) - f'(y)| \leq C|x-y|^{1/2+\epsilon}$, for some $\epsilon > 0$ and $C > 0$ not depending on f.

3 Testing monotonicity of the underlying density

To test the composite null hypothesis that the underlying density f is non-increasing, various approaches may be transferred from the regression setting, where this problem has been studied intensively for regression curves. The most applicable one seems to be the DIP-type test based on the supremum distance between the empirical distribution function and its concave majorant. Durot (2003) has used this approach for testing the composite null hypothesis that the regression curve is non-increasing. The limiting distribution of the supremum distance was obtained under the hypothesis of constant regression curves, which were shown to be least favorable. Woodroofe and Sun (1999) investigated a similar test statistic in the setting of density estimation to test the simple null hypothesis of uniformity, and obtained the limiting distribution under the null hypothesis. In both cases the test statistic is easy to calculate and it appears to be powerful at increasing alternatives. This suggests

$$S_n = \sqrt{n} \sup_{t\in[0,1]} \Big(\hat{F}_n(t) - F_n(t)\Big), \tag{3.9}$$

as a test statistic for the composite null hypothesis that the underlying density f is non-increasing. Similar to Durot (2003), the uniform distribution is least favorable among all non-increasing densities on $[0,1]$. The limit distribution of S_n under uniformity is given in the following theorem.

Theorem 3.1 *Let f be non-increasing on $[0,1]$ and S_n be defined as in (3.9).*

1. *For a sample $X_1, X_2, \dots, X_n$ from f and $U_1, U_2, \dots, U_n$ defined by $U_i = F(X_i)$, it holds that $S_n(X_1, \dots, X_n) \le S_n(U_1, \dots, U_n)$.*
2. *If $f = 1$, then S_n converges in distribution to $sup_{t\in[0,1]}\big(\hat{W}(t) - W(t)\big)$, where $\hat{W}$ denotes the least concave majorant of W on $[0,1]$.*

Note that if f satisfies additional smoothness conditions, then S_n is of smaller order. This follows immediately from Kiefer and Wolfowitz (1976), who showed that, if f is twice continuously differentiable, $(\log n)^{-1} n^{1/6} S_n$ tends to zero with probability one. The exact limit behavior of S_n for smooth f is still unknown. Computer simulations demonstrate that S_n is powerful at increasing alternatives similar to the ones considered in Woodroofe and Sun (1999). However, these simulations also demonstrate that S_n is not so powerful at increasing alternatives that are closer to the uniform density.

An alternative approach is to construct a test statistic on the basis of

$$\int_0^1 \big(\hat{F}_n(t) - F_n(t)\big)^k \, dt.$$

Its limiting distribution under uniformity can be obtained, but we cannot prove that the uniform distribution is least favorable. Nevertheless, one could consider the following slight modification

$$R_n = n^{k/2} \int_0^1 \big(\hat{F}_n(t) - F_n(t)\big)^k \, dF_n(t). \tag{3.10}$$

For R_n the uniform distribution is least favorable among all non-increasing densities on $[0,1]$, and its limit distribution under uniformity is given in the following theorem.

Theorem 3.2 *Let f be non-increasing on $[0,1]$ and R_n be defined as in (3.10).*

1. *For a sample $X_1, X_2, \dots, X_n$ from f and $U_1, U_2, \dots, U_n$ defined by $U_i = F(X_i)$, it holds that $R_n(X_1, \dots, X_n) \le R_n(U_1, \dots, U_n)$.*
2. *If $f = 1$, then both $n^{k/2} \int_0^1 \big(\hat{F}_n(t) - F_n(t)\big)^k dt$ and R_n converge in distribution to $\int_0^1 \big(\hat{W}(t) - W(t)\big)^k dt$, for all $k = 1, 2, \dots$, where $\hat{W}$ denotes the least concave majorant of W on $[0,1]$.*

Quantiles of the limiting distributions of R_n and S_n, as given in Theorems 3.1 and 3.2, can be obtained by means of computer simulations. They are given in Table 1. The test based on S_n is similar to the one proposed in Durot and Toquet (2002), and has the same limiting distribution. For this reason the quantiles for S_n are taken from that paper. Similar to S_n, the

Table 1: Simulated $(1-\alpha)$-quantiles of the limiting distributions of R_n and S_n.

α	R_n	S_n
0.99	0.812	1.696
0.95	0.655	1.461
0.90	0.581	1.320

limiting behavior of R_n will be different for smooth f. Undoubtedly, it can be deduced from Theorem 2.1 that $n^{1/6}(n^{k/6}R_n - \mu) = n^{1/6}\big(\int_0^1 A_n(t)^k\, dF_n(t) - \mu\big)$ is asymptotically normal with mean zero and variance σ^2, where μ and σ^2 are as in Theorem 2.1 with $g = f$.

Theorem 2.1 can be used to approximate the distribution of a test statistic based on the L_k-distance. The quantities depending on the process ζ can be obtained by simulation. However, the terms depending on the underlying density must be estimated. This requires estimation of f and of its derivative f'. Note that when f has almost flat parts, estimation of the terms with f' appearing in the denominator of μ and σ^2 causes difficulties. To this end we take $g = |f'|^{(k+1)/3}$ in Theorem 2.1. A possible test statistic then is

$$T_n = \frac{n^{1/6}}{\hat{\sigma}_n}\left(\int_0^1 A_n(t)^k |\hat{f}_n'(t)|^{(k+1)/3}\, dt - \hat{\mu}_n\right), \tag{3.11}$$

with

$$\hat{\mu}_n = E[\zeta(0)^k] \int_0^1 2^{k/3} \hat{f}_n(t)^{2k/3} |\hat{f}_n'(t)|^{1/3}\, dt, \tag{3.12}$$

$$\hat{\sigma}_n^2 = \int_0^\infty \mathrm{cov}(\zeta(0)^k, \zeta(s)^k)\, ds \int_0^1 2^{(2k+5)/3} \hat{f}_n(t)^{(4k+1)/3}\, dt, \tag{3.13}$$

where $\hat{f}_n$ and $\hat{f}_n'$ are estimates for f and its derivative. Because $A_n(t) = \mathcal{O}_p(1)$ (see Lemma 5.7), it follows that if $\hat{f}_n$ converges to f and $\hat{f}_n'$ to f' at a fast enough rate, then asymptotic normality from Theorem 2.1 carries over to T_n.

Theorem 3.3 *Suppose that f satisfies conditions (A1)-(A3). Let A_n by defined by (2.1) and T_n by (3.11). Define $\hat{\mu}_n$ and $\hat{\sigma}_n$ as in (3.12) and (3.13). Suppose that for $i = 0, 1$,*

(i) $\sup_{t\in[0,1]} |\hat{f}_n^{(i)}(t)| = O_p(1)$ *and* $1/\inf_{t\in[0,1]} |\hat{f}_n^{(i)}(t)| = O_p(1)$,

(ii) for some $q \geq 1$, $\int_0^1 |\hat{f}_n^{(i)}(t) - f^{(i)}(t)|^q\, dt = o_p(n^{-q/6})$.

Then for all $k \geq 1$,

$$T_n = \frac{n^{1/6}}{\hat{\sigma}_n}\left(\int_0^1 A_n(t)^k |\hat{f}_n'(t)|^{(k+1)/3}\, dt - \hat{\mu}_n\right),$$

converges in distribution to a standard normal random variable.

One possibility is to estimate f and f' by kernel estimators. Let $K : \mathbb{R} \to \mathbb{R}$ be a kernel function satisfying

(K1) K is a continuously differentiable symmetric probability density,

(K2) $K(u) = 0$ for $|u| \geq 1$,

(K3) $K'(u)$ and $uK'(u)$ are of bounded variation.

Define

$$\hat{f}_{n,h}(t) = \frac{1}{h}\int K\left(\frac{t-x}{h}\right)\,dF_n(x), \tag{3.14}$$

$$\hat{f}'_{n,h}(t) = \frac{1}{h^2}\int K'\left(\frac{t-x}{h}\right)\,dF_n(x). \tag{3.15}$$

If f is twice continuously differentiable, then under assumptions (K1)-(K3), the MISE optimal choice for h is $C_1(f)C_2(K)n^{-1/5}$ for estimating f by $\hat{f}_{n,h}$ (see for instance Prakasa Rao (1983)). The constant $C_1(f)$ can be estimated by several techniques, such as the method of cross validation or by using a reference family. If f is twice continuously differentiable, K satisfies (K1)-(K3), and $h = \mathcal{O}(n^{-1/5})$, then the conditions (i)-(ii) of Theorem 3.3 are satisfied by the kernel estimators defined in (3.14) and (3.15). See Lemmas 5.13 and 5.14.

One may wish to account for the inconsistency of $\hat{f}_{n,h}$ at the boundaries of the support of f. A simple family of boundary kernels that modifies K at the left boundary is the following linear multiple of K:

$$K^L_\alpha(u) = \frac{\nu_{2,\alpha}(K) - \nu_{1,\alpha}(K)u}{\nu_{0,\alpha}(K)\nu_{2,\alpha}(K) - \nu_{1,\alpha}(K)^2}K(u)1_{(-1,\alpha)}(u),$$

where $\nu_{\ell,\alpha}(K) = \int_{-1}^{\alpha} u^\ell K(u)\,du$ and $0 \leq \alpha = t/h < 1$. Using K^L_α instead of K in (3.14) leads to bias of the order $\mathcal{O}(h^2)$ uniformly for $t = \alpha h$ close to the left boundary (see for instance Wand and Jones (1995)). For the right boundary we use

$$K^R_\beta(u) = \frac{\nu_{2,\beta}(K) + \nu_{1,\beta}(K)u}{\nu_{0,\beta}(K)\nu_{2,\beta}(K) - \nu_{1,\beta}(K)^2}K(u)1_{(-\beta,1)}(u),$$

where $0 \leq \beta = (1-t)/h < 1$. Now define

$$K_t(u) = \begin{cases} K^L_{t/h}(u) & ,0 \leq t < h \\ K(u) & ,h \leq t \leq 1-h \\ K^R_{(1-t)/h}(u) & ,1-h < t \leq 1. \end{cases} \tag{3.16}$$

Then the bias of the corresponding kernel estimator

$$\hat{f}_{B,n,h}(t) = \frac{1}{h}\int K_t\left(\frac{t-x}{h}\right)\,dF_n(x), \tag{3.17}$$

is of the order $\mathcal{O}(h^2)$ everywhere. Let $K'_t(u) = dK_t(u)/du$ and define the corresponding kernel estimator for f' by

$$\hat{f}'_{B,n,h}(t) = \frac{1}{h^2}\int K'_t\left(\frac{t-x}{h}\right)\,dF_n(x) - \frac{1}{h}K_t\left(\frac{t}{h}\right)\hat{f}_{B,n,h}(0) - \frac{1}{h}K_t\left(\frac{t-1}{h}\right)\hat{f}_{B,n,h}(1). \tag{3.18}$$

The last two terms are subtracted in order to keep the bias of the order $\mathcal{O}(h)$. Using standard arguments from the theory of kernel estimators, it can be shown that if we take $\hat{f}_n = \hat{f}_{B,n,h}$ and $\hat{f}'_n = \hat{f}'_{B,n,h}$ as defined in (3.16), (3.17) and (3.18), where K satisfies (K1)-(K3) and $h = \mathcal{O}(n^{-1/5})$, then the conditions (i)-(ii) of Theorem 3.3 are satisfied. See Lemmas 5.13 and 5.14.

Remark 3.1 To show that the kernel estimators, as described above, satisfy conditions (i)-(ii) of Theorem 3.3, the condition (A1) on f can be relaxed somewhat. At the cost of some additional technicalities, conditions (i)-(ii) can be shown to hold if we require $|f'(x)-f'(y)| \leq C|x-y|^{3/4+\epsilon}$, for some $\epsilon > 0$ and $C > 0$ not depending on f.

Theorem 3.3 suggests that in order to test the null hypothesis that f is non-increasing at level α, we reject if $T_n > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution. It is easy to see that this test is not a test of level α. The probability of committing a type I error under the uniform distribution may only be bounded asymptotically by α when f' is estimated at rate $n^{1/2}$, which is far too restrictive. In that case rejection criteria must be corrected. This is also the case for underlying densities that are strictly decreasing, but nearly flat. However, this deficiency of T_n is completely made up by its high power on increasing alternatives very close to the uniform distribution. In order to have a test of level α one can use statistics R_n or S_n. They have smaller power close to the uniform, but the probability of committing a type I error does not exceed level α. A comparison of T_n with R_n and S_n can be found in Section 4. The use of T_n is advisable when the price of a false alarm is not high compared to the price of overlooking violation of f being non-increasing. This may be the case, for example, when testing the null hypothesis of a monotone intensity of a Poisson process describing disease cases or fallouts. Although the Poisson model is different from the one considered here, testing for a monotone intensity is equivalent to testing for a monotone density (see also Woodroofe and Sun (1999)).

4 A simulation study

We have done a small simulation study in order to compare the different test statistics and to illustrate their advantages and disadvantages. For the underlying distribution we have chosen a family of exponential distributions restricted to the interval $[0,1]$ with densities

$$f_\theta(x) = \frac{\theta e^{\theta x}}{e^\theta - 1} \quad \text{for } x \in [0,1], \tag{4.19}$$

where $\theta \in I\!\!R$. The value $\theta = 0$ corresponds to the uniform density. It is a transition point, in the sense that for $\theta < 0$ (or $\theta > 0$), the density is strictly decreasing (or increasing). We perform the simulation for $\theta = 0, \pm 1/20, \pm 1/10, \pm 1/5, \pm 1/3$ and ± 1. Note that for smaller $|\theta|$, the density is closer to the uniform. The case $\theta = 1$ was also considered by Woodroofe and Sun (1999).

We have simulated one thousand values of S_n, R_n and T_n, as defined in (3.9), (3.10) and (3.11), with $k = 1$. The constants depending on ζ in (3.12) and (3.13) have been determined by a separate simulation, where we found $E[\zeta(0)] \approx 0.521$ and $\int_0^\infty \mathrm{cov}(\zeta(0), \zeta(s))\, ds \approx 0.0171$. For $\hat{f}$ and $\hat{f}'$ we used kernel estimators $\hat{f}_{B,n,k}$ and $\hat{f}'_{B,n,h}$ as defined (3.17) and (3.18). The boundary kernel K_t, as defined in (3.16), was constructed with the triweight kernel $K(u) = \frac{35}{32}(1-u^2)^3 1_{\{|u|<1\}}$. Since, we modify both at the boundaries 0 and 1, the bandwidth should never exceed 1/2. For this reason we took $h = \min(C_1(f)C_2(K)n^{-1/5}, 1/2)$, where the constant $C_1(f)$ was estimated by using an exponential reference family.

In Table 2 we report simulated probabilities of rejecting the null hypothesis that f is non-increasing at significance level 0.05, for underlying density f_θ as defined in (4.19), and for sample sizes $n = 100, 1000$ and 10000. The probabilities are simulated for the values of θ

specified above. Critical values for R_n and S_n were taken from Table 1. For T_n we used the 95% percentile of the standard normal distribution.

The test based on T_n is very powerful, rejecting the null hypothesis already for small positive θ's. It is not a test of level α since for small negative θ, it rejects the null hypothesis too often. However, at these decreasing, but almost flat underlying densities, the probability of committing a type I error disappears rapidly with decreasing θ. Test statistics R_n and S_n do produce tests of level α. However, the test based on R_n is less powerful than T_n. The test based on statistic S_n appears to be least powerful.

We also compared the power of T_n to that of the tests proposed by WOODROOFE AND SUN (1999) (see their Table 3). At the alternative f_θ, with $\theta = 1$, it appears that T_n is preferable to their D-and P-test for reasonably large samples. For $n = 100$, we see that the D-test has power 0.838 and the P-test 0.787, whereas T_n has power 0.960. Furthermore, we also have simulated the power of T_n at underlying density $f(x) = 2x$. The comparison with the D- and P-test is given in Table 3.

5 Proofs

We first show that for $J = E, W$, a properly scaled version of F_n^J can be approximated by the process

$$Y_{nt}(s) = n^{1/6}\Big(W_n\big(F(t+n^{-1/3}s)\big) - W_n\big(F(t)\big)\Big) + \frac{1}{2}f'(t)s^2, \quad \text{for } -\infty < s < \infty. \tag{5.20}$$

plus linear term, where W_n is defined in (2.6).

Lemma 5.1 *Suppose that f satisfies conditions (A1)-(A3). Let $F_n^E = F_n$ and let F_n^W be defined as in (2.7). Then for $t \in (0,1)$ fixed, $J = E, W$ and $s \in [-tn^{1/3}, (1-t)n^{1/3}]$:*

$$n^{2/3}F_n^J(t+n^{-1/3}s) = Y_{nt}(s) + L_{nt}^J(s) + R_{nt}^J(s),$$

Table 2: Simulated probabilities of rejection at $\alpha = 0.05$.

θ	-1	-0.33	-0.2	-0.1	-0.05	0	0.05	0.1	0.2	0.33	1
						$n = 100$					
T_n	0.013	0.091	0.114	0.181	0.214	0.241	0.285	0.323	0.413	0.539	0.960
R_n	0.000	0.000	0.007	0.016	0.020	0.034	0.043	0.051	0.075	0.161	0.771
S_n	0.000	0.002	0.008	0.013	0.022	0.026	0.034	0.038	0.060	0.111	0.625
						$n = 1000$					
T_n	0.006	0.025	0.042	0.136	0.191	0.286	0.396	0.559	0.809	0.974	1.000
R_n	0.000	0.000	0.001	0.008	0.015	0.055	0.109	0.215	0.481	0.871	1.000
S_n	0.000	0.000	0.004	0.012	0.022	0.041	0.091	0.163	0.376	0.779	1.000
						$n = 10\,000$					
T_n	0.005	0.006	0.015	0.027	0.079	0.282	0.739	0.965	1.000	1.000	1.000
R_n	0.000	0.000	0.000	0.000	0.000	0.054	0.360	0.859	1.000	1.000	1.000
S_n	0.000	0.000	0.000	0.000	0.004	0.059	0.268	0.753	0.999	1.000	1.000

Table 3: Power of T_n compared to the D- and P-test.

n	D-test	P-test	T_n
10	0.542	0.588	0.496
20	0.835	0.875	0.906
40	0.984	0.992	0.995

where Y_{nt} is defined in (5.20), $L_{nt}^J(s)$ is linear in s, and where for all $k \geq 1$,

$$E \sup_{|s|\leq \log n} \left|R_{nt}^J(s)\right|^k = \mathcal{O}(n^{-k/3}(\log n)^{3k}),$$

uniformly in $t \in (0,1)$.

Proof: Taylor expansion together with (2.7) and (2.6) yields that

$$n^{2/3}F_n^W(t+n^{-1/3}s) = Y_{nt}(s) + L_{nt}^W(s) + R_{nt}^W(s),$$

with Y_{nt} as defined in (5.20), $L_{nt}^W(s)$ is linear in s:

$$L_{nt}^W(s) = n^{2/3}F(t) + n^{1/6}W_n(F(t)) + n^{1/3}f(t)s,$$

and $R_{nt}^W(s) = \frac{1}{6}n^{-1/3}f''(\theta_1)s^3$, for some $|\theta_1 - t| \leq n^{-1/3}|s|$. Similarly

$$\begin{aligned} n^{2/3}F_n^E(t+n^{-1/3}s) &= n^{2/3}F_n^W(t+n^{-1/3}s) \\ &\quad + n^{1/6}\left\{E_n(t+n^{-1/3}s) - B_n(F(t+n^{-1/3}s))\right\} \\ &\quad - n^{1/6}\xi_n\left\{F(t) + f(t)n^{-1/3}s + \frac{1}{2}f'(\theta_2)n^{-2/3}s^2\right\} \\ &= Y_{nt}(s) + L_{nt}^E(s) + R_{nt}^E(s), \end{aligned}$$

where $L_{nt}^E(s) = L_{nt}^W(s) - n^{1/6}\xi_n F(t) - n^{-1/6}\xi_n f(t)s$ is linear in s, and

$$R_{nt}^E(s) = R_{nt}^W(s) - n^{1/6}\left\{E_n(t+n^{-1/3}s) - B_n(F(t+n^{-1/3}s))\right\} - \frac{1}{2}n^{-1/2}\xi_n f'(\theta_2)s^2,$$

for some $|\theta_2 - t| \leq n^{-1/3}|s|$. It follows immediately from conditions (A1)-(A3) that:

$$\sup_{|s|\leq \log n} \left|R_{nt}^W(s)\right|^k \leq C_1 n^{-k/3}(\log n)^{3k}. \tag{5.21}$$

Note that

$$\sup_{|s|\leq \log n} \left|R_{nt}^E(s)\right| \leq \sup_{|s|\leq \log n} \left|R_{nt}^W(s)\right| + n^{1/6}S_n + \frac{1}{2}\sup|f'|n^{-1/2}(\log n)^2|\xi_n|,$$

where $S_n = \sup_{s\in\mathbb{R}} |E_n(s) - B_n(F(s))|$. From Kómlos, Major and Tusnády (1975) we have that

$$P\left\{S_n \geq n^{-1/2}(C\log n + x)\right\} \leq Ke^{-\lambda x},$$

for positive constants C, K, and λ. This implies that for all $k \geq 1$,

$$ES_n^k = \mathcal{O}(n^{-k/2}(\log n)^k). \tag{5.22}$$

Next use that for all $a, b > 0$ and $k \geq 1$

$$(a+b)^k \leq 2^k(a^k + b^k). \tag{5.23}$$

Then from conditions (A1)-(A3) together with (5.22) and (5.21) we find that

$$\begin{aligned} E \sup_{|s| \leq \log n} \left| R_{nt}^E(s) \right|^k &= \mathcal{O}\left(n^{-k/3}(\log n)^{3k}\right) + \mathcal{O}\left(n^{-k/3}(\log n)^{k}\right) + \mathcal{O}\left(n^{-k/2}(\log n)^{2k}\right) \\ &= \mathcal{O}\left(n^{-k/3}(\log n)^{3k}\right). \end{aligned}$$

This proves the lemma. ■

The next step is to approximate the moments of $A_n^J(t)$ by corresponding moments of the concave majorant of the process Y_{nt}. For this we need to show that the concave majorants of F_n^J on [0,1], and on a neighborhood of t are equal at t. This requires some results from Kulikov and Lopuhaä (2003). They are listed below for easy reference.

Lemma 5.2 *Let g and h be functions on an interval $B \subset \mathbb{R}$. Then*

1. *For any linear function $l(t) = at + b$ on B, we have $[\mathrm{CM}_B(g+l)](t) = [\mathrm{CM}_B g](t) + l(t)$ for all $t \in B$.*
2. *$(\mathrm{CM}_B g) + \inf_B h \leq \mathrm{CM}_B(g+h) \leq (\mathrm{CM}_B g) + \sup_B h$ on B.*
3. *Let $[a,b] \subset B \subset \mathbb{R}$ and suppose that $[\mathrm{CM}_{[a,b]}g](x_1) = [\mathrm{CM}_B g](x_1)$ and $[\mathrm{CM}_{[a,b]}g](x_2) = [\mathrm{CM}_B g](x_2)$, for $x_1 < x_2$ in $[a,b]$. Then $[\mathrm{CM}_{[a,b]}g](t) = [\mathrm{CM}_B g](t)$, for all $t \in [x_1, x_2]$.*

Proof: See Kulikov and Lopuhaä (2003).

Lemma 5.3 *Let Z be defined in (2.2) and, for $d > 0$, let $N(d)$ be the event that $[\mathrm{CM}_{\mathbb{R}}Z](s)$ and $[\mathrm{CM}_{[-d,d]}Z](s)$ are equal for $s \in [-d/2, d/2]$. Then there exist constants $C_1 > 0$ and $C_2 > 0$, such that for all d sufficiently large*

$$P(N(d)^c) \leq C_1 \exp\left(-C_2 d^{3/2}\right).$$

Proof: See Kulikov and Lopuhaä (2003).

Lemma 5.4 *For $d > 0$, let $I_{nt}(d) = [0,1] \cap [t - dn^{-1/3}, t + dn^{-1/3}]$. For $J = E, W$, $t \in (0,1)$, let $N_{nt}^J(d)$ be the event that $[\mathrm{CM}_{[0,1]}F_n^J](s)$ and $[\mathrm{CM}_{I_{nt}(d)}F_n^J](s)$ are equal for $s \in I_{nt}(d/2)$. Then, for any distribution function F satisfying conditions (A1)-(A3),*

$$P\left\{N_{nt}^J(d)^c\right\} \leq 8 \exp\left(-Cd^3\right),$$

where $C > 0$ does not depend on d, t and n.

Proof: See Kulikov and Lopuhaä (2003).

Lemma 5.5 *Suppose that f satisfies conditions (A1)-(A3). For $t \in (0,1)$ fixed, let Y_{nt} be defined as in (5.20). Let $A_n^E(t) = A_n(t)$ and $A_n^W(t)$ be defined in (2.1) and (2.8). Define $H_{nt} = [-n^{1/3}t, n^{1/3}(1-t)] \cap [-\log n, \log n]$. Then for all $k \geq 1$, and for $J = E, W$*

$$EA_n^J(t)^k = E\left[\mathrm{D}_{H_{nt}} Y_{nt}\right](0)^k + o(n^{-1/6})$$

uniformly for $t \in (0,1)$.

Proof: Let $I_{nt}(d)$ and $N_{nt}^J(d)$ be defined as in Lemma 5.4, and write $I_{nt} = I_{nt}(\log n)$ and $N_{nt}^J = N_{nt}^J(\log n)$. Then by property 3 of Lemma 5.2, on the event N_{nt}^J, the concave majorants $\mathrm{CM}_{[0,1]} F_n^J$ and $\mathrm{CM}_{I_{nt}} F_n^J$ coincide on $[t - n^{-1/3}\log n/2, t + n^{-1/3}\log n/2]$. In particular, they coincide at t, so that

$$A_n^J(t) 1_{N_{nt}^J} = n^{2/3} [\mathrm{D}_{I_{nt}} F_n^J](t) 1_{N_{nt}^J} \quad \text{for } J = E, W. \tag{5.24}$$

By definition $|A_n^E(t)| \leq 2n^{2/3}$ and A_n^W is bounded by $2n^{2/3}\left(1 + n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right)$, so that

$$\begin{aligned} E\left|A_n^J(t)^k - n^{2k/3}[\mathrm{D}_{I_{nt}} F_n^J](t)^k\right| 1_{(N_{nt}^J)^c} &\leq 2^{k+1} n^{2k/3} E\left(1 + n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right)^k 1_{(N_{nt}^J)^c} \\ &\leq 2^{k+1} n^{2k/3} \left\{E\left(1 + n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right)^{2k}\right\}^{1/2} \left\{P\left((N_{nt}^J)^c\right)\right\}^{1/2}. \end{aligned}$$

Next, use (5.23) together with the fact that all moments of $\sup_{s\in[0,1]}|W_n(s)|$ are finite. Then it follows from Lemma 5.4 that

$$\begin{aligned} EA_n^J(t)^k &= n^{2k/3} E[\mathrm{D}_{I_{nt}} F_n^J](t)^k + E\left(A_n^J(t)^k - n^{2k/3}[\mathrm{D}_{I_{nt}} F_n^J](t)^k\right) 1_{(N_{nt}^J)^c} \\ &= n^{2k/3} E[\mathrm{D}_{I_{nt}} F_n^J](t)^k + n^{2k/3}\mathcal{O}(e^{-\frac{1}{2}C_2(\log n)^3}), \end{aligned}$$

uniformly for $t \in (0,1)$. From Lemmas 5.1 and 5.2, we have for $s \in H_{nt} = n^{1/3}(I_{nt} - t)$:

$$n^{2/3}\left[\mathrm{D}_{I_{nt}} F_n^J\right](t) = \left[\mathrm{D}_{H_{nt}}(Y_{nt} + R_{nt}^J)\right](0) = \left[\mathrm{D}_{H_{nt}} Y_{nt}\right](0) + \Delta_{nt},$$

where $\Delta_{nt} = \left[\mathrm{D}_{H_{nt}}(Y_{nt} + R_{nt}^J)\right](0) - \left[\mathrm{D}_{H_{nt}} Y_{nt}\right](0)$. We find that

$$EA_n^J(t)^k = E\left[\mathrm{D}_{H_{nt}} Y_{nt}\right](0)^k + \epsilon_{nt} + n^{2k/3}\mathcal{O}\left(e^{-\frac{1}{2}C_2(\log n)^3}\right), \tag{5.25}$$

where, by application of the mean value theorem,

$$|\epsilon_{nt}| \leq kE|\theta_{nt}|^{k-1}|\Delta_{nt}| \leq k\left\{E|\theta_{nt}|^{2k-2}\right\}^{1/2}\left\{E|\Delta_{nt}|^2\right\}^{1/2}, \tag{5.26}$$

with $|\theta_{nt} - [\mathrm{D}_{H_{nt}} Y_{nt}](0)| \leq |\Delta_{nt}|$. Since $H_{nt} \subset [-\log n, \log n]$, by application of (5.23)

$$E|\theta_{nt}|^{2k-2} \leq 4^{2k-2}\left(E\sup_{|s|\leq\log n}|Y_{nt}(s)|^{2k-2} + E|\Delta_{nt}|^{2k-2}\right), \tag{5.27}$$

where according to property 2 of Lemma 5.2 together with Lemma 5.1, for all $k \geq 1$

$$E|\Delta_{nt}|^k \leq 2^k E\sup_{|s|\leq\log n}|R_{nt}^J(s)|^k = \mathcal{O}\left(n^{-k/3}(\log n)^{3k}\right), \tag{5.28}$$

uniformly for $t \in (0,1)$. On the other hand, for $|s| \leq \log n$, there exist constants $C_3, C_4 > 0$ that only depend on f, such that

$$\sup_{|s|\leq \log n} |Y_{nt}(s)| \leq \sup_{|s|\leq C_3 \log n} |W_n(s)| + C_4(\log n)^2 \stackrel{d}{=} (C_3 \log n)^{1/2} \sup_{|s|\leq 1} |W(s)| + C_4(\log n)^2.$$

Because all moments of $\sup_{|s|\leq 1} |W(s)|$ are finite, from (5.26), (5.27) and (5.28) we conclude that $\epsilon_{nt} = \mathcal{O}\big(n^{-1/3}(\log n)^{2k+1}\big)$. Together with (5.25) this proves the lemma. ■

By uniform continuity of Brownian motion on compacta, the process $Y_{nt}(s)$ is close to the process $W_n(f(t)s) + \frac{1}{2}f'(t)s^2$. In view of Theorem 2.1 this difference must be of smaller order than $n^{-1/6}$. Unfortunately, it does not suffice to bound the difference between the concave majorants by

$$\sup_{|s|\leq \log n} \left| W_n\left(n^{1/3}\left(F(t+n^{-1/3}s) - F(t)\right)\right) - W_n\big(f(t)s\big)\right|,$$

because according to the properties of the modulus of continuity for Brownian motion, this is of order $\mathcal{O}(n^{-1/6}\log n)$. However, Lemma 5.7 follows from the next lemma that ensures that the two concave majorants at zero are sufficiently close. We only need this lemma for continuous g, but with a little more effort a similar result can be obtained for non-continuous g.

Lemma 5.6 *Let g be a continuous function on an interval $B \subset \mathbb{R}$. Let $0 \in B^\circ$ and let $\phi : \mathbb{R} \to \mathbb{R}$ be invertible with $\phi(0) = 0$. Let $\sup_B g < \infty$ and suppose there exists an $\alpha \in [0, 1/2]$ such that*

$$1 - \alpha \leq \frac{\phi(t)}{t} \leq 1 + \alpha, \tag{5.29}$$

for all $t \in B \backslash \{0\}$. Then

$$\left| \left[\mathrm{CM}_{\phi^{-1}(B)}(g\circ\phi)\right](0) - \left[\mathrm{CM}_B g\right](0)\right| \leq 4\alpha\left\{\sup_B g - \left[\mathrm{CM}_B g\right](0)\right\}.$$

Proof: Consider the function $h(t) = g(t) - \sup_B g$. For $a < b$, let $[a, b] \subset B$ be an interval containing zero. Then with property (5.29), t and $\phi(t)$ have the same sign. Hence, $\phi^{-1}(a) < \phi^{-1}(b)$ and $0 \in [\phi^{-1}(a), \phi^{-1}(b)]$. This yields the following inequality

$$\frac{1+\alpha}{1-\alpha}\cdot\frac{h(a)b - h(b)a}{b-a} \leq \frac{h(a)\phi^{-1}(b) - h(b)\phi^{-1}(a)}{\phi^{-1}(b) - \phi^{-1}(a)} \leq \frac{1-\alpha}{1+\alpha}\cdot\frac{h(a)b - h(b)a}{b-a}. \tag{5.30}$$

First assume that both $\mathrm{CM}_B h$ and $\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)$ have non-empty segments containing zero. Let $[\tau_1, \tau_2] \subset B$, with $\tau_1 < \tau_2$, be the segment of $\mathrm{CM}_B h$ that contains zero. Similarly, let $[\xi_1, \xi_2] \subset \phi^{-1}(B)$ be the segment of $\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)$ that contains zero, with $\xi_1 < \xi_2$. Denote $t_i = \phi^{-1}(\tau_i)$ and $x_i = \phi(\xi_i)$, for $i = 1, 2$, so that $t_1 < t_2$ and $x_1 < x_2$. Consider the line between $(x_1, h(x_1))$ and $(x_2, h(x_2))$. Since $[x_1, x_2] \subset B$, the intercept at zero of this line must be below $[\mathrm{CM}_B h](0)$:

$$\frac{h(x_1)x_2 - h(x_2)x_1}{x_2 - x_1} \leq [\mathrm{CM}_B h](0) = \frac{h(\tau_1)\tau_2 - h(\tau_2)\tau_1}{\tau_2 - \tau_1}. \tag{5.31}$$

Similarly, consider the line between $\big(t_1,(h\circ\phi)(t_1)\big)$ and $\big(t_2,(h\circ\phi)(t_2)\big)$. Since $[t_1,t_2]\subset\phi^{-1}(B)$, the intercept at zero of this line must be below $[\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0)$:

$$\frac{(h\circ\phi)(t_1)t_2-(h\circ\phi)(t_2)t_1}{t_2-t_1}\leq[\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0)=\frac{(h\circ\phi)(\xi_1)\xi_2-(h\circ\phi)(\xi_2)\xi_1}{\xi_2-\xi_1},$$

or equivalently,

$$\frac{h(\tau_1)\phi^{-1}(\tau_2)-h(\tau_2)\phi^{-1}(\tau_1)}{\phi^{-1}(\tau_2)-\phi^{-1}(\tau_1)}\leq[\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0)=\frac{h(x_1)\phi^{-1}(x_2)-h(x_2)\phi^{-1}(x_1)}{\phi^{-1}(x_2)-\phi^{-1}(x_1)}.$$

Together with (5.31) and (5.30), this implies that

$$\begin{aligned}\frac{1+\alpha}{1-\alpha}[\mathrm{CM}_B h](0) &\leq \frac{h(\tau_1)\phi^{-1}(\tau_2)-h(\tau_2)\phi^{-1}(\tau_1)}{\phi^{-1}(\tau_2)-\phi^{-1}(\tau_1)}\leq[\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0)\\ &\leq \frac{1-\alpha}{1+\alpha}\cdot\frac{h(x_1)x_2-h(x_2)x_1}{x_2-x_1}\leq\frac{1-\alpha}{1+\alpha}[\mathrm{CM}_B h](0).\end{aligned}$$

Now use that $(1-\alpha)/(1+\alpha)\geq 1-4\alpha$ and $(1+\alpha)/(1-\alpha)\leq 1+4\alpha$, for $\alpha\in[0,1/2]$ and apply property 2 of Lemma 5.2 to the function $g(t)=h(t)+\sup_B g$. This finishes the proof for the case that both $\mathrm{CM}_B h$ and $\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)$ have non-empty segments containing zero.

If this is not the case, for $\epsilon>0$ sufficiently small, such that $[-\epsilon,\epsilon]\subset B$, define

$$g_\epsilon(t)=\begin{cases}g(t) & \text{if } t\in B\setminus[-\epsilon,\epsilon]\\ g(0)+(g(0)-g(-\epsilon))t/\epsilon & \text{if } t\in[-\epsilon,0]\\ g(0)+(g(\epsilon)-g(0))\phi^{-1}(t)/\phi^{-1}(\epsilon) & \text{if } t\in[0,\epsilon].\end{cases}$$

Then g_ϵ is continuous and linear on $[-\epsilon,0]$ and the function $g_\epsilon\circ\phi$ is linear on $[0,\phi^{-1}(\epsilon)]$. This implies that for the corresponding function $h_\epsilon=g_\epsilon-\sup_B g_\epsilon$, both $\mathrm{CM}_B h$ and $\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)$ have non-empty segments containing zero. Next, let $\delta>0$ arbitrary and choose $\epsilon>0$ sufficiently small such that $\sup|g_\epsilon-g|\leq\delta$. Then, again by property 2 of Lemma 5.2 and by (5.29), it follows that $|[\mathrm{CM}_B g](0)-[\mathrm{CM}_B g_\epsilon](0)|\leq\sup_{t\in[-\epsilon,\epsilon]}|g(t)-g_\epsilon(t)|\leq\delta$, and similarly

$$\begin{aligned}\left|\left[\mathrm{CM}_{\phi^{-1}(B)}(g\circ\phi)\right](0)-\left[\mathrm{CM}_{\phi^{-1}(B)}(g_\epsilon\circ\phi)\right](0)\right| &\leq \sup_{t\in[\phi^{-1}(-\epsilon),\phi^{-1}(\epsilon)]}|(g\circ\phi)(t)-(g_\epsilon\circ\phi)(t)|\\ &= \sup_{t\in[-\epsilon,\epsilon]}|g(t)-g_\epsilon(t)|\leq\delta,\end{aligned}$$

where $\delta>0$ can be chosen arbitrarily small. ■

Lemma 5.7 *Suppose that f satisfies conditions (A1)-(A3). Let $t\in(0,1)$ and let ζ be defined as in (2.3). Let $A_n^E(t)=A_n(t)$ and $A_n^W(t)$ be defined in (2.1) and (2.8). Then for all $k\geq 1$, and for $J=E,W$,*

$$EA_n^J(t)^k=\left(\frac{2f(t)^2}{|f'(t)|}\right)^{k/3}E\zeta(0)^k+o(n^{-1/6}),$$

uniformly in $t\in(n^{-1/3}\log n,1-n^{-1/3}\log n)$, and

$$EA_n^J(t)^k\leq\left(\frac{2f(t)^2}{|f'(t)|}\right)^{k/3}E\zeta(0)^k+o(n^{-1/6}),$$

uniformly in $t\in(0,1)$.

Proof: For $t \in (0,1)$ fixed let Y_{nt} be defined as in (5.20) and let

$$Z_{nt}(s) = W_n(f(t)s) + \frac{1}{2}f'(t)s^2. \tag{5.32}$$

Let $a_{nt} = \max(0, t - n^{-1/3}\log n)$ and $b_{nt} = \min(1, t + n^{-1/3}\log n)$. Define the interval $J_{nt} = \left[n^{1/3}\left(F(a_{nt}) - F(t)\right)/f(t),\, n^{1/3}\left(F(b_{nt}) - F(t)\right)/f(t)\right]$ and the mapping

$$\phi_{nt}(s) = \frac{n^{1/3}\big(F(t+n^{-1/3}s) - F(t)\big)}{f(t)}.$$

Then $H_{nt} = \phi_{nt}^{-1}(J_{nt}) = [n^{1/3}(a_{nt}-t), n^{1/3}(b_{nt}-t)]$, and there exists a constant $C_1 > 0$ only depending on f, such that for all $s \in H_{nt}$, we have $1 - \alpha_n \le \phi_{nt}(s)/s \le 1 + \alpha_n$, where $\alpha_n = C_1 n^{-1/3}\log n$. By definition of Z_{nt} and Y_{nt},

$$(Z_{nt} \circ \phi_{nt})(s) = Y_{nt}(s) + \frac{1}{2}f'(t)s^2\left(\frac{\phi_{nt}(s)^2}{s^2} - 1\right).$$

Since $H_{nt} \subset [-\log n, \log n]$, according to property 2 of Lemma 5.2, there exists constant $C_2 > 0$ only depending on f, such that

$$\big|[\mathrm{D}_{H_{nt}}Y_{nt}](0) - [\mathrm{D}_{H_{nt}}(Z_{nt}\circ\phi_{nt})](0)\big| \le C_2 n^{-1/3}(\log n)^3. \tag{5.33}$$

Now apply Lemma 5.6 with $g = Z_{nt}$, $\phi = \phi_{nt}$, $\alpha = \alpha_n$ and $B = J_{nt}$. This yields that

$$\big|[\mathrm{D}_{H_{nt}}(Z_{nt}\circ\phi_{nt})](0) - [\mathrm{D}_{J_{nt}}Z_{nt}](0)\big| \le 8\alpha_n \sup_{s\in\mathbb{R}}|Z_{nt}(s)|.$$

Together with (5.33) we conclude that there exists a constant $C > 0$ only depending on f, such that

$$|[\mathrm{D}_{H_{nt}}Y_{nt}](0) - [\mathrm{D}_{J_{nt}}Z_{nt}](0)| \le Cn^{-1/3}\log n\left((\log n)^2 + \sup_{s\in\mathbb{R}}|Z_{nt}(s)|\right). \tag{5.34}$$

Similar to the proof of Lemma 5.5, this implies that

$$E[\mathrm{D}_{H_{nt}}Y_{nt}](0)^k = E[\mathrm{D}_{J_{nt}}Z_{nt}](0)^k + \epsilon_{nt}, \tag{5.35}$$

where $|\epsilon_{nt}| \le k\left\{E|\theta_{nt}|^{2k-2}\right\}^{1/2}\left\{E|\Delta_{nt}|^2\right\}^{1/2}$, with $\Delta_{nt} = [\mathrm{D}_{H_{nt}}Y_{nt}](0) - [\mathrm{D}_{J_{nt}}Z_{nt}](0)$ and $|\theta_{nt} - [\mathrm{D}_{H_{nt}}Y_{nt}](0)| \le |\Delta_{nt}|$. Note that with $c_1(t)$ and $c_2(t)$ as defined in (2.5), by Brownian scaling one has

$$c_1(t)Z_{nt}(c_2(t)s) \overset{d}{=} Z(s). \tag{5.36}$$

Since $P\{\sup_{t\in\mathbb{R}}(W(t) - t^2) > x\} \le 4\exp(-x^{3/2}/2)$ (see for instance Lemma 3.3 in Kulikov and Lopuhaä (2003)), it follows that for all $k \ge 1$,

$$E\left(\sup_{s\in\mathbb{R}}|Z_{nt}(s)|\right)^k \le CE\left(\sup_{s\in\mathbb{R}}|Z(s)|\right)^k < \infty,$$

for a constant $C > 0$ only depending on f. From (5.34) we conclude that for all $k \ge 1$

$$E|\Delta_{nt}|^k = \mathcal{O}(n^{-k/3}(\log n)^{3k}). \tag{5.37}$$

Similar to the proof of Lemma 5.5, using an inequality similar to (5.27) together with (5.37), we find that $\epsilon_{nt} = \mathcal{O}(n^{-1/3}(\log n)^{2k+1})$, so that from (5.35) we get

$$E[\mathrm{D}_{H_{nt}} Y_{nt}](0)^k = E[\mathrm{D}_{J_{nt}} Z_{nt}](0)^k + \mathcal{O}(n^{-1/3}(\log n)^{2k+1}).$$

Together with Lemma 5.5 and scaling property (5.36), we find that

$$EA_n^J(t)^k = c_1(t)^{-k} E\left[\mathrm{D}_{I\!R} Z\right](0)^k + c_1(t)^{-k} E\left(\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{I\!R} Z\right](0)^k\right) + o(n^{-1/6}), \quad (5.38)$$

where $I_{nt} = c_2(t)^{-1} J_{nt}$. First note that for any $t \in (0,1)$, on the interval I_{nt}, the concave majorant $\mathrm{CM}_{I_{nt}} Z$ always lies below $\mathrm{CM}_{I\!R} Z$. Because I_{nt} contains 0, this implies that

$$EA_n^J(t)^k \le c_1(t)^{-k} E\left[\mathrm{D}_{I\!R} Z\right](0)^k + o(n^{-1/6}),$$

uniformly for $t \in (0,1)$.

When $t \in (n^{-1/3}\log n, 1 - n^{-1/3}\log n)$, there exist an $M > 0$, only depending on f, such that $[-M\log n, M\log n] \subset I_{nt}$. Note that on the interval $[-M\log n, M\log n]$ we always have $\mathrm{CM}_{[-M\log n, M\log n]} Z \le \mathrm{CM}_{I_{nt}} Z \le \mathrm{CM}_{I\!R} Z$. Write $N_{nM} = N(M\log n)$, with $N(d)$ as defined in Lemma 5.3. On the event N_{nM}, we have $\left[\mathrm{CM}_{[-M\log n, M\log n]} Z\right](0) = \left[\mathrm{CM}_{I_{nt}} Z\right](0) = \left[\mathrm{CM}_{I\!R} Z\right](0)$. Hence

$$\begin{aligned} \left| E\left(\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{I\!R} Z\right](0)^k\right)\right| &\le E\left|\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{I\!R} Z\right](0)^k\right| 1_{N_{nM}^c} \\ &\le 2^{k+1} E\left(\sup_{s\in I\!R} |Z(s)|\right)^k 1_{N_{nM}^c} \le 2^{k+1}\left\{E\left(\sup_{s\in I\!R}|Z(s)|\right)^{2k}\right\}^{1/2} \left\{P(N_{nM}^c)\right\}^{1/2}. \end{aligned}$$

Since $E(\sup|Z|)^{2k} < \infty$, it follows from Lemma 5.3 that $E\left(\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{I\!R} Z\right](0)^k\right) = o(n^{-1/6})$. Together with (5.38) and the fact that $\zeta = \mathrm{D}_{I\!R} Z$ this proves the lemma. ∎

Lemma 5.8 *Suppose that f satisfies conditions (A1)-(A3). Let $A_n^E = A_n$ and A_n^W be defined by (2.1) and (2.8). Then for all $k \ge 1$, we have $\int_0^1 \left|A_n^E(t)^k - A_n^W(t)^k\right| dt = o_p(n^{-1/6})$.*

Proof: By Markov's inequality is suffices to prove that $E\left|A_n^E(t)^k - A_n^W(t)^k\right| = o(n^{-1/6})$ uniformly in $t \in (0,1)$. Let $I_{nt}(d)$ and $N_{nt}^J(d)$ be defined as in Lemma 5.4. Write $I_{nt} = I_{nt}(\log n)$ and $N_{nt}^J = N_{nt}^J(\log n)$, and let $K_{nt} = N_{nt}^E \cap N_{nt}^W$. Then according to (5.24):

$$\begin{aligned} E\left|A_n^E(t)^k - A_n^W(t)^k\right| &= n^{2k/3} E\left|[D_{I_{nt}} F_n^E](t)^k - [D_{I_{nt}} F_n^W](t)^k\right| 1_{K_{nt}} \\ &\quad + E\left|A_n^E(t)^k - A_n^W(t)^k\right| 1_{K_{nt}^c}. \end{aligned} \quad (5.39)$$

We first bound the second expectation the right hand side of (5.39). We have that

$$\begin{aligned} E\left|A_n^E(t)^k - A_n^W(t)^k\right| 1_{K_{nt}^c} &\le EA_n^E(t)^k 1_{K_{nt}^c} + EA_n^W(t)^k 1_{K_{nt}^c} \\ &\le \left\{EA_n^E(t)^{2k}\right\}^{1/2} \left\{P(K_{nt}^c)\right\}^{1/2} + \left\{EA_n^W(t)^{2k}\right\}^{1/2}\left\{P(K_{nt}^c)\right\}^{1/2}, \end{aligned}$$

where, according to Lemma 5.4, $P(K_{nt}^c) \le 16 e^{-C(\log n)^3}$ uniformly in $t \in (0,1)$. Since from Lemma 5.7 we know that $EA_n^J(t)^{2k}$ are bounded uniformly in n and $t \in (0,1)$, we conclude that

$$E\left|A_n^E(t)^k - A_n^W(t)^k\right| 1_{K_{nt}^c} = \mathcal{O}(e^{-\frac{1}{2}C(\log n)^3}), \quad (5.40)$$

uniformly in $t \in (0,1)$.

To bound the first expectation in (5.39), apply the mean value theorem to write

$$\begin{aligned} & n^{2k/3}\left|[D_{I_{nt}}F_n^E](t)^k - [D_{I_{nt}}F_n^W](t)^k\right| 1_{K_{nt}} \\ & \leq k|\theta_{nt}|^{k-1}n^{2/3}\left|[D_{I_{nt}}F_n^E](t) - [D_{I_{nt}}F_n^W](t)\right| 1_{K_{nt}} \\ & \leq k\left(A_n^E(t)^{k-1} + A_n^W(t)^{k-1}\right) n^{2/3}\left|[D_{I_{nt}}F_n^E](t) - [D_{I_{nt}}F_n^W](t)\right|. \end{aligned} \tag{5.41}$$

By Lemmas 5.1 and 5.2, $n^{2/3}\left|[D_{I_{nt}}F_n^E](t) - [D_{I_{nt}}F_n^W](t)\right| \leq |R_{nt}^E(0)| + |R_{nt}^W(0)|$. Hence, together with (5.41), the first expectation in (5.39) can be bounded by

$$k\left\{E\left(A_n^E(t)^{k-1} + A_n^W(t)^{k-1}\right)^2\right\}^{1/2}\left\{E\left(|R_{nt}^E(0)| + |R_{nt}^W(0)|\right)^2\right\}^{1/2},$$

From Lemma 5.7 together with (5.23), it follows that the first expectation is bounded uniformly for $t \in (0,1)$. According to Lemma 5.1, the second expectation is of the order $\mathcal{O}(n^{-1/3}(\log n)^3)$. Together with (5.40) this proves the lemma. ∎

Lemma 5.9 *Suppose that f satisfies conditions (A1)-(A3). The process $\{A_n^W(t) : t \in [0,1]\}$ is strong mixing process with mixing function*

$$\alpha_n(d) = 12e^{-C_1 n d^3},$$

where $C_1 > 0$ only depends on f. More specifically, for $d > 0$,

$$\sup|P(A \cap B) - P(A)P(B)| \leq \alpha_n(d),$$

where the supremum is taken over all sets $A \in \sigma\{A_n^W(s) : 0 < s \leq t\}$ and $B \in \sigma\{A_n^W(u) : t + d \leq u < 1\}$.

Proof: Let $t \in (0,1)$ arbitrary and take $0 < s_1 \leq s_2 \leq \cdots \leq s_k = t < t + d = u_1 \leq u_2 \leq \cdots \leq u_l < 1$. Consider events

$$\begin{aligned} E_1 &= \left\{A_n^W(s_1) \in B_1, \ldots, A_n^W(s_k) \in B_k\right\}, \\ E_2 &= \left\{A_n^W(u_1) \in C_1, \ldots, A_n^W(u_l) \in C_l\right\}, \end{aligned}$$

for Borel sets $B_1, \ldots, B_k$ and $C_1, \ldots, C_l$ of $\mathbb{R}$. Note that cylinder sets of the form E_1 and E_2 generate the σ-algebras $\sigma\{A_n^W(s) : 0 < s \leq t\}$ and $\sigma\{A_n^W(u) : t + d \leq u < 1\}$, respectively. Define the event

$$\begin{aligned} S = \Big\{ & \left[CM_{[0,1]}F_n^W\right](u) = \left[CM_{[0,t+d/2]}F_n^W\right](u) \text{ for any } u \in [0,t] \\ & \text{and } \left[CM_{[0,1]}F_n^W\right](u) = \left[CM_{[t+d/2,1]}F_n^W\right](u) \text{ for any } u \in [t+d,1]\Big\}. \end{aligned}$$

Let $E_1' = E_1 \cap S$ and $E_2' = E_2 \cap S$. Then by independency of the increments of the process F_n^W, the events E_1' and E_2' are independent. Therefore by means of Lemma 5.4

$$|P(E_1 \cap E_2) - P(E_1)P(E_2)| \leq 3P\left(S^c\right) \leq 48e^{-Cd^3 n}$$

for some constant $C > 0$ that only depends on f. This proves the lemma. ∎

From Lemmas 5.8 and 5.7 it follows immediately that for proving asymptotic normality of $n^{1/6}\int_0^1 \left(A_n(t)^k - EA_n(t)^k\right) g(t)\,dt$, it suffices to prove that its Brownian version

$$T_n^W = n^{1/6}\int_0^1 \left(A_n^W(t)^k - EA_n^W(t)^k\right) g(t)\,dt, \tag{5.42}$$

is asymptotically normal. The proof runs along the lines of the proof of Theorem 4.1 in Groeneboom, Hooghiemstra and Lopuhaä (1999) and needs two lemmas that bound covariances by the mixing coefficient. The lemmas are analogous to Theorems 17.2.1 and 17.2.2 in Ibragimov and Linnik (1971) and can be proven similarly, since stationarity is not essential in these theorems.

Lemma 5.10 *If X is measurable with respect to $\{\sigma\{A_n^W(s) : 0 < s \leq t\}$ and Y is measurable with respect to $\{\sigma\{A_n^W(u) : t + d \leq u < 1\}$ $(d > 0)$, and if $|X| \leq C_1$ and $|Y| \leq C_2$ a.s., then*

$$|E(XY) - E(X)E(Y)| \leq 4C_1C_2\alpha_n(d).$$

Lemma 5.11 *If X is measurable with respect to $\{\sigma\{A_n^W(s) : 0 < s \leq t\}$ and Y is measurable with respect to $\{\sigma\{A_n^W(s) : t + d \leq u < 1\}$ $(d > 0)$, and if for some $\delta > 0$,*

$$E|X|^{2+\delta} \leq C_3 \quad \text{and} \quad E|Y|^{2+\delta} \leq C_4,$$

then

$$|E(XY) - E(X)E(Y)| \leq C_5 \left(\alpha_n(d)\right)^{\delta/(2+\delta)},$$

where $C_5 > 0$ only depends on C_3 and C_4.

We first derive the asymptotic variance of T_n^W. To this end we introduce the Brownian version of the process ζ_{nt} defined in (2.4). For $t \in (0, 1)$ fixed and $t + c_2(t)sn^{-1/3} \in (0, 1)$,

$$\zeta_{nt}^W(s) = c_1(t)A_n^W(t + c_2(t)sn^{-1/3}), \tag{5.43}$$

where A_n^W is defined in (2.8) and $c_1(t)$ and $c_2(t)$ are defined in (2.5). From the fact that ζ_{nt} converges to ζ in distribution (see Theorem 4.1 in Kulikov and Lopuhaä (2003)) and Lemma 5.8, it follows immediately that the process

$$\{\zeta_{nt}^W(s) : s \in I\!R\} \to \{\zeta(s) : s \in I\!R\} \text{ in distribution.} \tag{5.44}$$

Furthermore, note that Lemma 5.7 implies that for every $m = 1, 2, \ldots$ there exists a constant $M > 0$ such that $EA_n^W(t)^{km} < M$, uniformly in $n = 1, 2 \ldots$ and $t \in (0, 1)$. Hence it follows from Markov's inequality, that for all $m = 1, 2, \ldots$ there exists a constant $M' > 0$

$$P\{|\zeta_{nt}^W(s)|^k > y\} \leq \frac{M'}{y^m},$$

uniformly in $n = 1, 2 \ldots$, $t \in (0, 1)$ and $t + c_2(t)sn^{-1/3} \in (0, 1)$. This guarantees uniform integrability of the sequence $\zeta_{nt}^W(s)^k$ for s, t and k fixed, so that together with (5.44) it implies convergence of moments of $(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k)$ to the corresponding moments of $(\zeta(0)^k, \zeta(s)^k)$. This leads to the following lemma.

Lemma 5.12 *Suppose that f satisfies conditions (A1)-(A3). Then for any function g that is continuous on $[0,1]$, and any $k \geq 1$,*

$$\operatorname{var}\left(n^{1/6}\int_0^1 A_n^W(t)^k g(t)dx\right) \to \int_0^1 \frac{2^{(2k+5)/3} f(t)^{(4k+1)/3}}{|f'(t)|^{(2k+2)/3}} g(t)^2\,dt \int_0^\infty \operatorname{cov}(\zeta(0)^k, \zeta(s)^k)\,ds.$$

Proof: We have with ζ_{nt}^W as defined in (5.43),

$$\begin{aligned}
&\operatorname{var}\left(n^{1/6}\int_0^1 A_n^W(t)^k g(t)dx\right)\\
&= 2n^{1/3}\int_0^1\int_u^1 \operatorname{cov}\left(A_n^W(t)^k, A_n^W(u)^k\right) g(t)g(u)\,dt\,du\\
&= 2\int_0^1 \frac{c_2(t)}{c_1(t)^{2k}}\int_0^{n^{1/3}(1-t)/c_2(t)} \operatorname{cov}(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k) g(t)g(t+c_2(t)sn^{-1/3})\,dt\,ds,
\end{aligned}$$

by change of variables of integration $u = t + c_2(t)sn^{-1/3}$. As noted above for s and t fixed,

$$\operatorname{cov}(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k) \to \operatorname{cov}(\zeta(0)^k, \zeta(s)^k).$$

Lemma 5.7 implies that uniformly in $n = 1, 2, \ldots$, s and t, we have $E|\zeta_{nt}^W(0)|^{3k} \leq C_3$ and $E|\zeta_{nt}^W(s)|^{3k} \leq C_4$. Hence by Lemma 5.11, it follows that

$$\operatorname{cov}(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k) \leq C_5\alpha_n\left(n^{-1/3}c_2(t)s\right)^{1/3} \leq D_1\exp(-D_2|s|^3),$$

where $D_1, D_2 > 0$ do not depend on n, s and t. Substituting $c_1(t)$, $c_2(t)$ as defined in (2.5), and using that g is uniformly bounded on $[0,1]$, it follows by dominated convergence that

$$\operatorname{var}\left(n^{1/6}\int_0^1 A_n^W(t)^k g(t)dx\right) \to \int_0^1 \frac{2^{(2k+5)/3} f(t)^{(4k+1)/3}}{|f'(t)|^{(2k+2)/3}} g(t)^2\,dt \int_0^\infty \operatorname{cov}(\zeta(0)^k, \zeta(s)^k\,ds. \quad \blacksquare$$

Proof of Theorem 2.1: It suffices to prove the statement for T_n^W as defined in (5.42). Define

$$X_n(t) \stackrel{def}{=} \left(A_n^W(t)^k - EA_n^W(t)^k\right) g(t).$$

Let

$$L_n = n^{-1/3}(\log n)^3, \quad M_n = n^{-1/3}\log n, \quad N_n = \left[\frac{1}{L_n + M_n}\right],$$

where $[x]$ denotes the integer part of x. We divide interval $[0,1]$ into blocks of alternating length

$$\begin{aligned}
A_j &= [(j-1)(L_n+M_n), (j-1)(L_n+M_n)+L_n],\\
B_j &= [(j-1)(L_n+M_n)+L_n, j(L_n+M_n)],
\end{aligned}$$

where $1 \leq j \leq N_n$. Now write $T_n^W = S_n' + S_n'' + R_n$, where

$$\begin{aligned}
S_n' &= n^{1/6}\sum_{j=1}^{N_n}\int_{A_j} X_n(t)\,dt,\\
S_n'' &= n^{1/6}\sum_{j=1}^{N_n}\int_{B_j} X_n(t)\,dt,\\
R_n &= n^{1/6}\int_{N_n(L_n+M_n)}^1 X_n(t)\,dt.
\end{aligned}$$

According to Lemma 5.7 and the Cauchy-Schwarz inequality, for all $s, t \in (0,1)$,

$$E|X_n(s)X_n(t)| \leq C, \tag{5.45}$$

where C is uniform with respect to s, t and n. Together with a fact that length of the interval of integration for R_n is $O(n^{-1/3}(\log n)^3)$ this shows $E|R_n| \to 0$ and hence $R_n = o_p(1)$.

Next we show that contribution of integrals over small blocks is negligible. To this end consider

$$E(S_n^{''})^2 = n^{1/3}\sum_{j=1}^{N_n} E\left(\int_{B_j} X_n(t)\,dt\right)^2 + n^{1/3}\sum_{i\neq j}\int_{B_i}\int_{B_j} EX_n(s)X_n(t)\,ds\,dt.$$

We have that

$$|EX_n(s)X_n(t)| = |g(s)g(t)||\mathrm{cov}(A_n^W(s)^k, A_n^W(t)^k)| \leq D_1 e^{-D_2 n|s-t|^3},$$

where $D_1, D_2 > 0$ do not depend s, t and n, using the fact that g in uniformly bounded on $[0,1]$ together with Lemma 5.11. Moreover, for $s \in B_i$ and $t \in B_j$, we have $|s-t| \geq n^{-1/3}(\log n)^3$. Since $N_n = \mathcal{O}(n^{1/3}/(\log n)^3)$ this implies that

$$\left|n^{1/3}\sum_{i\neq j}\int_{B_i}\int_{B_j} EX_n(s)X_n(t)\,ds\,dt\right| \leq n^{1/3}N_n^2 M_n^2 D_1 e^{-D_2(\log n)^9} \to 0.$$

Hence, using (5.45) we obtain

$$E(S_n^{''})^2 = \mathcal{O}(n^{1/3}N_n M_n^2) + o(1) \to 0,$$

so that the contribution of the small blocks is negligible.

Define

$$Y_j = n^{1/6}\int_{A_j} X_n(t)\,dt \quad \text{and} \quad \sigma_n^2 = \mathrm{var}\left(\sum_{j=1}^{N_n} Y_j\right),$$

so that $S_n^{'} = \sum_{j=1}^{N_n} Y_j$ and $\sigma_n^2 = \mathrm{var}(S_n^{'})$. We have

$$\begin{aligned}
&\left|E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{N_n} Y_j\right\} - \prod_{j=1}^{N_n} E\exp\left\{\frac{iu}{\sigma_n}Y_j\right\}\right| \\
&\quad \leq \sum_{k=2}^{N_n}\left|E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{k} Y_j\right\} - E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{k-1} Y_j\right\}E\exp\left\{\frac{iu}{\sigma_n}Y_k\right\}\right| \\
&\quad \leq 4(N_n - 1)\alpha_n(M_n),
\end{aligned}$$

where the last inequality follows from Lemma 5.10. Observe that $(N_n - 1)\alpha_n(M_n) \to 0$, which means that we can apply the Central Limit Theorem to independent copies of Y_j. Asymptotic normality of S_n' follows if we can show that the independent copies of the Y_j's satisfy the Lindeberg condition, i.e., for all $\epsilon > 0$,

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{\{|Y_j| > \epsilon\sigma_n\}} \to 0,$$

as $n \to \infty$. Note that by the Markov inequality $EY_j^2 1_{\{|Y_j|>\epsilon\sigma_n\}} \le E|Y_j|^3/(\epsilon\sigma_n)$. Again using the Cauchy-Schwarz inequality and uniform boundedness of the moments of $|X_n(t)|$ we obtain

$$\sup_{1\le j\le N_n} E|Y_j|^3 = n^{1/2}\mathcal{O}(|A_j|^3) = \mathcal{O}(n^{-1/2}(\log n)^9).$$

Hence

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{|Y_j|>\epsilon\sigma_n} \le \frac{1}{\epsilon\sigma_n^3} N_n \sup_{1\le j\le N_n} E|Y_j|^3 = \mathcal{O}(\sigma_n^{-3} n^{-1/6}(\log n)^6).$$

Note that

$$\sigma_n^2 = \mathrm{var}(S_n^{'}) = \mathrm{var}\left(T_n^W\right) + \mathrm{var}(S_n^{''} + R_n) - 2ET_n^W(S_n^{''} + R_n).$$

Using the already obtained results $E(S_n^{''})^2 = o(1)$ and $ER_n^2 = o(1)$, together with the Cauchy-Schwarz inequality, we conclude that

$$\mathrm{var}(S_n^{''} + R_n) = E(S_n^{''})^2 + ER_n^2 + 2E(S_n^{''}R_n) \to 0,$$

and that according to the Lemma 5.12

$$ET_n^W(S_n^{''} + R_n) \le \sqrt{E\left(T_n^W\right)^2 \mathrm{var}(S_n^{''} + R_n)} \to 0.$$

So we find that $\sigma_n^2 = \mathrm{var}(S_n^{'}) = \sigma^2 + o(1)$, which implies

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{\{|Y_j|>\epsilon\sigma_n\}} = o(n^{-1/6}(\log n)^6) \to 0. \quad \blacksquare$$

Proof of Theorem 3.1: Let G_n be the empirical distribution function of the U_i's and let $\hat{G}_n$ be its least concave majorant on $[0,1]$. Then $F_n(t) = G_n(F(t))$ for all $t \in [0,1]$. When f is non-increasing, then F is concave, so that $\hat{G}_n(F(t))$ is also concave. Moreover, $\hat{G}_n(F(t))$ lies above $G_n(F(t)) = F_n(t)$. Since $\hat{F}_n(t)$ is the least concave function on [0,1] that lies above F_n, it follows that $\hat{F}_n(t) \le \hat{G}_n(F(t))$. We find that for all $t \in [0,1]$,

$$\hat{F}_n(t) - F_n(t) \le \hat{G}_n(F(t)) - G_n(F(t)). \tag{5.46}$$

It follows that $S_n(X_1, X_2, \ldots, X_n) \le S_n(U_1, U_2, \ldots, U_n)$. When $f = 1$, then $F(t) = t$, so that according to property 2 of Lemma 5.2,

$$n^{1/2}(\hat{F}_n(t) - F_n(t)) = n^{1/2}\big(\hat{F}_n(t) - t\big) - \big(F_n(t) - t\big) = [\mathrm{D}_{[0,1]}E_n](t), \tag{5.47}$$

where $E_n(t)$ is the uniform empirical process. Since the mapping $h \mapsto \sup_{t\in[0,1]}[\mathrm{D}_{[0,1]}h](t)$ is continuous, it follows that S_n converges in distribution to $\sup_{t\in[0,1]}[\mathrm{D}_{[0,1]}B](t)$, where B denotes Brownian bridge. Because $B(t)$ has the same distribution as $W(t) - tW(1)$ and, according to property 2 of Lemma 5.2, $\mathrm{D}_{[0,1]}$ is invariant under addition of linear functions, this proves Theorem 3.1. $\blacksquare$

Proof of Theorem 3.2: Using $F_n(t) = G_n(F(t))$ and (5.46), we find

$$\begin{aligned}\int_0^1 \big(\hat{F}_n(t) - F_n(t)\big)\, dF_n(t) \quad &\le \quad \int_0^1 \big(\hat{G}_n(F(t)) - G_n(F(t))\big)\, dG_n(F(t)) \\ &= \quad \int_0^1 [\mathrm{D}_{[0,1]}G_n](t)\, dG_n(t),\end{aligned}$$

which means that $R_n(X_1, X_2, \dots, X_n) \leq R_n(U_1, U_2, \dots, U_n)$. When $f = 1$, then similar to the proof of Theorem 3.1, using that the mapping $h \mapsto \int_0^1 [\mathrm{D}_{[0,1]}h](t)\,dt$ is continuous, it follows that

$$n^{k/2} \int_0^1 \left(\hat{F}_n(t) - F_n(t)\right)^k dt \to \int_0^1 \left(\hat{W}(t) - W(t)\right)^k dt.$$

To prove the same for R_n, it suffices to show

$$n^{k/2} \int_0^1 (\hat{F}_n(t) - F_n(t))^k d(F_n(t) - F(t)) = o_p(1).$$

To this end, let $\mathcal{G}_n$ be the class of functions $(F_1 - F_2)^k$, where $F_1 \geq F_2$ are distribution functions satisfying $\sup|F_1 - F_2| \leq n^{-1/2}\log n$. According to Theorem 3.1, we have that $(\hat{F}_n - F_n)^k$ is in $\mathcal{G}_n$ with probability tending to one. Therefore, we can restrict ourself to proving

$$\sup_{h \in \mathcal{G}_n} \int_0^1 h(t)\, d(F_n(t) - F(t)) = o_p(n^{-k/2}). \tag{5.48}$$

First, note that the total variation of $(F_1 - F_2)^k$ in $\mathcal{G}_n$ is

$$V = \mathrm{TV}((F_1 - F_2)^k) \leq k \sup(F_1 - F_2)^{k-1} \mathrm{TV}(F_1 - F_2) \leq 2kn^{-(k-1)/2}(\log n)^{k-1}.$$

Therefore, $(F_1 - F_2)^k$ in $\mathcal{G}_n$ is of bounded variation and may be represented as a difference of two monotone functions, both bounded by V. Hence, if $\mathcal{F}$ denotes this class, then $\mathcal{G}_n \subseteq \{f_1 - f_2; f_1, f_2 \in \mathcal{F}\}$. This implies that the entropy with bracketing $H_B(\delta, \mathcal{G}_n, P)$ with respect to $L_2(P)$-norm is bounded as

$$H_B(\delta, \mathcal{G}_n, P) \leq 2H_B(\delta/(2V), \mathcal{F}, P) \leq Dn^{-(k-1)/2}(\log n)^{k-1}/\delta, \tag{5.49}$$

where the constant $D > 0$ does not depend on $\delta > 0$ and P. Next, we will apply Theorem 5.11 in van de Geer (2000). Application of this theorem involves a suitable bound on the generalized entropy with bracketing $\mathcal{H}_{B,K}(\delta, \mathcal{G}_n, P)$ with respect to ρ_K, defined by

$$\rho_K^2(g) = 2K^2 \int \left(e^{|g|/K} - 1 - |g|/K\right)^2 dP$$

(see Definition 5.1 in van de Geer (2000)). With $K = 4n^{-k/2}(\log n)^k$, it follows from Lemma 5.10 in van de Geer (2000) that

$$\mathcal{H}_{B,K}(\delta, \mathcal{G}_n, P) \leq H_B(\delta/\sqrt{2}, \mathcal{G}_n, P) \leq 2Dn^{-(k-1)/2}(\log n)^{k-1}/\delta. \tag{5.50}$$

Furthermore, it is easy to see that there exists a constant $B > 0$, such that for any $g \in \mathcal{G}_n$,

$$\rho_K(g) \leq B\rho_{L_2(P)}(g) \leq Bn^{-k/2}(\log n)^k.$$

Now, (5.48) follows from Theorem 5.11 in van de Geer (2000), with $K = 4n^{-k/2}(\log n)^k$, $R = Bn^{-k/2}(\log n)^k$, $a = C(BD)^{1/2}n^{-(2k-1)/4}(\log n)^{(2k-1)/2}$, $C_1 = 1$, and $C_0 = C\sqrt{2}$, where C is a universal constant. ∎

Proof of Theorem 3.3: First note that for $\gamma = (k+1)/3$,

$$\left|\int_0^1 A_n(t)^k \left(|\hat{f}_n'(t)|^\gamma - |f'(t)|^\gamma\right) dt\right| \leq \gamma \sup_{t \in [0,1]} |\xi_t|^{\gamma-1} \int_0^1 A_n(t)^k |\hat{f}_n'(t) - f'(t)|\, dt, \tag{5.51}$$

where ξ_t is between $|\hat{f}_n'(t)|$ and $|f'(t)|$. Since both $\sup|\hat{f}_n'| = \mathcal{O}_p(1)$ and $1/\inf|\hat{f}_n'| = \mathcal{O}_p(1)$, it follows that $\sup|\xi_t|^{\gamma-1} = \mathcal{O}_p(1)$ for any γ. By application of Hölders inequality, the integral on the right hand side is bounded by

$$\left(\int_0^1 A_n(t)^{pk}\,dt\right)^{1/p}\left(\int_0^1 |\hat{f}_n'(t) - f'(t)|^q\,dt\right)^{1/q},$$

where $p \geq 1$, is chosen such that $1/p + 1/q = 1$. According to Lemma 5.7 the first integral is of the order $\mathcal{O}_p(1)$, and according to condition (ii), the second term is of the order $o_p(n^{-1/6})$. Next, consider

$$\begin{aligned}
&\left|\int_0^1 \hat{f}_n(t)^{2k/3}|\hat{f}_n'(t)|^{1/3}\,dt - \int_0^1 f(t)^{2k/3}|f'(t)|^{1/3}\,dt\right| \\
&\quad\leq \int_0^1 \hat{f}_n(t)^{2k/3}\left||\hat{f}_n'(t)|^{1/3} - |f'(t)|^{1/3}\right|\,dt + \int_0^1 |f'(t)|^{1/3}\left|\hat{f}_n(t)^{2k/3} - f(t)^{2k/3}\right|\,dt \\
&\quad\leq \sup(\hat{f}_n)^{2k/3}\int_0^1 \left||\hat{f}_n'(t)|^{1/3} - |f'(t)|^{1/3}\right|\,dt + \sup|f'|^{1/3}\int_0^1 \left|\hat{f}_n(t)^{2k/3} - f(t)^{2k/3}\right|\,dt
\end{aligned}$$

Conditions (A1)-(A3) imply that $\sup|f'|$ is bounded and $\sup|\hat{f}_n| = \mathcal{O}_p(1)$. The two integrals can be treated in the same way as in (5.51). It follows that $\mu_n - \mu = o_p(n^{-1/6})$. Finally, $\sigma_n - \sigma = o_p(n^{-1/6})$ can be shown similarly. This proves the theorem. ■

Lemma 5.13 *Let f and K satisfy (A1)-(A3) and (K1)-(K3). Suppose that $h = \mathcal{O}(n^{-1/5})$, then for $i = 0, 1$,*

(i) $\sup_{t\in[0,1]}|\hat{f}_{n,h}^{(i)}(t)| = O_p(1)$ *and* $1/\inf_{t\in[0,1]}|\hat{f}_{n,h}^{(i)}(t)| = O_p(1)$,

(ii) for any $1 \leq q < 6/5$, $\int_0^1 |\hat{f}_n^{(i)}(t) - f^{(i)}(t)|^q\,dt = o_p(n^{-q/6})$.

Proof: Following the proof of Prakasa Rao (1983) page 38, we have that

$$\sup_{t\in[0,1]}|\hat{f}_{n,h}(t) - E\hat{f}_{n,h}(t)| \leq \frac{1}{h}\sup_{y\in[0,1]}|F_n(y) - F(y)|\,\mathrm{TV}(K),$$

where $\mathrm{TV}(K)$ denotes the total variation of K. Note that since K is differentiable, it is of bounded variation, so that $\mathrm{TV}(K) < \infty$. This implies that the above supremum is of the order $\mathcal{O}_p(h^{-1}n^{-1/2}) = \mathcal{O}_p(n^{-3/10})$. Similarly,

$$\sup_{t\in[0,1]}|\hat{f}_{n,h}'(t) - E\hat{f}_{n,h}'(t)| \leq \frac{1}{h^2}\sup_{y\in[0,1]}|F_n(y) - F(y)|\,\mathrm{TV}(K') = \mathcal{O}_p(h^{-2}n^{-1/2}) = \mathcal{O}_p(n^{-1/10}).$$

Next, we use that

$$E\hat{f}_{n,h}(t) = \int_{-1}^1 K(u)f(t-hu)\,du \tag{5.52}$$

and

$$E\hat{f}_{n,h}'(t) = \int_{-1}^1 K(u)f'(t-hu)\,du. \tag{5.53}$$

Then, it follows for $i = 0, 1$, that

$$\begin{aligned}\sup_{t\in[0,1]} |\hat{f}^{(i)}_{n,h}(t)| &\leq \sup_{t\in[0,1]} |\hat{f}^{(i)}_{n,h}(t) - E\hat{f}^{(i)}_{n,h}(t)| + \sup_{t\in[0,1]} |E\hat{f}^{(i)}_{n,h}(t)| \\ &\leq \mathcal{O}_p(n^{-1/10}) + \sup_{t\in[0,1]} |f^{(i)}(t)| = O_p(1).\end{aligned}$$

The infimum can treated similarly, using that

$$\begin{aligned}\inf_{t\in[0,1]} |\hat{f}^{(i)}_{n,h}(t)| &\geq \inf_{t\in[0,1]} |E\hat{f}^{(i)}_{n,h}(t)| - \sup_{t\in[0,1]} |\hat{f}^{(i)}_{n,h}(t) - E\hat{f}^{(i)}_{n,h}(t)| \\ &\geq \frac{1}{2} \inf_{t\in[0,1]} |f^{(i)}(t)| - \sup_{t\in[0,1]} |\hat{f}^{(i)}_{n,h}(t) - E\hat{f}^{(i)}_{n,h}(t)|.\end{aligned}$$

This proves (i).

For (ii), use the triangle inequality for the L_q-norm $\|\cdot\|_q$:

$$\|\hat{f}^{(i)}_{n,h} - f^{(i)}\|_q \leq \|\hat{f}^{(i)}_{n,h} - E\hat{f}^{(i)}_{n,h}\|_q + \|E\hat{f}^{(i)}_{n,h} - f^{(i)}\|_q. \tag{5.54}$$

For the second term, we can write

$$\left(\|E\hat{f}^{(i)}_{n,h} - f^{(i)}\|_q\right)^q = \int_0^1 |E\hat{f}^{(i)}_{n,h}(t) - f^{(i)}(t)|^q \, dt.$$

This integral can be decomposed into three integrals over the intervals $[0, h)$, $[h, 1-h]$, and $(1-h, 1]$. According to conditions (A1)-(A3), (5.52), and (5.53), the integrals over $[0, h)$ and $(1-h, 1]$ are of order $\mathcal{O}(h)$, and for $h \leq t \leq 1-h$, we use that

$$E\hat{f}_{n,h}(t) = \int_{-1}^1 K(u) f(t - hu) \, du = f(t) + \frac{h^2}{2} \int_{-1}^1 f''(\xi_{t,u}) u^2 K(u) \, du \tag{5.55}$$

and

$$E\hat{f}'_{n,h}(t) = \int_{-1}^1 K(u) f'(t - hu) \, du = f'(t) - h \int_{-1}^1 f''(\xi_{t,u}) u K(u) \, du. \tag{5.56}$$

This implies that also

$$\int_h^{1-h} |E\hat{f}^{(i)}_{n,h}(t) - f^{(i)}(t)|^q \, dt = \mathcal{O}(h).$$

It follows that for $i = 0, 1$,

$$\|E\hat{f}^{(i)}_{n,h} - f^{(i)}\|_q = \mathcal{O}(h^{1/q}) = o(n^{-1/6}).$$

To bound the first term in (5.54), write

$$\left(\|\hat{f}^{(i)}_{n,h} - E\hat{f}^{(i)}_{n,h}\|_q\right)^q = \int_0^1 |\hat{f}^{(i)}_{n,h}(t) - E\hat{f}^{(i)}_{n,h}(t)|^q \, dt.$$

Again we can decompose the integral into three integrals over the intervals $[0, h)$, $[h, 1-h]$, and $(1-h, 1]$. On $[0, h)$ and $(1-h, 1]$ we bound

$$|\hat{f}^{(i)}_{n,h}(t) - E\hat{f}^{(i)}_{n,h}(t)| \leq \sup_{t\in[0,1]} |\hat{f}^{(i)}_{n,h}(t) - E\hat{f}^{(i)}_{n,h}(t)| = \mathcal{O}_p(n^{-1/10}).$$

This implies that the integrals over $[0, h)$ and $(1-h, 1]$ are of the order $\mathcal{O}_p(hn^{-q/10}) = \mathcal{O}(n^{-(q+2)/10})$. For the integral over $[h, 1-h]$, we write

$$\int_h^{1-h} |\hat{f}_{n,h}^{(i)}(t) - E\hat{f}_{n,h}^{(i)}(t)|^q \, dt \leq \left(\int_h^{1-h} \left(\hat{f}_{n,h}^{(i)}(t) - E\hat{f}_{n,h}^{(i)}(t) \right)^2 \, dt \right)^{q/2}.$$

Now, also use that

$$\text{var}\hat{f}_{n,h}(t) \leq \frac{1}{nh} \int_{-1}^{1} K(u)^2 f(t-hu) \, du \tag{5.57}$$

and that

$$\text{var}\hat{f}'_{n,h}(t) \leq \frac{1}{nh^3} \int_{-1}^{1} K'(u)^2 f(t-hu) \, du. \tag{5.58}$$

According to (5.57) and (5.55), we have that

$$E \int_h^{1-h} \left(\hat{f}_{n,h}(t) - E\hat{f}_{n,h}(t) \right)^2 \, dt \quad \leq \quad \frac{\sup|f|}{nh} \int_{-1}^{1} K(u)^2 \, du = \mathcal{O}(n^{-4/5}).$$

Similarly, by means of (5.58) and (5.56),

$$E \int_h^{1-h} \left(\hat{f}'_{n,h}(t) - E\hat{f}'_{n,h}(t) \right)^2 \, dt \quad \leq \quad \frac{\sup|f|}{nh^3} \int_{-1}^{1} K'(u)^2 \, du = \mathcal{O}(n^{-2/5}).$$

It follows that for $i = 0, 1$

$$\int_h^{1-h} |\hat{f}_{n,h}^{(i)}(t) - E\hat{f}_{n,h}^{(i)}(t)|^q \, dt \leq \left(\int_h^{1-h} |\hat{f}_{n,h}^{(i)}(t) - E\hat{f}_{n,h}^{(i)}(t)|^2 \, dt \right)^{q/2} = \mathcal{O}(n^{-q/5}).$$

It follows that

$$\|E\hat{f}_{n,h}^{(i)} - f^{(i)}\|_q = \left(\mathcal{O}_p(hn^{-q/10}) + \mathcal{O}(n^{-q/5}) \right)^{1/q} = o_p(n^{-1/6}),$$

which proves (ii). ■

For boundary kernels we prove the following lemma. Both properties imply the two conditions of Theorem 3.3.

Lemma 5.14 *Let f and K satisfy (A1)-(A3) and (K1)-(K3). Suppose that $h \downarrow 0$ such that $nh^4 \to \infty$, then for $i = 0, 1$,*

(i) $\sup_{t \in [0,1]} |\hat{f}_{B,n,h}^{(i)}(t) - f^{(i)}(t)| = o_p(1)$.

(ii) $\int_0^1 (\hat{f}_{B,n,h}^{(i)}(t) - f^{(i)}(t))^2 \, dt = o_p(n^{-1/3})$.

Proof: For the boundary kernel estimator $\hat{f}_{B,n,h}$ we can use the same Taylor expansion (5.52), when $h \leq t \leq 1-h$. For $0 \leq t = \alpha h < h$, we have

$$E\hat{f}_{B,n,h}(t) = \int_{-1}^{\alpha} K_\alpha^L(u) f(t-hu) \, du = f(t) + \frac{h^2}{2} \int_{-1}^{\alpha} f''(\xi_{t,u}) u^2 K_\alpha^L(u) \, du,$$

and for $1-h < t = 1-\beta h \leq 1$, we have

$$E\hat{f}_{B,n,h}(t) = \int_{-\beta}^{1} K_\beta^R(u) f(t-hu)\,du = f(t) + \frac{h^2}{2}\int_{-\beta}^{1} f''(\xi_{t,u})u^2 K_\beta^R(u)\,du.$$

Hence, incorporating the definition of K_t given in (3.16), this means that for every $t \in [0,1]$,

$$E\hat{f}_{B,n,h}(t) = f(t) + \frac{h^2}{2}\int_{-1}^{1} f''(\xi_{t,u})u^2 K_t(u)\,du. \tag{5.59}$$

For the boundary kernel estimator of the derivative

$$\hat{f}'_{B,n,h}(t) = \frac{1}{h^2}\int K'_t\left(\frac{t-x}{h}\right)\,dF_n(x) - \frac{1}{h}K_t\left(\frac{t}{h}\right)\hat{f}_{B,n,h}(0) - \frac{1}{h}K_t\left(\frac{t-1}{h}\right)\hat{f}_{B,n,h}(1),$$

we use the Taylor expansion (5.53), when $h \leq t \leq 1-h$. Note that in that case $K_t = K$, and that $K_t(t/h)$ and $K_t((1-t)/h)$ are both zero. For $0 \leq t = \alpha h < h \leq 1/2$,

$$\begin{aligned}
K'_t\left(\frac{t-1}{h}\right) &= 0\\
\frac{1}{h}K_t\left(\frac{t}{h}\right)E\hat{f}_{B,n,h}(0) &= \frac{1}{h}K_\alpha^L(\alpha)f(0) + \frac{h}{2}K_\alpha^L(\alpha)\int_{-1}^{0} f''(\xi_{0,u})u^2 K_0^L(u)\,du
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{h^2}E\int K'_t\left(\frac{t-x}{h}\right)\,dF_n(x) &= \frac{1}{h}\int_{-1}^{\alpha} K'_t(u)f(t-hu)\,du\\
&= \left[\frac{1}{h}K_\alpha^L(u)f(t-hu)\right]_{-1}^{\alpha} + \int_{-1}^{\alpha} K_t(u)f'(t-hu)\,du\\
&= \frac{1}{h}K_\alpha^L(\alpha)f(0) + f'(t) - h\int_{-1}^{\alpha} f''(\xi_{t,u})uK_t(u)\,du.
\end{aligned}$$

It follows that for $0 \leq t = \alpha h < h \leq 1/2$,

$$E\hat{f}'_{B,n,h}(t) = f'(t) + \frac{h}{2}K_\alpha^L(\alpha)\int_{-1}^{0} f''(\xi_{0,u})u^2 K_0^L(u)\,du - h\int_{-1}^{\alpha} f''(\xi_{t,u})uK_t(u)\,du,$$

where $|\xi_{0,u}| \leq hu$ and $|\xi_{t,u} - t| \leq hu$. For $1/2 \leq 1-h < t = 1-\beta h \leq 1$,

$$\begin{aligned}
K'_t\left(\frac{t}{h}\right) &= 0\\
\frac{1}{h}K_t\left(\frac{t-1}{h}\right)E\hat{f}_{B,n,h}(1) &= \frac{1}{h}K_\beta^R(-\beta)f(1) + \frac{h}{2}K_\beta^R(-\beta)\int_{0}^{1} f''(\xi_{1,u})u^2 K_0^R(u)\,du
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{h^2}E\int K'_t\left(\frac{t-x}{h}\right)\,dF_n(x) &= \frac{1}{h}\int_{-\beta}^{1} K'_t(u)f(t-hu)\,du\\
&= \left[\frac{1}{h}K_\beta^R(u)f(t-hu)\right]_{-\beta}^{1} + \int_{-\beta}^{1} K_t(u)f'(t-hu)\,du\\
&= \frac{1}{h}K_\beta^R(-\beta)f(1) + f'(t) - h\int_{-\beta}^{1} f''(\xi_{t,u})uK_t(u)\,du.
\end{aligned}$$

It follows that for $1/2 \leq 1-h < t = 1-\beta h \leq 1$,

$$E\hat{f}'_{B,n,h}(t) = f'(t) + \frac{h}{2}K^R_\beta(-\beta)\int_0^1 f''(\xi_{1,u})u^2K^R_0(u)\,du - h\int_{-\beta}^1 f''(\xi_{t,u})uK_t(u)\,du,$$

where $|\xi_{1,u}| \leq hu$ and $|\xi_{t,u} - t| \leq hu$. Putting this together, we obtain for every $t \in [0,1]$,

$$\begin{aligned} E\hat{f}'_{B,n,h}(t) &= f'(t) - h\int_{-1}^1 f''(\xi_{t,u})uK_t(u)\,du \\ &\quad + \frac{h}{2}K_t\left(\frac{t}{h}\right)\int_{-1}^1 f''(\xi_{0,u})u^2K_0(u)\,du \\ &\quad + \frac{h}{2}K_t\left(\frac{t-1}{h}\right)\int_{-1}^1 f''(\xi_{1,u})u^2K_1(u)\,du. \end{aligned} \tag{5.60}$$

Furthermore, there exist constants $A, B > 0$ not depending on t, such that

$$|K_t(u)| \leq (A + B|u|)K(u). \tag{5.61}$$

We proceed as in the proof of Lemma 5.13. First note that

$$\begin{aligned} |K_t(x) - K_t(y)| &= c_{1,t}|K(x) - K(y)| + c_{2,t}|xK(x) - yK(y)| \\ &\leq C_1|K(x) - K(y)| + C_2|xK(x) - yK(y)|, \end{aligned}$$

for some $C_1 > 0$ and $C_2 > 0$ that do not depend on t. Hence, if $L(u) = uK(u)$ then

$$\sup_{t\in[0,1]} \mathrm{TV}(K_t) \leq C_1\mathrm{TV}(K) + C_2\mathrm{TV}(L).$$

Because $L'(u) = uK'(u) + K(u)$, it follows that L is of bounded variation. This means that

$$\sup_{t\in[0,1]} |\hat{f}_{B,n,h}(t) - E\hat{f}_{B,n,h}(t)| \leq \frac{1}{h}\sup_{y\in[0,1]}|F_n(y) - F(y)| \sup_{t\in[0,1]}\mathrm{TV}(K_t) = \mathcal{O}(n^{-1/2}h^{-1}).$$

On the other hand according to (5.59),

$$\sup_{t\in[0,1]} |E\hat{f}_{B,n,h}(t) - f(t)| \leq \frac{1}{2}h^2 \sup|f''| \sup_{t\in[0,1]}\int_{-1}^1 u^2K_t(u)\,du = \mathcal{O}(h^2).$$

This proves (i) for $i = 0$. Similarly, for the estimator of the derivative first note that

$$|K'_t(x) - K'_t(y)| \leq C_1|K'(x) - K'(y)| + C_2|xK'(x) - yK'(y)| + C_2|K(x) - K(y)|,$$

so that if $M(u) = uK'(u)$, then

$$\sup_{t\in[0,1]} \mathrm{TV}(K'_t) \leq C_1\mathrm{TV}(K') + C_2\mathrm{TV}(M) + C_2\mathrm{TV}(K).$$

This implies that

$$\begin{aligned} &\sup_{t\in[0,1]} |\hat{f}'_{B,n,h}(t) - E\hat{f}'_{B,n,h}(t)| \\ &\leq \frac{1}{h^2}\sup_{y\in[0,1]}|F_n(y) - F(y)| \left(\sup_{t\in[0,1]}\mathrm{TV}(K'_t) + \sup_{t,u\in[0,1]}|K'_t(u)| \sup_{t\in[0,1]}\mathrm{TV}(K_t)\right) \\ &= \mathcal{O}_p(n^{-1/2}h^{-2}). \end{aligned}$$

With (5.60),

$$\sup_{t\in[0,1]} |E\hat{f}'_{B,n,h}(t) - f'(t)| \le C_3 h,$$

for some $C_3 > 0$. This proves (i) for $i = 1$.

For (ii) we also need that

$$\text{var}\hat{f}_{B,n,h}(t) \le \frac{C_1}{nh}, \tag{5.62}$$

for some $C_1 > 0$ only depending on K and f. For the derivative we can write

$$\begin{aligned}\hat{f}'_{B,n,h}(t) &= \frac{1}{h^2}\int K'_t\left(\frac{t-x}{h}\right)dF_n(x) - \frac{1}{h}K_t\left(\frac{t}{h}\right)\hat{f}_{B,n,h}(0) - \frac{1}{h}K_t\left(\frac{t-1}{h}\right)\hat{f}_{B,n,h}(1)\\ &= \frac{1}{h^2}\int L_t\left(\frac{t-x}{h}\right)dF_n(x),\end{aligned}$$

where

$$L_t(u) = K'_t(u) - K'_t\left(\frac{t}{h}\right)K_0\left(u - \frac{t}{h}\right) - K'_t\left(\frac{t-1}{h}\right)K_1\left(u + \frac{1-t}{h}\right).$$

This means that

$$\text{var}\hat{f}'_{B,n,h}(t) \le \frac{1}{nh^3}\int_{-1}^{1} L_t(u)^2 f(t-hu)\,du \le \frac{C_2}{nh^3}, \tag{5.63}$$

for some $C_2 > 0$ only depending on K and f. According to (5.62) and (5.59), we have that

$$\begin{aligned}E\int_0^1 \left(\hat{f}_{B,n,h}(t) - f(t)\right)^2 dt &\le \frac{C_1}{nh} + \frac{\sup|f''|^2}{4}h^4\left(\int_{-1}^1 u^2K_t(u)\,du\right)^2\\ &= \mathcal{O}(n^{-4/5}),\end{aligned}$$

which proves (i) for $i = 0$. Similarly, by means of (5.63) and (5.60),

$$E\int_0^1 \left(\hat{f}'_{B,n,h}(t) - f'(t)\right)^2 dt \le \frac{C_2}{nh^3} + C_3h^2 = \mathcal{O}(n^{-2/5}),$$

which proves (ii) for $i = 1$.

References

Bowman, A.W., Jones, M.C. and Gijbels, I. (1998) Testing monotonicity of regression. *J. Comp. Graph. Statist.* **7** 489–500.

Cheng, M.Y. and Hall, P. (1997) Calibrating the excess mass and dip tests of modality. *J. R. Statist. Soc.*B **60** 579-589.

Cohen, A. and Sackrowitz, H.B. (1993) Evaluating tests for increasing intensity of a Poisson process. *Technometrics* **35** 446-448.

Dümbgen, L. and Spokoiny, V.G. (2001) Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124-152.

Durot, C. and Toquet, A.S. (2002) On the distance between the empirical process and its concave majorant in a monotone regression framework. *Ann. Inst. H. Poincaré Probab. Statist.* **39** 217–240.

Durot, C. (2003) A Kolmogorov-type test for monotonicity of regression. *Statist. Probab. Lett.* **63** 425–433.

van de Geer, S. *Empirical processes in M-estimation*, Cambridge University Press.

Ghosal, S., Sen, A. and van der Vaart, A.W. (2000) Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082.

Groeneboom, P., Hooghiemstra, G. and Lopuhaä, H.P. (1999) Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316-1347.

Hall, P. and Heckman, N.E. (2000) Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39.

Hartigan, J.A. and Hartigan, P.M. (1985) The Dip test of unimodality. *Ann. Statist.* **13** 70–84.

Ibragimov, I.A. and Linnik, Y.V. (1971) *Independent and stationary sequences of random variables.* Wolters-Nordhoff, Groningen.

Kiefer, J. and Wolfowitz, J. (1976) Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. verw. Gebiete* **34** 73-85.

Kómlos, J., Major, P., and Tusnády, G. (1975) An approximation of partial sums of independent RV's and the sample DF. *Z. Wahrsch. verw. Gebiete* **32** 111-131.

Kulikov, V.N. and Lopuhaä, H.P. (2003) The limit process of the difference between the empirical distribution function and its concave majorant. *To be submitted.*

Kulikov, V.N. and Lopuhaä, H.P. (2004) Testing for a monotone density using L_k-distances between the empirical distribution function and its concave majorant. Report Eurandom, 2004.

Mammen, E., Maron, J.S. and Fisher, N.I. (1992) Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Theory Related Fields* **91** 115–132.

Müller, H.G. and Sawitzki, G. (1991) Excess mass estimates and tests for multimodality. *J. Am. Statist. Assoc.* **86** 738-746.

Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation.* Academic Press, Orlando.

Robertson, Wright and Dijkstra (1988) *Order Restricted Statistical Inference.* Wiley, New York.

Schlee, W. (1982) Nonparametric tests of the monotonicity and convexity of regression. In: Gnedenko, B.V., Puri, M.L., Vincze, I. (Eds.), *Nonparametric Statistical Inference*, Vol.2, North-Holland, Amsterdam, 823–836.

Silverman, B.W. (1981) Using kernel density estimates to investigate multi-modality. *J. R. Statist. Soc.*B **43** 97-99.

Wang, Y. (1994) The limit distribution of the concave majorant of an empirical distribution function. *Statistics and Probability Letters* **20** 81-84.

Woodroofe, M. and Sun, J.S. (1999) Testing uniformity versus a monotone density. *Ann. Statist.* **27** 338-360.