

UC San Diego

Recent Work

Title

ON THE BEHAVIOR OF NONPARAMETRIC DENSITY AND SPECTRAL DENSITY ESTIMATORS AT ZERO POINTS OF THEIR SUPPORT

Permalink

<https://escholarship.org/uc/item/40q0z0tz>

Author

Politis, Dimitris

Publication Date

2012-12-01

ON THE BEHAVIOR OF NONPARAMETRIC DENSITY AND SPECTRAL DENSITY ESTIMATORS AT ZERO POINTS OF THEIR SUPPORT

EFSTATHIOS PAPARODITIS AND DIMITRIS N. POLITIS

ABSTRACT. The asymptotic behavior of nonparametric estimators of the probability density function of an i.i.d. sample and of the spectral density function of a stationary time series have been studied in some detail in the last 50-60 years. Nevertheless, an open problem remains to date, namely the behavior of the estimator when the target function happens to vanish at the point of interest. In the paper at hand we fill this gap, and show that asymptotic normality still holds true but with a super-efficient rate of convergence. We also provide two possible applications where these new results can be found useful in practice.

1. INTRODUCTION

Nonparametric estimation of a probability density function (p.d.f.) via local averaging, i.e., kernel smoothing, was developed in the late 1950s following the development of nonparametric estimation of the spectral density function (s.d.f.) of a stationary time series; cf. [1] [2] [4] [10] [11] [12] [13] [14]. Despite the fact that the settings are very different, the two subjects share many similarities that are brought forth in the monograph by Rosenblatt [16].

To describe them, let the kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric, square-integrable function satisfying $\int K(u)du = 1$. Let d be the smallest natural number for which $\int u^d K(u)du \neq 0$ where the integral is assumed to exist; then, the kernel K is said to have order d . By the assumed symmetry, the order of a kernel will be at least 2 as long as $\int u^2 K(u)du < \infty$; in fact, if $K(u) \geq 0$ everywhere, then K will have exactly order 2. Also define the bandwidth $h > 0$, and the re-scaled kernel $K_h(\cdot) = h^{-1}K(\cdot/h)$.

Date: November 27, 2012.

Key words and phrases. Central limit theorem, kernel smoothing, nonparametric estimation, overdifferencing, time series.

Consider the following two settings:

- (A) Let $X_1, X_2, \dots, X_n$ be independent, identically distributed (i.i.d.) with common distribution F that is absolutely continuous. Define the p.d.f. $f(x) = F'(x)$; then, the kernel smoothed estimator of $f(x)$ is $\hat{f}_n(x) = n^{-1} \sum_{j=1}^n K_h(x - X_j)$.

Under regularity conditions, for any point x such that $f(x) > 0$, it is well known that letting $h \rightarrow 0$ but with $hn \rightarrow \infty$ as $n \rightarrow \infty$ we have that

$$(1.1) \quad \sqrt{nh}(\hat{f}(x) - E\hat{f}(x)) \Rightarrow N(0, f(x) \int K^2(u)du)$$

where " $\Rightarrow$ " denotes weak convergence.

- (B) Let $X_1, X_2, \dots, X_n$ be a stretch of a second-order stationary time series $\{X_t, t \in \mathbb{Z}\}$ with s.d.f. defined as $f(\lambda) = (2\pi)^{-1} \sum_{s \in \mathbb{Z}} \gamma(s) e^{-is\lambda}$ for $\lambda \in [-\pi, \pi]$; here $\gamma(s) = \text{Cov}(X_t, X_{t+s})$ is the autocovariance at lag s . The kernel smoothed estimator of $f(\lambda)$ is $\hat{f}_n(\lambda) = n^{-1} \sum_{j \in \mathbb{Z}} K_h(\lambda - \lambda_j) I_n(\lambda_j)$ where $I_n(\lambda) = (2\pi n)^{-1} |\sum_{j=1}^n X_j \exp\{-i\lambda j\}|^2$ is the periodogram, and $\lambda_j = 2\pi j/n$.

Under regularity conditions, for any point λ such that $f(\lambda) > 0$, it is well known that letting $h \rightarrow 0$ but with $hn \rightarrow \infty$ as $n \rightarrow \infty$ we have that

$$(1.2) \quad \sqrt{nh}(\hat{f}(\lambda) - E\hat{f}(\lambda)) \Rightarrow N(0, (1 + \eta_\lambda) f^2(\lambda) \int K^2(u)du)$$

where $\eta_\lambda = 1$ if $\lambda = 0 \pmod{\pi}$ and $\eta_\lambda = 0$ otherwise.

Thus, although the p.d.f. and s.d.f. estimators are obtained under very different conditions, their asymptotic behavior is very similar; both estimators are $\sqrt{nh}$ consistent, and their asymptotic variance depends on the value of the function (p.d.f case) to be estimated or its square (s.d.f. case), as well as the L_2 norm of the kernel K .

Despite the fact that the problem has been so well studied in the last 50-60 years, the asymptotic behavior of the above estimators has been unknown when the point x (resp. λ) of interest is such that $f(x) = 0$ (resp. $f(\lambda) = 0$). The paper at hand attempts to fill this gap. In particular, we will show in the sequel that in this case the aforementioned similarities in the asymptotic behavior of the p.d.f. and s.d.f. estimators break down but a new similarity emerges: both estimators become *super-efficient*, i.e., they achieve rates of convergence faster than usual.

In the following we investigate the asymptotic behavior of these kernel estimators at so-called *zero points* of f , that is, at points of their support that are in the set

$$N_f = \{x : f(x) = 0 \text{ and } \exists \epsilon > 0 \text{ such that } f(y) > 0 \text{ for all } y \in (x - \epsilon, x + \epsilon) \setminus \{x\}\}.$$

In Section 2, the asymptotic theory for p.d.f. estimation is developed, and an application is discussed. Section 3 has the theory for s.d.f. estimation, and another application. All proofs are placed in Section 4.

2. DENSITY ESTIMATION IN THE I.I.D. CASE

2.1. Asymptotic Results. The following theorem describes the properties of the estimator $\hat{f}_n(x_0)$ when $x_0 \in N_f$ and $X_1, X_2, \dots, X_n$ are i.i.d. with p.d.f. f , i.e., case (A) of the Introduction.

Theorem 2.1. *Let $x_0 \in N_f \neq \emptyset$, and assume that $f \in \mathcal{C}^{(r)}$, the set of r -times continuously differentiable functions, with $r = 2m$ for some $m \in \mathbb{N}$. Assume that at least one of the derivatives $f^{(2)}(x), f^{(4)}(x), \dots, f^{(r)}(x)$ is non-zero at $x = x_0$, and let $r_0 = \min\{s : f^{(s)}(x_0) \neq 0 \text{ where } s = 2k \text{ for some natural number } k \leq m\}$. Finally, assume the integrals $\int u^{r_0} K(u) du$ and $\int u^{r_0} K^2(u) du$ are both finite.*

1. *If $h \sim C n^{-\delta}$ for some constants $C > 0$ and $0 < \delta < 1$, then the following assertions are true as $n \rightarrow \infty$.*
 - (i) $E\hat{f}_n(x_0) = f(x_0) + \frac{1}{r_0!} h^{r_0} f^{(r_0)}(x_0) \int u^{r_0} K(u) du + O(h^{r_0+2})$.
 - (ii) $\text{Var}(\hat{f}_n(x_0)) = \frac{h^{r_0-1}}{nr_0!} f^{(r_0)}(x_0) \int u^{r_0} K^2(u) du + O(n^{-1} h^{r_0})$
2. *If $h \sim C n^{-\delta}$ for some constants $C > 0$ and $0 < \delta < 1/(r_0 + 1)$, then as $n \rightarrow \infty$,*

$$\sqrt{\frac{n}{h^{r_0-1}}}(\hat{f}_n(x_0) - E\hat{f}_n(x_0)) \Rightarrow N(0, \tau^2(x_0)),$$

$$\text{where } \tau^2(x_0) = \frac{1}{r_0!} f^{(r_0)}(x_0) \frac{1}{2} \int u^{r_0} K^2(u) du.$$

Part 1(i) of the above theorem shows that the Bias of $\hat{f}_n(x_0)$, has the same form whether $x_0 \in N_f$ or not. For example, for a 2nd order kernel with $f^{(2)}(x_0) \neq 0$ we have $\text{Bias}(\hat{f}_n(x_0)) = O(h^2)$.

Nevertheless, Parts 1(ii) and 2 of Theorem 2.1 show a striking departure from the usual case of eq. (1.2). To see why, consider again a 2nd order kernel with $f^{(2)}(x_0) \neq 0$; if $x_0 \in N_f$, then the nonparametric estimator $\hat{f}_n(x_0)$ converges at the rate $n^{(1+\delta)/2}$ which is not only faster than the rate $\sqrt{nh} = n^{(1-\delta)/2}$ of eq. (1.2)—it is even faster than the parametric rate $n^{1/2}$. Hence, when $x_0 \in N_f$, the nonparametric estimator $\hat{f}_n(x_0)$ is *super-efficient*.

2.2. An Application. Suppose that it is conjectured that the common p.d.f. f of $X_1, \dots, X_n$ has a zero point within its support. This might happen, for instance, if—unbeknownst to the practitioner—the data were generated as $X_i = \epsilon_i \zeta_i$ where the ζ_i are i.i.d. with χ^2 distribution with $k > 2$ degrees of freedom, and ϵ_i are i.i.d. with $\text{Prob}\{\epsilon_i = 1\} = \text{Prob}\{\epsilon_i = -1\} = 1/2$; in this case, the support of f is the whole real line but $f(0) = 0$.

So it may be of interest to devise a nonparametric test of the null hypothesis $H : f(0) = 0$ vs. $\bar{H} : f(0) > 0$. Such a test will reject H when the kernel estimator $\hat{f}_n(\lambda_0)$ is found significantly different from zero for $\lambda_0 = 0$. It is apparent that in order to conduct this test, the distribution of $\hat{f}_n(0)$ is required under the null hypothesis that $f(0) = 0$.

Theorem 2.1 is now found useful as it provides a large-sample approximation for the required null distribution from which the test's critical region could be computed. Nevertheless, the issue of optimal choice of the bandwidth h associated with the estimator $\hat{f}_n(0)$ is raised. To briefly discuss this, suppose for concreteness that K is a 2nd order, nonnegative kernel. In this case, under the alternative hypothesis, the bandwidth $h \sim C n^{-1/5}$ is optimal with respect to minimizing the Mean Squared Error of $\hat{f}_n(0)$; to see this, combine the large-sample variance from (1.1) with the bias from Part 1(i) of Theorem 2.1 which is of the same form whether $f(0) = 0$ or not. But under the null hypothesis, Part 1 of Theorem 2.1 implies that the choice $h \sim C n^{-\delta}$ with δ close to one is optimal. However, since an accurate estimator under the alternative hypothesis results into an increased power of the test, it is intuitive that the optimal bandwidth choice under the alternative must still be used, i.e., $h \sim C n^{-1/5}$. Future work may help substantiate the above intuitive recommendation.

3. SPECTRAL DENSITY ESTIMATION

3.1. Asymptotic Results: zero mean case. Consider now the setup of case (B) of the Introduction, i.e., $X_1, X_2, \dots, X_n$ is a stretch from a strictly stationary time series $\{X_t, t \in \mathbb{Z}\}$ with $\mu = E(X_t)$ and $\gamma(s) = \text{Cov}(X_t, X_{t+s})$, $s \in \mathbb{Z}$. Assume $E(X_t^4) < \infty$, and $\sum_{h_1, h_2, h_3 \in \mathbb{Z}} |c_4(X_0, X_{h_1}, X_{h_2}, X_{h_3})| < \infty$ where $c_4(X_0, X_{h_1}, X_{h_2}, X_{h_3})$ is the fourth order cumulant function. Define then the fourth order cumulant spectral density function as

$$f_4(\lambda_1, \lambda_2, \lambda_3) = (2\pi)^{-3} \sum_{h_1 \in \mathbb{Z}} \sum_{h_2 \in \mathbb{Z}} \sum_{h_3 \in \mathbb{Z}} c_4(h_1, h_2, h_3) e^{-i \sum_{j=1}^3 \lambda_j h_j}.$$

In this subsection, we assume for simplicity that $\mu = 0$.

Theorem 3.1. *Let $\lambda_0 \in N_f \neq \emptyset$, assume that f is four-times continuously differentiable with its second derivative at λ_0 satisfying $f^{(2)}(\lambda_0) \neq 0$. Also assume that K is a 2nd order kernel, and the integrals $\int u^2 K(u) du$ and $\int u^4 K^2(u) du$ are both finite. If $h \sim C n^{-\delta}$ for some constants $C > 0$ and $0 < \delta < 1/2$, then the following assertions are true as $n \rightarrow \infty$.*

- (i) $E\hat{f}_n(\lambda_0) = f(\lambda_0) + \frac{1}{2}h^2 f^{(2)}(\lambda_0) \int u^2 K(u) du + o(h^2).$
- (ii)

$$\begin{aligned} \text{Var}[\hat{f}_n(\lambda_0)] &= \frac{1}{n} f_4(\lambda_0, -\lambda_0, -\lambda_0) (1 + o(1)) \\ &\quad + \frac{h^3}{4n} (f^{(2)}(\lambda_0))^2 \frac{1}{2\pi} \int u^4 K^2(u) du + O(hn^{-2}). \end{aligned}$$

As in the p.d.f. setup of Theorem 2.1, part (i) of the above theorem shows that the Bias of $\hat{f}_n(\lambda_0)$ has the same form whether $\lambda_0 \in N_f$ or not.

Part (ii) Theorem 3.1 shows that if $\lambda_0 \in N_f$, then $\hat{f}_n(\lambda_0)$ is again super-efficient. What is surprising here is that its rate of convergence depends on whether the fourth order cumulant spectral density vanishes or not at the triplet $(\lambda_0, -\lambda_0, -\lambda_0)$. If it does not vanish, the rate is the parametric one of $\sqrt{n}$; if it does, the rate is even faster, i.e., $\sqrt{n/h^3} = n^{(1+3\delta)/2}$. For example, this faster rate is achieved when the time series is Gaussian, in which case f_4 vanishes everywhere. The reason for the slower (although still super-efficient) rate when $f_4(\lambda_0, -\lambda_0, -\lambda_0) \neq 0$ is that, in this case, the correlation between periodogram ordinates at Fourier frequencies can not be treated as negligible although it is of order $O(1/n)$; see the proof of Theorem 3.1.

To derive the asymptotic distribution of the spectral density estimator $\hat{f}_n(\lambda_0)$, some assumptions are needed concerning the dependence structure of the underlying process. Among the different available alternatives, we will assume the physical dependence condition introduced by Wu (2005). In particular, we assume that

$$X_j = R(\dots, \varepsilon_{j-1}, \varepsilon_j),$$

where R is a measurable function and $\{\varepsilon_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables. Let ε'_0 have the same distribution as ε_0 , and be independent of $\{\varepsilon_j, j \in \mathbb{Z}\}$. Define $X_{j,0} = R(\dots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_j)$ and consider the physical dependence measure

$$\delta_{j,p} = \left(E|X_j - X_{j,0}|^p \right)^{1/p}, \quad \text{where } p > 0.$$

To define $\delta_{j,p}$ for $p > 0$, it is obviously required that $E|X_j|^p < \infty$. Since we have already assumed $E(X_t^4) < \infty$, we will use the physical dependence measure $\delta_{j,4}$ in the following.

Theorem 3.2. *Assume the same assumptions as those of Theorem 3.1, including $E(X_t) = 0$ and $E(X_t^4) < \infty$. If $\sum_{j=0}^{\infty} \delta_{j,4} < \infty$, then*

$$c_n(\widehat{f}_n(\lambda_0) - E\widehat{f}_n(\lambda_0)) \Rightarrow N(0, \tau^2(\lambda_0)),$$

as $n \rightarrow \infty$, where

$$c_n = \sqrt{n} \quad \text{and} \quad \tau^2(\lambda_0) = f_4(\lambda_0, -\lambda_0, -\lambda_0) \quad \text{when} \quad f_4(\lambda_0, -\lambda_0, -\lambda_0) \neq 0,$$

and

$$c_n = \sqrt{n/h^3} \quad \text{and} \quad \tau^2(\lambda_0) = (f^{(2)}(\lambda_0))^2 \frac{1}{8\pi} \int u^4 K^2(u) du \quad \text{when} \quad f_4(\lambda_0, -\lambda_0, -\lambda_0) = 0.$$

3.2. Asymptotic Results: general case. We now revisit the results of the previous subsection *without* the simplifying assumption that the mean μ is zero. First note that an equivalent way to write the s.d.f. estimator is $\widehat{f}_n(\lambda) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \kappa(hs) \hat{\gamma}(s) e^{-is\lambda}$ where κ is related to the kernel K via a Fourier series $K(\lambda) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \kappa(s) e^{-is\lambda}$; see [3] for more details. In the above, the sample autocovariance is given by $\hat{\gamma}(s) = n^{-1} \sum_{t=1}^{n-|k|} (X_t - \bar{X}_n)(X_{t+|k|} - \bar{X}_n)$ where $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ is the sample mean.

Consider for a moment the $\tilde{\gamma}(k) = n^{-1} \sum_{t=1}^{n-|k|} (X_t - \mu)(X_{t+|k|} - \mu)$, and the associated $\tilde{f}_n(\lambda) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \kappa(hs) \tilde{\gamma}(s) e^{-is\lambda}$. Defining the mean zero series $Y_t = X_t - \mu$, it is easy to see that Theorems 3.1 and 3.2 are valid as stated (even if μ is nonzero) for the random function $\tilde{f}_n$ in place of $\widehat{f}_n$.

However, $\tilde{f}_n$ is not a proper statistic since μ is unknown. The question to be addressed in this section is to what extent $\widehat{f}_n$ (that uses the data series centered at the sample mean $\bar{X}_n$) is close to the quantity $\tilde{f}_n$ (that uses the data series centered at the true mean μ).

A starting point is to investigate how close $\bar{X}_n$ is to μ . To do this, we require an assumption that is slightly stronger than assuming f has $[r]$ continuous derivatives, namely:

$$(3.1) \quad \sum_{k=-\infty}^{\infty} |k|^r |\gamma(k)| < \infty \quad \text{for some real number } r \geq 1.$$

If $Z_1, Z_2, \dots$ is a sequence of random variables, and $a_1, a_2, \dots$ is a sequence of real numbers, the notation $Z_n = O_{L_2}(a_n)$ will be used to denote that the L_2 norm of Z_n is

of order $O(a_n)$, i.e., that $\sqrt{EZ_n^2} = O(a_n)$. By Chebyshev's inequality, it is immediate that $Z_n = O_{L_2}(a_n)$ implies the order “in probability” $Z_n = O_P(a_n)$.

Lemma 3.1. *Assume eq. (3.1).*

- (i) *If $f(0) > 0$, then $\bar{X}_n = \mu + O_{L_2}(1/\sqrt{n})$.*
- (ii) *If $f(0) = 0$, then $\bar{X}_n = \mu + O_{L_2}(n^{-(1+1 \wedge r)/2})$ where $\wedge$ denotes minimum.*

Part (ii) of Lemma 3.1 indicates that $\bar{X}_n$ is *super-efficient* for μ if $f(0) = 0$; this phenomenon had been noticed before—see [9] and the references therein. For example, condition $f(0) = 0$ implies $\bar{X}_n = \mu + O_{L_2}(1/n)$. This super-efficiency is very handy when analyzing the properties of $\hat{\gamma}(s)$ and $\hat{f}(w)$. To see this, let $M_n = \bar{X}_n - \mu$, and note that

$$(3.2) \quad \hat{\gamma}(k) = \tilde{\gamma}(k) - M_n^2 + M_n \left(n^{-1} \sum_{t=1}^{k-1} Y_t + n^{-1} \sum_{t=k+1}^n Y_t \right).$$

From (3.2) and Lemma 3.1, the following lemma is immediate.

Lemma 3.2. *Assume eq. (3.1), and let $\lambda \in [-\pi, \pi]$.*

- (i) *If $f(0) > 0$, then $\hat{\gamma}(s) - \tilde{\gamma}(s) = O_{L_2}(1/n)$ uniformly in s , and $\hat{f}(\lambda) - \tilde{f}(\lambda) = O_{L_2}(1/(hn))$ uniformly in λ .*
- (ii) *If $f(0) = 0$, then $\hat{\gamma}(s) - \tilde{\gamma}(s) = O_{L_2}(1/n^2)$ uniformly in s , and $\hat{f}(\lambda) - \tilde{f}(\lambda) = O_{L_2}(1/(hn^2))$ uniformly in λ .*

The following theorem is now immediate.

Theorem 3.3. *Assume the same assumptions as those of Theorem 3.1, including $E(X_t^4) < \infty$ but without the assumption $E(X_t) = 0$. If $\sum_{j=0}^{\infty} \delta_{j,4} < \infty$, then*

$$c_n(\hat{f}_n(\lambda_0) - E\hat{f}_n(\lambda_0)) \Rightarrow N(0, \tau^2(\lambda_0)),$$

as $n \rightarrow \infty$, where

$$c_n = \sqrt{n} \quad \text{and} \quad \tau^2(\lambda_0) = f_4(\lambda_0, -\lambda_0, -\lambda_0) \quad \text{when} \quad f_4(\lambda_0, -\lambda_0, -\lambda_0) \neq 0,$$

and

$$c_n = \sqrt{n/h^3} \quad \text{and} \quad \tau^2(\lambda_0) = (f^{(2)}(\lambda_0))^2 \frac{1}{8\pi} \int u^4 K^2(u) du \quad \text{when} \quad f_4(\lambda_0, -\lambda_0, -\lambda_0) = 0$$

with one caveat: if $0 \neq \lambda_0 \in N_f$ and $f_4(\lambda_0, -\lambda_0, -\lambda_0) = 0$ but $f(0) > 0$, the following additional condition is required: either $\delta < 1/5$ and/or K has compact support.

For a 2nd order kernel, a bandwidth choice corresponding to $\delta < 1/5$ is suboptimal as it typically leads to oversmoothing; in this connection, using a kernel with compact support would by-pass this potential difficulty. Note, however, that the above limitation is of concern only in the case where $f(0) > 0$ but $0 \neq \lambda_0 \in N_f$ with $f_4(\lambda_0, -\lambda_0, -\lambda_0) = 0$. Obviously, such a situation can not occur if the point of interest $\lambda_0 \in N_f$ is the origin; the next subsection provides an application in this interesting case.

3.3. An Application. Suppose that the stationary time series data $X_1, \dots, X_n$ have been obtained as the first differences of another observed process $\{Y_t\}$ that is assumed to have a ‘unit root’ [5]; this is common practice with financial series since (log)-price series often exhibit ‘random walk’-type trajectories. If this assumption is correct, then differencing is necessary in order to reduce the problem to a stationary setting. However, if $\{Y_t\}$ does *not* have a unit root but is itself stationary, then differencing is unnecessary and the process $X_t = Y_t - Y_{t-1}$ is said to be *overdifferenced*. Interesting phenomena manifest themselves with overdifferenced data, the crucial reason being that overdifferencing implies $f(0) = 0$; see Rosenblatt [15] for an early account on issues related to overdifferencing.

To discuss a concrete example, it is of interest to devise a nonparametric test of the null hypothesis $H : \{Y_t\}$ is stationary vs. $\bar{H} : \{Y_t\}$ has a unit root. This is a *reversal* of the usual unit root testing problem where the null hypothesis is $\bar{H}$; see [5] for a review. Since H implies $f(0) = 0$, a nonparametric test will reject H when the kernel estimator $\hat{f}_n(\lambda_0)$ is found significantly different from zero for $\lambda_0 = 0$. It is apparent that in order to conduct this test, the distribution of $\hat{f}_n(0)$ is required under the null hypothesis that $f(0) = 0$.

Theorem 3.3 is now useful as it provides a large-sample approximation for the required null distribution from which the test’s critical region could be computed. Nevertheless, in addition to the issue of optimal choice of the bandwidth h as discussed in Section 2.2, here we have the additional complication of a switching behavior of the large-sample variance of $\hat{f}_n(0)$ according to whether $f_4(0, 0, 0)$ is zero or not. Future work may shed light on these important practical questions.

4. PROOFS

Proof of Theorem 2.1: Note that $E\hat{f}_n(x_0) = \int K(u)f(x_0 - uh)du$. Plugging-in a Taylor series expansion of $f(x_0 - uh)$ around $f(x_0)$, and taking into account

the symmetry of the kernel K and the fact that $f(x_0) = 0$ for $x_0 \in N_f$ proves Assertion 1(i).

Assertion 1(ii) follows using similar arguments and the expression

$$\text{Var}(\widehat{f}_n(x_0)) = n^{-1}h^{-2} \left(\int K^2(u)f(x_0 - uh)d uh - \left(\int K(u)f(x_0 - uh)d uh \right)^2 \right).$$

To establish Assertion 2 via Liapunov's condition, it suffices to show that for some $\epsilon > 0$, the sum $\sum_{j=1}^n E|Z_{n,j}|^{2+\epsilon}$ converges to zero as $n \rightarrow \infty$, where

$$Z_{n,j} = \frac{1}{\tau(x_0)n^{1/2}h^{(r_0+1)/2}} \left(K((x_0 - X_j)/h) - EK((x_0 - X_j)/h) \right).$$

We then have,

$$\sum_{j=1}^n E|Z_{n,j}|^{2+\epsilon} \leq O(1) \frac{1}{n^{\epsilon/2}h^{[(r_0+1)(\epsilon+2)]/2}} E \left[K^{2+\epsilon}((x_0 - X_1)/h) \right] = O(n^{-\epsilon/2}h^{-[(r_0+1)\epsilon]/2}).$$

But the above tends to zero under the assumed condition $nh^{(r_0+1)} \rightarrow \infty$. $\square$

The proof of Theorem 3.1 uses the following lemma from Krogstad [7].

Lemma 4.1. *Let $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ be a zero mean fourth order stationary process with autocovariance function $\gamma(k) = E(X_t X_{t+k})$ satisfying $\sum_{k \in \mathbb{Z}} |k| |\gamma(k)| < \infty$, second order spectral density $f(\lambda)$ and fourth order cumulant spectral density $f_4(\lambda_1, \lambda_2, \lambda_3)$, $\lambda, \lambda_i \in [-\pi, \pi]$, $i = 1, 2, 3$. Then, for $\lambda_r = 2\pi r/n$, $r \in \mathbb{Z}$, we have*

$$\begin{aligned} \text{Cov}(I_n(\lambda_j), I_n(\lambda_k)) = & \mathbf{1}(|\lambda_j| = |\lambda_k|) \left\{ (1 + \eta_{\lambda_j}) E^2(I_n(\lambda_j)) + O(\log^2(n)n^{-2}) \right\} \\ & + n^{-1} f_4(\lambda_j, \lambda_k, \lambda_{-k})(1 + o(1)) + O(n^{-2}), \end{aligned}$$

where the $O(\cdot)$ and $o(\cdot)$ terms are uniform in λ_j, λ_k , and $\mathbf{1}(A)$ is the indicator of set A .

Proof of Theorem 3.1: Assertion (i) follows from the expression $E\widehat{f}_n(\lambda_0) = (2\pi)^{-1} \int K_h(\lambda_0 - x)f(x)dx + O(n^{-1})$, and using the symmetry of the kernel K and the fact that $f(x) = 0$ for $x \in N_f$.

To show Assertion (ii), use Lemma 4.1 and notice that

$$\begin{aligned} \text{Var}(\widehat{f}_n(\lambda_0)) &= n^{-2} \sum_j \sum_k K_h(\lambda_0 - \lambda_j) K_h(\lambda_0 - \lambda_k) \left\{ \mathbf{1}(|\lambda_j| = |\lambda_k|) E^2(I_n(\lambda_j)) + O(\log^2(n)/n^2) \right\} \\ &\quad + n^{-1} f_4(\lambda_j, \lambda_k, -\lambda_k)(1 + o(1)) \Big\} + O(n^{-2}) \\ &= V_{1,n} + V_{2,n} + O(n^{-2}) \end{aligned}$$

with an obvious notation for $V_{1,n}$ and $V_{2,n}$. Using $E(I_n(\lambda_j)) = f(\lambda_j) + O(n^{-1})$ for $j \neq 0$, we get that

$$\begin{aligned} V_{1,n} &= (nh)^{-2} \sum_j K^2((\lambda_0 - \lambda_j)/h)(1 + \eta_{\lambda_j})f^2(\lambda_j) + O(hn^{-2}) \\ &= \frac{h^3}{n} \frac{1}{2\pi} \int u^4 K^2(u) du \left(f''(\lambda_0) \right)^2 + O(hn^{-2}), \end{aligned}$$

where the last equality follows by a Taylor series expansion of $f^2(\lambda)$ around $f^2(\lambda_0)$ and taking into account that for $\lambda \in N_f$,

$$f^2(\lambda) = (f^2(\lambda))' = (f^2(\lambda))'' = (f^2(\lambda))''' = 0 \quad \text{and} \quad (f^2(\lambda))'''' = 6(f''(\lambda))^2 \neq 0.$$

Furthermore,

$$\begin{aligned} V_{2,n} &= \frac{1}{n^3} \sum_j \sum_k K_h(\lambda_0 - \lambda_j) K_h(\lambda_0 - \lambda_k) (f_4(\lambda_j, \lambda_k, -\lambda_k) + o(1)) \\ &= \frac{1}{n} f_4(\lambda_0, \lambda_0, -\lambda_0) (1 + o(1)) + o(n^{-1}) \end{aligned}$$

and Assertion (ii) is proven. $\square$

Proof of Theorem 3.2: Let

$$a_{n,r} = \frac{c_n}{2\pi n^2} \sum_j K_h(\lambda_0 - \lambda_j) e^{-i\lambda_j r}$$

and note that

$$c_n(\widehat{f}(\lambda_0) - E(\widehat{f}(\lambda_0))) = \sum_{t=1}^n \sum_{s=1}^n a_{n,t-s} X_t X_s$$

is a quadratic form. Hence, it suffices to show that the conditions of Theorem 6 of Liu and Wu [8] are satisfied. To do this, let $b_{n,r} = c_n(2\pi n)^{-2} \sum_j K_h(\lambda_0 - \lambda_j)$, $s_n^2 = \sum_{t=1}^n b_{n,t}^2$ and $\sigma_n^2 = \eta_{\lambda_0} \sum_{k=1}^n \sum_{t=1}^n b_{n,t-k}^2$. We then have to show the following:

$$\begin{aligned} \text{(i)} \quad & \frac{\max_{0 \leq t \leq n} b_{n,t}^2}{s_n^2} = o(1), \\ \text{(ii)} \quad & \frac{ns_n^2}{\sigma_n^2} = O(1), \\ \text{(iii)} \quad & \frac{\sum_{k=1}^n \sum_{t=1}^{k-1} \left| \sum_{j=1+k}^n a_{n,k-j} a_{n,t-j} \right|^2}{\sigma_n^4} = o(1), \\ \text{(iv)} \quad & \frac{\sum_{k=1}^n |b_{n,k} - b_{n,k-1}|^2}{s_n^2} = o(1). \end{aligned}$$

To see (i) notice that $s_n^2 = O(c_n^2 n^{-1})$ and verify that $\max_{0 \leq t \leq n} b_{n,t}^2 / s_n^2 = O(n^{-1}) \rightarrow 0$. Requirement (ii) follows since $ns_n^2 = O(1)$ and $\sigma_n^2 = O(c_n^2)$. To establish (iii) notice that because $\sum_{j=r}^n e^{ij\lambda_s} = (n-r)$ if $\lambda_s = 0$ and zero otherwise, we have

$$\begin{aligned} & \frac{\sum_{k=1}^n \sum_{t=1}^{k-1} \left| \sum_{j=1+k}^n a_{n,k-j} a_{n,t-j} \right|^2}{\sigma_n^4} \\ &= \frac{1}{(2\pi)^4 n^4} \sum_{k=1}^n \sum_{t=1}^{k-1} \left\{ n^{-2} \sum_{j=1+k}^n \sum_{r_1} \sum_{r_2} K_h(\lambda_0 - \lambda_{r_1}) K_h(\lambda_0 - \lambda_{r_2}) e^{-i\lambda_{r_1}(k-j) - i\lambda_{r_2}(t-j)} \right\}^2 \\ &= \frac{1}{n^6} \sum_{k=1}^n \sum_{t=1}^{k-1} (n-k)^2 \left\{ \frac{1}{n} \sum_r K_h(\lambda_0 - \lambda_r) K_h(\lambda_0 + \lambda_r) e^{-i\lambda_r(k-t)} \right\}^2 \\ &= O(n^{-2} h^{-2}) \rightarrow 0. \end{aligned}$$

Finally, (iv) follows immediately by the definition of the $b_{n,r}$'s. $\square$

Proof of Lemma 3.1: Part (i) is well known. To show part (ii), note that

$$\text{Var}(\sqrt{n}\bar{X}_n) = \sum_{k=-n}^n \left(1 - \frac{|k|}{n}\right) \gamma(k) = A - B$$

where $A = \sum_{|k| \leq n} \gamma(k)$, and $B = \sum_{k=-n}^n \frac{|k|}{n} \gamma(k)$.

But $A = \sum_{k=-\infty}^{\infty} \gamma(k) - \sum_{|k| > n} \gamma(k) = -\sum_{|k| > n} \gamma(k)$ since the assumption $f(0) = 0$ implies $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$. Hence,

$$|A| = \left| \sum_{|k| > n} \gamma(k) \right| \leq 2 \sum_{k > n} |\gamma(k)| \leq 2 \sum_{k > n} \left(\frac{k^r}{n^r} \right) |\gamma(k)| = O(1/n^r).$$

Similarly, writing $\frac{|k|}{n} = \left(\frac{|k|}{n} \right)^{1 \wedge r} \left(\frac{|k|}{n} \right)^{1 - 1 \wedge r}$ we have

$$|B| \leq \sum_{k=-n}^n \left(\frac{|k|}{n} \right)^{1 \wedge r} \left(\frac{|k|}{n} \right)^{1 - 1 \wedge r} |\gamma(k)| \leq \sum_{k=-n}^n \left(\frac{|k|}{n} \right)^{1 \wedge r} |\gamma(k)| = O(1/n^{1 \wedge r})$$

using eq. (3.1) and that fact that $|k| \leq n$ in the above sum. $\square$

Proof of Theorem 3.3: Note that the assumption that $f \in \mathcal{C}^{(4)}$ implies eq. (3.1) with $r = 3$; see [6]. So the proof of Theorem 3.3 is immediate from Lemma 3.1 and the fact that—as already mentioned—Theorem 3.2 is already valid as stated for $\tilde{f}_n$ in place of $\hat{f}_n$ in the setting of a general (and unknown) μ .

REFERENCES

- [1] Bartlett, M.S. (1946), On the theoretical specification of sampling properties of autocorrelated time series, *J. Roy. Statist. Soc., Suppl.*, 8, 27-41.
- [2] Bartlett, M.S. (1948). Smoothing periodograms from time series with continuous spectra, *Nature*, vol. 161, pp. 686-687.
- [3] Brockwell, P. and Davis, R.A. (1991). *Time Series: Theory and Methods, 2nd Ed.*, Springer, New York.
- [4] Daniell, P.J. (1946). Discussion of paper by M.S. Bartlett. *J. Roy. Statist. Soc. Suppl.*, vol. 8, pp. 88-90.
- [5] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- [6] Katznelson, Y. (1968). *An Introduction to Harmonic Analysis*, Dover, New York.
- [7] Krogstad, H.E. (1982), On the covariance of the periodogram, *J. Time Series Anal.*, vol. 3, 195-207.
- [8] Liu, W. and Wu, W.B. (2010), Asymptotics of spectral density estimates, *Econometric Theory*, vol. 26, pp. 1218-1245.
- [9] McElroy, T. and Politis, D.N. (2012). Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory, *Econometric Theory*, vol. 28, pp. 471-481.
- [10] Parzen, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Ann. Math. Statist.*, Vol. 28, No. 2., pp. 329-348.
- [11] Parzen, E. (1961), Mathematical Considerations in the Estimation of Spectra, *Technometrics*, vol. 3, 167-190.
- [12] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Stat.*, vol. 33, pp. 1065–1076, 1962.
- [13] Priestley, M.B. (1962), Basic considerations in the estimation of spectra, *Technometrics*, vol. 4, 551-564.
- [14] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Stat.*, vol. 27, no. 3, pp. 832–837.
- [15] Rosenblatt, M. (1961). Independence and dependence, *Proc. Fourth Berkeley Symposium on Math. Statist. and Prob.*, University of California Press, vol. 2, pp. 431-443.
- [16] Rosenblatt, M. (1991), *Stochastic Curve Estimation*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 3, Institute of Mathematical Statistics, Hayward CA.
- [17] Wu, W.B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences*, vol. 102, pp. 14150–14154.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CYPRUS, 1678 NICOSIA, CYPRUS

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA—SAN DIEGO, LA JOLLA, CA 92093-0112, USA