EMPIRICAL SPECTRAL PROCESSES AND THEIR APPLICATIONS TO STATIONARY POINT PROCESSES

By Michael Eichler

Universität Heidelberg

Abstract

We consider empirical spectral processes indexed by classes of functions for the case of stationary point processes. Conditions for the measurability and equicontinuity of these processes and a weak convergence result are established. The results can be applied to the spectral analysis of point processes. In particular, we discuss the application to parametric and nonparametric spectral density estimation.

1 Introduction

In the context of spectral analysis of time series, Dahlhaus (1988) introduced empirical processes where the spectral distribution function of a stationary process takes the part of the probability distribution. The asymptotic theory of these empirical spectral processes provides a method for proving limit theorems for statistics which depend on the spectral distribution.

In this paper, we are interested in empirical spectral processes derived from stationary point processes. Here a point process on $\mathbb{R}$ is defined as a random counting measure N where $N(A)$ denotes the number of point events occuring in some Borel set $A \subset \mathbb{R}$ [cf. Daley and Vere-Jones (1988)]. In a fundamental paper by Brillinger (1972), it was shown that the spectral analysis of such processes based on finite Fourier transforms leads to similar results as in time series analysis. As an important difference to the case of time series, we note that the cumulant spectra of point processes are functions on $\mathbb{R}$ which do not vanish for high frequencies and thus are not $\mathcal{L}^2$-integrable.

Consider a stationary point process N on $\mathbb{R}$. If N satisfies certain mixing conditions, the spectral density f_2 of N exists and is given by

$$f_2(\lambda) = \int_{\mathbb{R}} \exp(-i\lambda u) dC_2'(u), \qquad \lambda \in \mathbb{R}. \tag{1.1}$$

Here C_2' denotes the reduced cumulant measure of second order, which is defined by the equation

$$\mathrm{cum}\{N(A_1), N(A_2)\} = \int_{A_1} \int_{A_2} dC_2'(t_1 - t_2) dt_2 \tag{1.2}$$

AMS 1991 subject classifications. Primary 60F05; secondary 62M15, 60G55.

Key words and phrases. Point processes, empirical spectral process, functional central limit theorem, spectral density estimation.

for all $A_1, A_2 \in \mathcal{B}$. Now many interesting functionals in spectral analysis can be written in the form

$$A(\phi_\theta) = \int_{\mathbb{R}} \phi_\theta(\lambda) f_2(\lambda) d\lambda$$

with parameter $\theta \in \Theta \subset \mathbb{R}^p$. Examples we have in mind are the spectral distribution function $F_2(\alpha) = \int_0^\alpha f_2(\lambda)d\lambda$, the covariance density function $q_2(u) = \int_{\mathbb{R}} \exp(iu\lambda)(f_2(\lambda) - p/2\pi)d\lambda$ and the variance time curve $V(t) = \operatorname{var}\{N((0,t])\}$ [e.g. Brillinger (1975)].

If the process has been observed on the interval $[0, T]$, the spectral density can be estimated by the periodogram

$$I^{(T)}(\lambda) = \{2\pi H_2^{(T)}(0)\}^{-1}\, d^{(T)}(\lambda) d^{(T)}(-\lambda),$$

where

$$d^{(T)}(\lambda) = \int_{\mathbb{R}} h^{(T)}(t) \exp(-i\lambda t)[dN(t) - \hat{p}^{(T)}\, dt]$$

is the finite Fourier transform of the point process, $h^{(T)}(t) = h(t/T)$ is a data taper with Fourier transforms

$$H_k^{(T)}(\lambda) = \int_{\mathbb{R}} h^{(T)}(t)^k \exp(-i\lambda t) dt$$

and

$$\hat{p}^{(T)} = \{H_1^{(T)}(0)\}^{-1} \int_{\mathbb{R}} h^{(T)}(t) dN(t)$$

is an estimate for the mean intensity p of the process. The taper function $h : \mathbb{R} \to \mathbb{R}$ is of bounded variation, vanishes outside the interval $[0, 1]$ and should be smooth with $h(0) = h(1) = 0$. However, our results also include the classical case where $h(t) = 1_{[0,1]}(t)$. We further define

$$H_k = \int_{\mathbb{R}} h(t)^k dt.$$

Substituting the periodogram $I^{(T)}$ for the spectral density, we obtain as an estimate for $A(\phi_\theta)$

$$A^{(T)}(\phi_\theta) = \int_{\mathbb{R}} \phi_\theta(\lambda) I^{(T)}(\lambda) d\lambda.$$

For finitely many θ such quadratic statistics have been studied e.g. by Brillinger (1972, 1978) and Tuan (1981). In these papers, the asymptotic normality of the estimate $A^{(T)}(\phi_\theta)$ has been derived for the nontapered case.

The present paper deals with the case where the parameter space Θ consists of infinitely many parameters. More generally, we establish a functional central limit theorem for the empirical spectral process

$$E_T^{(w)}(g) = \sqrt{T} \int_{\mathbb{R}} g(\lambda)[I^{(T)}(\lambda) - f_2(\lambda)] w(\lambda) d\lambda,$$

where $g : \mathbb{R} \to \mathbb{C}$ is from a suitable class of functions. The weight function w introduced for technical reasons should take values in $[0, 1]$ such that high frequencies are weighted down or cut off. If there exists a smooth function $\phi : \mathbb{R} \to \mathbb{R}$ with Fourier transform $\hat{\phi}$ such that $w(\lambda) = |\hat{\phi(\lambda)}|^2$, the product $f_2(\lambda)w(\lambda)$ can be viewed as the spectral density of the smoothed stochastic process $\int_{\mathbb{R}} \phi(t-u)dN(u)$.

In Section 2, we obtain our main result on the weak convergence of $E_T^{(w)}$ by proving the measurability and stochastic equicontinuity of $E_T^{(w)}$ and the weak convergence of its finite dimensional distributions. As in Dahlhaus (1988), we use a proof for the stochastic equicontinuity which is based on uniform bounds for the moments of the increments of $E_T^{(w)}$. However, our method of deriving these bounds is different as problems arise from the nonintegrability of point process spectra. The derivation is technical and therefore put into an appendix. The conditions for measurability are stated in Theorem 2.2. This result is also valid in the case of stationary time series.

In Section 3, we give some applications of these results to the statistical analysis of point processes. In particular, we discuss parametric and nonparametric spectral density estimates obtained by maximizing an approximation to the log likelihood function.

2 Weak convergence of the empirical spectral process

For some measurable function $w : \mathbb{R} \to \mathbb{R}$, let $\mathcal{L}^2_w(\mathbb{R})$ denote the space of all complex valued functions g on $\mathbb{R}$ for which the seminorm

$$\rho_w(g) = \Big(\int_{\mathbb{R}} |g(\lambda)|^2 w(\lambda) d\lambda \Big)^{1/2}$$

is finite. Further, if $\mathcal{F}$ is a subset of $\mathcal{L}^2_w(\mathbb{R})$, let $\mathcal{X}$ be the space of all bounded, complex valued functions on $\mathcal{F}$ which are uniformly continuous with respect to the seminorm. We equip $\mathcal{X}$ with the Borel-field $\mathcal{B}_{\mathcal{X}}$ generated by the open sets corresponding to the uniform norm $\|x\|_\infty = \sup |x(g)|$ for $x \in \mathcal{X}$.

Now, if the spectrum f_2 is bounded, it follows from the Cauchy-Schwarz inequality and the boundedness of $I^{(T)}$ that the sample paths $E_T^{(w)}(\omega, \cdot)$ are uniformly continuous with respect to ρ_w. Therein the empirical spectral processes differ from ordinary empirical processes, which in general have discontinuous sample paths. The difference is important as we make use of continuity for proving the measurability of $E_T^{(w)}$ with respect to $\mathcal{B}_{\mathcal{X}}$.

The limit process of $E_T^{(w)}$ for $T \to \infty$ is defined by its finite dimensional distributions. Therefore, we call a stochastic process $E_{f_2}^{(w)}$ a spectral process if its sample paths are in $\mathcal{X}$ almost surely and its finite dimensional distributions are normal with mean zero and

$$\operatorname{cov}\{E_{f_2}^{(w)}(g), E_{f_2}^{(w)}(h)\}$$

$$= \frac{2\pi H_4}{H_2^2} \int_{\mathbb{R}^2} g(\lambda) w(\lambda) \overline{h(\mu)} w(\mu) f_4(\lambda, -\lambda, \mu) d\lambda d\mu$$
$$+ \frac{2\pi H_4}{H_2^2} \int_{\mathbb{R}} g(\lambda) w(\lambda) \Big(\overline{h(\lambda)} w(\lambda) + \overline{h(-\lambda)} w(-\lambda) \Big) f_2(\lambda)^2 d\lambda, \quad (2.1)$$

where f_4 is the cumulant spectrum of order four of the point process N. The higher order cumulant spectra f_k and the corresponding reduced cumulant measures C_k' are defined analogously to (1.1) and (1.2), respectively [cf. Brillinger (1972)].

For the results in this paper, we need to impose conditions on the strength of the dependence of the data and on the size of the index class $\mathcal{F}$. The latter is determined by the covering number of $\mathcal{F}$, which we denote by

$$N(\delta, \rho_w, \mathcal{F}) = \inf\{m \in \mathbb{N} | \exists g_1, \ldots, g_m \in \mathcal{L}_w^2(\mathbb{R}) \, \forall g \in \mathcal{F} : \min_{1 \leq k \leq m} \rho_w(g - g_k) < \delta\}$$

[e.g. Pollard (1984)]. If $\mathcal{F}$ is a totally bounded subset of $\mathcal{L}_w^2(\mathbb{R})$, $N(\delta, \rho_w, \mathcal{F})$ is finite for all $\delta > 0$.

Assumptions

(A1) N is an orderly, stationary point process on $\mathbb{R}$ with finite mean intensity p and reduced cumulant measures C_k' such that there exists a constant C with

$$\int_{\mathbb{R}^{k-1}} (1 + |u_j|) |dC_k'(u_1, \ldots, u_{k-1})| \leq C^k$$

for all $j \in \{1, \ldots, k-1\}$ and $k \geq 2$.

(A2) $h : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable function of bounded variation with $h(x) = 0$ for all $x \notin [0, 1]$.

(A3) $w : \mathbb{R} \to \mathbb{R}$ is nonnegative, bounded and $\mathcal{L}^1$-integrable.

(A4) $\mathcal{F}$ is a totally bounded subset of $\mathcal{L}_w^2(\mathbb{R})$ such that for all $g \in \mathcal{F}$ the product $g \cdot w$ is bounded and the covering numbers of $\mathcal{F}$ satisfy

$$\int_0^1 [\log\{N(u, \rho_w, \mathcal{F})^2 / u\}]^2 du < \infty.$$

We now state our main theorem.

Theorem 2.1 *Suppose that Assumptions (A1) - (A4) hold. Then the empirical spectral process $E_T^{(w)}(g)$, $g \in \mathcal{F}$ converges weakly on $\mathcal{X}$ to the spectral process $E_{f_2}^{(w)}(g)$, $g \in \mathcal{F}$.*

Proof. We will prove the stochastic equicontinuity and the measurability of the empirical spectral process and the weak convergence of its finite dimensional distributions to that of $E_{f_2}^{(w)}$. Then the weak convergence of $E_T^{(w)}$ follows by Theorem 10.2

in Pollard (1990), in which the outer measure $\mathbb{P}^*$ can be replaced by the measure $\mathbb{P}$ due to the measurability of $E_T^{(w)}$. □

For the proof of the measurability of $E_T^{(w)}$, let $(\Omega, \mathcal{A}, \mathbb{P})$ be the underlying probability space.

Theorem 2.2 *Suppose that Assumptions (A1) - (A3) hold and let $\mathcal{F}$ be a totally bounded subset of $\mathcal{L}_w^2(\mathbb{R})$. Then the empirical spectral process $E_T^{(w)}(\cdot, g)$, $g \in \mathcal{F}$ is a measurable mapping into $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$.*

Proof. We first prove the measurability of the finite dimensional projections of the empirical spectral process. For fixed $g \in \mathcal{F}$, we have due to dominated convergence

$$E_T^{(w)}(\omega, g) = \lim_{k\to\infty} \int_{-k}^{k} g(\lambda)[I^{(T)}(\omega, \lambda) - f_2(\lambda)]w(\lambda)d\lambda$$

pointwise for all $\omega \in \Omega$. Thus, it suffices to show measurability of the integrals on the right hand side. For this, let $C[-k,k]$ denote the space of all continuous functions on $[-k,k]$ endowed with the topology of uniform convergence. Then the corresponding Borel-field $\mathcal{B}_{C[-k,k]}$ and the σ-field generated by the projections $\pi_t(x) = x(t)$ coincide. Hence the periodogram $I^{(T)}$ is a measurable mapping into $C[-k,k]$ since all projections $I^{(T)}(\lambda)$ are measurable. Further, it follows from the Cauchy-Schwarz inequality that the mapping

$$x \mapsto \sqrt{T} \int_{[-k,k]} g(\lambda)w(\lambda)[x(\lambda) - f_2(\lambda)]d\lambda$$

is continuous and thus $(\mathcal{B}_{C[-k,k]}, \mathcal{B}_{\mathbb{C}})$-measurable where $\mathcal{B}_{\mathbb{C}}$ is the Borel-field of $\mathbb{C}$. This now implies the $(\mathcal{A}, \mathcal{B}_{\mathbb{C}})$-measurability of the above integrals.

Now since $\mathcal{F}$ is totally bounded, $\mathcal{X}$ is separable and thus the measurability of the empirical spectral process follows from the uniform continuity of its sample paths and the measurability of its finite dimensional projections. □

Note that for the uniform continuity of the sample paths it is sufficient that the spectrum f_2 is bounded. Therefore, the assertion of the theorem holds also in the case of stationary time series under the assumptions stated in Dahlhaus (1988).

Theorem 2.3 *Suppose that Assumptions (A1) - (A4) hold. Then the empirical spectral process $E_T^{(w)}(g)$, $g \in \mathcal{F}$ is stochastically equicontinuous, i.e. for each $\eta > 0$ and $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\limsup_{T\to\infty} \mathbb{P}\{\sup_{[\delta]} |E_T^{(w)}(g-h)| > \eta\} < \varepsilon$$

where $[\delta] = \{(g,h) \in \mathcal{F}^2 \mid \rho_w(g-h) < \delta\}$.

The proof of Theorem 2.3 is technical and therefore put into the Appendix. For the next theorem, we only require finiteness of the integrals in Assumption (A1).

Theorem 2.4 *Suppose that Assumptions (A1) - (A3) hold. Further, let* $g_1, \ldots, g_k \in \mathcal{L}^2_w(\mathbb{R})$ *be such that the products* $g_j \cdot w$ *are bounded. Then*

$$\{E_T^{(w)}(g_j)\}_{j=1,\ldots,k} \xrightarrow{\mathcal{D}} \{E_{f_2}^{(w)}(g_j)\}_{j=1,\ldots,k}.$$

A similar central limit theorem has been proved in the nontapered case by Tuan (1981). However, in order to prove the same result in the tapered case, which is of much practical importance [e.g. Dahlhaus (1990)], we require the concept of $L^{(T)}$-functions, which are used to deal with data tapers [cf. Dahlhaus (1983, 1990)]. We therefore give a sketch of a proof, which is put into the Appendix, as we use the same techniques as in the proof of Theorem 2.3.

3 Application to the spectral analysis of point processes

In this section, we present some applications of the above results to the statistical analysis of point processes. Throughout this section, we assume that (A1) and (A2) hold.

Example 3.1 Let $\mathcal{F} = \{1_{[0,\lambda]} | \lambda \in [0, \lambda_0]\}$ and $w(\lambda) = 1_{[0,\lambda_0]}(\lambda)$. Then, $\mathcal{F}$ satisfies Assumption (A4) and we obtain a functional limit theorem for the empirical spectral distribution function on the interval $[0, \lambda_0]$. More generally, we can set $\mathcal{F} = \{1_D | D \in \mathcal{D}\}$ where $\mathcal{D}$ is a Vapnik-Cervonenkis class [e.g. Gänssler (1983), p 22] of subsets $D \subset [0, \lambda_0]$ to get the same result for the empirical spectral measure $\int_D f_2(\lambda) d\lambda$ for all $D \in \mathcal{D}$.

Example 3.2 For the estimation of the covariance density of a point process, we consider

$$\hat{q}_2^{(w)}(u) = \int_{\mathbb{R}} \exp(iu\lambda)[I^{(T)}(\lambda) - \hat{p}^{(T)}/2\pi] w(\lambda) d\lambda \tag{3.1}$$

with symmetric and smooth weight function w such that $w(0) = 1$ and $w(\lambda) = O(\{1 + |\lambda|\}^{-3-\epsilon})$ for some $\epsilon > 0$. Further, we define $q_2^{(w)}$ by (3.1) with f_2 and p substituted for $I^{(T)}$ and $\hat{p}^{(T)}$, respectively. Then, $q_2^{(w)}$ is related to the true covariance density by

$$q_2^{(w)}(u) = (2\pi)^{-1} \int_{\mathbb{R}} \hat{w}(u - v) q_2(v) dv, \tag{3.2}$$

where $\hat{w}$ is the Fourier transform of w. Thus, the weight function in the frequency domain corresponds to a smoothing kernel in the time domain.

Let $\mathcal{F} = \{g_u | u \in [0, u_0]\}$ with $g_u(\lambda) = \exp(iu\lambda)$. Since $\rho_w(g_u - g_v) \leq C|u - v|$ for some constant $C > 0$, the class $\mathcal{F}$ satisfies Assumption (A4) and therefore by Theorem 2.1 we obtain the functional convergence

$$\sqrt{T}(\hat{q}_2^{(w)}(u) - q_2^{(w)}(u)) = E_T^{(w)}(g_u) + \sqrt{T}\frac{\hat{w}(u)}{2\pi}(\hat{p}^{(T)} - p) \xrightarrow{\mathcal{D}} Z^{(w)}(u)$$

for $u \in [0, u_0]$, where $Z^{(w)}(u)$ is normally distributed with mean zero and covariance

$$\begin{aligned}
\operatorname{cov}\{Z^{(w)}(u), Z^{(w)}(v)\} &= \frac{2\pi H_4}{H_2^2} \int_{\mathbb{R}^2} e^{iu\lambda - iv\mu} w(\lambda) w(\mu) f_4(\lambda, -\lambda, \mu) d\lambda d\mu \\
&+ \frac{2\pi H_4}{H_2^2} \int_{\mathbb{R}} (e^{i(u+v)\lambda} + e^{i(u-v)\lambda}) w(\lambda)^2 f_2(\lambda)^2 d\lambda \\
&+ \frac{H_3}{H_1 H_2} \int_{\mathbb{R}} (e^{iu\lambda} \hat{w}(v) + e^{iv\lambda} \hat{w}(u)) w(\lambda) f_3(\lambda, -\lambda) d\lambda \\
&+ \frac{\hat{w}(u)\hat{w}(v)}{2\pi} f_2(0). \qquad (3.3)
\end{aligned}$$

As we can see now, the weight function w balances variance and smoothness of the estimate: As the bandwidth of the smoothing kernel in (3.2) increases, the weight function gets more concentrated and the variance decreases.

The above result can be used to derive a simultaneous confidence band for $\hat{q}_2^{(w)}(u)$, $u \in [0, u_0]$. Application of the continuous mapping theorem yields

$$\sqrt{T} \sup_{u \in [0,u_0]} |\hat{q}_2^{(w)}(u) - q_2^{(w)}(u)| \xrightarrow{\mathcal{D}} \sup_{u \in [0,u_0]} |Z^{(w)}(u)|.$$

Then

$$\left\{ \hat{q}_2^{(w)}(u) \pm T^{-1/2} z_\alpha^{(w)} \right\}_{u \in [0,u_0]} \qquad (3.4)$$

is an asymptotic simultaneous confidence band, where $z_\alpha^{(w)}$ denotes the upper $\alpha \cdot 100$ percentile point of the limit statistic $\sup_{u \in [0,u_0]} |Z^{(w)}(u)|$. The problem now is to obtain the distribution of $\sup_{u \in [0,u_0]} |Z^{(w)}(u)|$, which appears to be extremely difficult. However, if we generate realisations $Z_1^{(w)}, \ldots, Z_B^{(w)}$ of the process $Z^{(w)}$ on a suitably fine grid, we can use the empirical distribution of $\sup_{u \in [0,u_0]} |Z_b^{(w)}(u)|$, $b = 1, \ldots, B$ as an approximation. For this, we have to estimate the covariances (3.3). For the first integral, a consistent estimator has been presented by Taniguchi (1982), the other integrals can be estimated similarly.

As an illustration, we apply this method to some data which describe the state of activity of a computer: An event has been recorded whenever the computer changed its state from busy to idle which was defined by the absence of any user interaction for more than five minutes. The data set consists of 1539 events which occured in an interval of length $T = 1695236\,\mathrm{s}$ (about 20 days).

Figure 1 shows the covariance density estimate $\hat{q}_2^{(w)}$ and the corresponding simultaneous confidence band (3.4) with $\alpha = 0.05$ for the data. Here, the weight function has been set to $w(\lambda) = \exp(-\lambda^2\sigma^2/2)$ with $\sigma = 500\,\mathrm{s}$, which corresponds to a Gaussian smoothing kernel. The covariance density exhibits two significant positive peaks: One for time lags smaller than 3 h and another for a delay of about 7 d. The former peak indicates that the process tends to form clusters, while the latter suggests some weekly structure in the data.

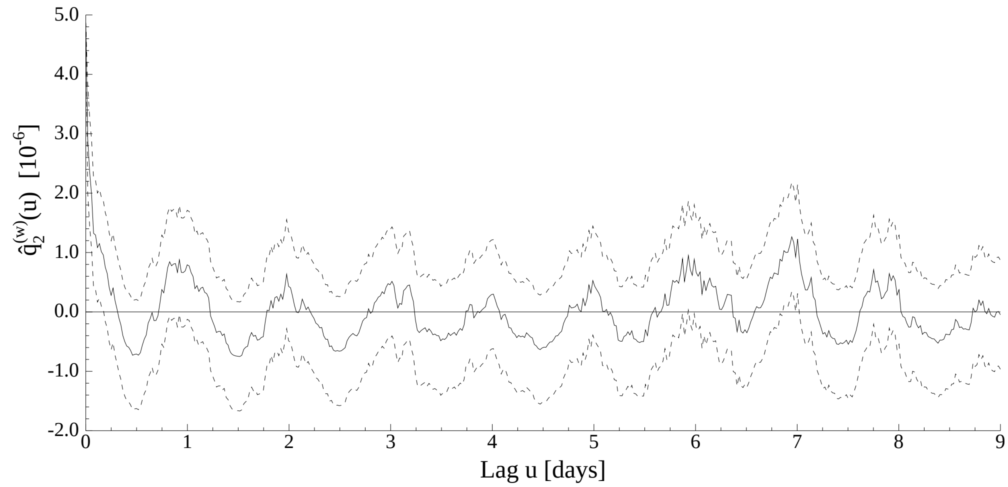

Figure 1: Estimated covariance density $\hat{q}_2^{(w)}$ (solid line) with simultaneous 5 % confidence bands (dashed lines) for the computer data.

We now turn to the problem of spectral density estimation. Suppose that N is a stationary point process with spectral density f_2^*. Given a realization of the process on the interval $[0, T]$, we want to fit a spectral density $f \in \mathcal{F}_0$ to the data. It is well known [e.g. Brillinger (1972)] that the random variables $I^{(T)}(2\pi j/T)$ are asymptotically independent and exponentially distributed with mean $f_2^*(2\pi j/T)$. This suggests approximating the log likelihood function by

$$\mathcal{L}_M^{(T)}(f) = -\frac{1}{T}\sum_{j=0}^{M}\Big\{\log f(2\pi j/T) + \frac{I^{(T)}(2\pi j/T)}{f(2\pi j/T)}\Big\}$$

and estimating f_2^* by maximizing $\mathcal{L}_M^{(T)}(f)$ with respect to $f \in \mathcal{F}_0$. This approach has been proposed by Hawkes and Adamopoulos (1973) and further discussed by Brillinger (1975) and Tuan (1981) for parametric families, for which the procedure is a point process version of a procedure suggested by Whittle (1953) for the analysis of time series.

Subsequently, we will use the following continuous version of $\mathcal{L}_M^{(T)}$

$$\mathcal{L}^{(T)}(f) = -\int_{\mathbb{R}}\Big\{\log f(\lambda) + \frac{I^{(T)}(\lambda)}{f(\lambda)}\Big\}w(\lambda)d\lambda \tag{3.5}$$

with $w : \mathbb{R} \to \mathbb{R}$ satisfying Assumption (A3) and $w(\lambda) = 0$ for all $\lambda < 0$. Let $f^{(T)}$ denote a sequence of functions maximizing $\mathcal{L}^{(T)}$. Substituting f_2^* in (3.5) for $I^{(T)}$, we obtain the corresponding theoretical function $\mathcal{L}(f)$, which is maximized by f_2^*. The next theorem states the consistency of $f^{(T)}$ as an estimate for f_2^*.

Theorem 3.3 *Let $\mathcal{F}_0$ be a subset of $\mathcal{L}_w^2(\mathbb{R})$ with envelope $F \in \mathcal{L}_w^2(\mathbb{R})$ and $f_2^* \in \mathcal{F}_0$. Then, if $\mathcal{F} = \{f^{-1} | f \in \mathcal{F}_0\}$ satisfies Assumption (A4),*

$$\rho_w(f^{(T)} - f_2^*) \xrightarrow{P} 0 \quad as \quad T \to \infty.$$

Proof. As in Example 3.2 in Dahlhaus (1988), we obtain from Theorem 2.1 and the Continuous Mapping Theorem [cf. Pollard (1984), Theorem IV.12]

$$\sup_{f\in\mathcal{F}_0} |\mathcal{L}^{(T)}(f) - \mathcal{L}(f)| = \sup_{f\in\mathcal{F}_0} |T^{-1/2} E_T^{(w)}(f^{-1})| \xrightarrow{P} 0,$$

which implies $\mathcal{L}(f^{(T)}) - \mathcal{L}(f^*) \xrightarrow{P} 0$. Now if $\mathcal{L}(f_n)$ converges to $\mathcal{L}(f_2^*)$ for a deterministic sequence f_n, we obtain with a Taylor expansion that $\rho_w(f_n - f_2^*) \to 0$, which proves the result. □

Uniform convergence of $f^{(T)}$ to f_0 requires further assumptions about $\mathcal{F}_0$. If, for example, $\mathcal{F}_0$ is equicontinuous, then for any compact set $K \subset \mathbb{R}$

$$\sup_{\lambda\in K} |f^{(T)}(\lambda) - f_2^*(\lambda)| \xrightarrow{P} 0 \quad \text{as} \quad T \to \infty.$$

In the next example, we present an explicit nonparametric function class which satisfies the requirements of Theorem 3.3.

Example 3.4 Consider the class $\mathcal{F}_{r,\alpha}(S) = \mathcal{F}_{r,\alpha}(S; c_0, \ldots, c_r, c)$ of smooth functions f on $S \subset \mathbb{R}$ such that

$$|f^{(i)}(x)| \le c_i$$

for $0 \le i \le r$ and

$$|f^{(r)}(x) - f^{(r)}(y)| \le c|x-y|^\alpha.$$

Further, suppose that $w(x) \le C(1+|x|)^{-\gamma}$ where $\gamma > 2(r+\alpha)+1$ and $r+\alpha > 2$. Then $\mathcal{F}_{r,\alpha}(\mathbb{R}^+)$ fulfills Assumption (A4). This can be seen by constructing an ε-covering of $\mathcal{F}_{r,\alpha}(\mathbb{R}^+)$ from ε_k-coverings of $\mathcal{F}_{r,\alpha}(I_k)$, $I_k = [k, k+1)$ with $k \le k_*$ for some large $k_* \in \mathbb{N}$. Using the uniform norm on I_k, the entropy of $\mathcal{F}_{r,\alpha}(I_k)$ is of order $O(\varepsilon_k^{-1/(r+\alpha)})$ [cf. Kolmogorov and Tikhomirov (1961)]. If we choose $\varepsilon_k = \varepsilon k^\beta$ with $r+\alpha < \beta < (\gamma-1)/2$ and k_* such that $1/2 \cdot \varepsilon^{-2/(\gamma-1)} \le k_* \le \varepsilon^{-1/\beta}$, then ε_k increases sufficiently fast to guarantee

$$\log N(\varepsilon, \rho_w, \mathcal{F}_{r,\alpha}(\mathbb{R}^+)) \le J\varepsilon^{-\frac{1}{r+\alpha}},$$

Now the function class $\mathcal{F} = \{f^{-1} | f \in \mathcal{F}_{r,\alpha}(\mathbb{R}^+),\ f \ge c\}$ with $c > 0$ is itself a subset of some class $\tilde{\mathcal{F}}_{r,\alpha}(\mathbb{R}^+)$ with different constants, and it therefore also fulfills Assumption (A4).

A stronger result than Theorem 3.3 can be obtained in the case where we fit a parametric model given by the class of spectral densities $\mathcal{F}_0 = \{f_\theta | \theta \in \Theta\}$. When dealing with parametric estimation, a point of view is to regard the parametric model only as an approximation to the true process. Therefore, we do not assume that the true spectral density f_2^* belongs to the model class $\mathcal{F}_0$. If we measure the distance between a fitted model specified by θ and the true process by $-\mathcal{L}(\theta)$, the parameter θ_0 which maximizes $\mathcal{L}(\theta)$ then determines the best approximation to the true process. By maximizing $\mathcal{L}^{(T)}(\theta)$, we obtain an estimate $\hat{\theta}^{(T)}$ for θ_0.

Assumptions Let $\mathcal{F}_0 = \{f_\theta | \theta \in \Theta\}$ be a parametric family.

(B1) $\mathcal{L}(\theta)$ has an unique maximum θ_0 which is an interior point of $\Theta \subseteq \mathbb{R}^d$.

(B2) There exists a compact subset Θ_* of Θ such that

$$\liminf_{T\to\infty} \mathbb{P}\{\mathcal{L}(\theta_0) - \inf_{\theta\in\Theta_*^C} \mathcal{L}^{(T)}(\theta) > \varepsilon_0\} = 1$$

for some $\varepsilon_0 > 0$. Further, the functions $f_\theta(\lambda)$ are continuous in $(\theta, \lambda) \in \Theta_* \times \mathbb{R}$ and there exist constants c_1, c_2 such that $0 < c_1 \leq f_\theta(\lambda) \leq c_2 < \infty$ for all $\theta \in \Theta_*$ and $\lambda \in \mathbb{R}$.

(B3) $\mathcal{F} = \{f_\theta^{-1} | \theta \in \Theta_*\}$ satisfies Assumption (A4).

(B4) f_θ^{-1} admits continuous first and second derivatives with respect to θ in a neighborhood $U(\theta_0)$ of θ_0, denoted by the vector ∇f_θ^{-1} and the matrix $\nabla^2 f_\theta^{-1}$, respectively. The families $\{\nabla f_\theta^{-1} | \theta \in U(\theta_0)\}$ and $\{\nabla^2 f_\theta^{-1} | \theta \in U(\theta_0)\}$ satisfy Assumption (A4).

Theorem 3.5 *Assume (B1) to (B3). Then we have $\hat{\theta}^{(T)} \xrightarrow{P} \theta_0$. If additionally (B4) holds, then*

$$\sqrt{T}(\hat{\theta}^{(T)} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W_{\theta_0}^{-1} \Sigma_{\theta_0} W_{\theta_0}^{-1}),$$

where

$$\begin{aligned} \Sigma_{\theta_0} &= 2\pi H_4 H_2^{-2} \int_{\mathbb{R}^2} \nabla f_{\theta_0}^{-1}(\lambda) \nabla f_{\theta_0}^{-1}(\mu)' f_4^*(\lambda, -\lambda, \mu) w(\lambda) w(\mu) d\lambda d\mu \\ &\quad + 2\pi H_4 H_2^{-2} \int_{\mathbb{R}} \nabla f_{\theta_0}^{-1}(\lambda) \nabla f_{\theta_0}^{-1}(\lambda)' f_2^*(\lambda)^2 w(\lambda)^2 d\lambda \end{aligned}$$

and

$$W_{\theta_0} = \int_{\mathbb{R}} (f_2^*(\lambda) - f_{\theta_0}(\lambda)) \nabla^2 f_{\theta_0}^{-1}(\lambda) w(\lambda) d\lambda + \int_{\mathbb{R}} \nabla \log f_{\theta_0}(\lambda) \nabla \log f_{\theta_0}(\lambda)' w(\lambda) d\lambda.$$

Proof. The result follows directly from Theorem 2.1. The proof is similar to that in Dahlhaus (1988). □

Example 3.6 Tuan (1981) suggested to approximate the spectral density by rational functions of the form

$$f_\theta(\lambda) = \frac{p_\theta(\lambda)}{q_\theta(\lambda)} = \frac{p}{2\pi} \cdot \frac{\lambda^n + a_1 \lambda^{n-1} + \ldots + a_n}{\lambda^n + b_1 \lambda^{n-1} + \ldots + b_n}$$

where $\theta = (p, a_1, \ldots, a_n, b_1, \ldots, b_n)$ is the unknown parameter. Suppose the parameter space Θ is compact. Then if p_θ and q_θ have no zeros in $\mathbb{R}^+$, the parametric class $\mathcal{F}_0$ satisfies Assumptions (B1) - (B4).

To motivate this remark, consider functions of the form $g(x, y) = p(x) - yq(x)$ where p and q are polynomials of fixed order. Since these functions form a finite dimensional vector space, the class of sets $\{(x, y) | g(x, y) \geq 0\}$ is a Vapnik-Cervonenkis (VC) class (cf. Pollard (1984), Lemma II.18). Then, the graphs $G(f_\theta^{-1})$ form also a VC-class and the Approximation Lemma (cf. Pollard (1984), Lemma II.25) yields (B3). The conditions on the derivatives are checked similarly.

Appendix

A key role in our proof of Theorem 2.3 is played by the function $L^{(T)} : \mathbb{R} \to \mathbb{R}$, $T \in \mathbb{R}^+$ which is given by

$$L^{(T)}(\alpha) := \begin{cases} T, & |\alpha| \leq 1/T \\ 1/|\alpha|, & |\alpha| > 1/T \end{cases} . \tag{A.1}$$

A similar, but periodic function was introduced in Dahlhaus (1983) as a tool for handling the cumulants of discrete time series statistics. The above function $L^{(T)}$ is the corresponding version for time-continuous stochastic processes. Its properties are summarized in the following lemma.

Lemma A.1 *Suppose w fulfills Assumption (A3). Further, let $\alpha, \beta, \gamma, \zeta \in \mathbb{R}$ and $p \in \mathbb{N}$. We obtain with constants K_p independent of T:*
(i) $L^{(T)}(\alpha)$ *is monotone increasing in* $T \in \mathbb{R}^+$ *and decreasing in* $\alpha \in \mathbb{R}^+$.
(ii) $L^{(T)}(c\alpha) \leq c^{-1} L^{(T)}(\alpha)$ *for all* $c \in (0, 1]$.
(iii) $L^{(T)}(\beta+\alpha)L^{(T)}(\gamma-\alpha) \leq L^{(T)}(\frac{\beta+\gamma}{2})L^{(T)}(\gamma-\alpha) + L^{(T)}(\beta+\alpha)L^{(T)}(\frac{\beta+\gamma}{2})$.
(iv) $\int_{\mathbb{R}} L^{(T)}(\alpha) w(\zeta+\alpha) d\alpha \leq K_1 \log(T)$ *for* $T \geq e$.
(v) $\int_{\mathbb{R}} L^{(T)}(\beta+\alpha)L^{(T)}(\gamma-\alpha) w(\zeta+\alpha) d\alpha \leq K_1 \log(T) L^{(T)}(\beta+\gamma)$ *for* $T \geq e$.
(vi) $\int_{\mathbb{R}} L^{(T)}(\alpha)^p d\alpha \leq K_p T^{p-1}$.
(vii) $\int_{\mathbb{R}} L^{(T)}(\beta+\alpha)^p L^{(T)}(\gamma-\alpha)^p d\alpha \leq K_p T^{p-1} L^{(T)}(\beta+\gamma)^p$.

PROOF. The proofs are straightforward and similar to those in Dahlhaus (1983). But unlike its periodic counterpart, the function $L^{(T)}$ is not $\mathcal{L}^1$-integrable, and the inequalities (iv) and (v) therefore require an $\mathcal{L}^1$-integrable weight function w. □

Using the definition of $L^{(T)}$, we can now derive an upper bound for the Fourier transform of a data taper. Let $V(h)$ denote the total variation of the function h. Then if h is of bounded variation, simple calculations yield the inequality

$$\int_{\mathbb{R}} |h(t+u_1)\cdots h(t+u_k) - h(t)^k| dt \leq \|h\|_\infty^{k-1} V(h)(|u_1| + \ldots + |u_k|). \tag{A.2}$$

From this, we obtain for the Fourier transform of $h^{(T)}$

$$|H_k^{(T)}(\alpha)| \leq 1/2 \cdot \int_{\mathbb{R}} |h(t)^k - h(t-\pi/\alpha)^k| dt \leq 1/2 \cdot \|h\|_\infty^{k-1} V(h) k\pi |\alpha|^{-1}.$$

On the other hand, we have $|H_k^{(T)}(\alpha)| \leq \|h\|_\infty^k T$. Hence, we obtain as an upper bound

$$|H_k^{(T)}(\alpha)| \leq K^k L^{(T)}(\alpha) \tag{A.3}$$

for all $k \in \mathbb{N}$ and a constant K independent of T.

For sake of simplicity, we assume for the proof of Theorem 2.3 that the mean intensity p is known and therefore replace $d^{(T)}(\lambda)$ by

$$\tilde{d}^{(T)}(\lambda) = \int_{\mathbb{R}} h^{(T)}(t) \exp(-i\lambda t)[dN(t) - p\, dt].$$

At the end of the proof, we indicate the modifications needed for the case that p is estimated by $\hat{p}^{(T)}$.

The next lemma, which is a slightly stronger version of Theorem 4.1 in Brillinger (1972), gives an approximation for the cumulants of $\tilde{d}^{(T)}(\lambda)$.

Lemma A.2 *Under Assumptions (A1) and (A2), there exists a constant $c > 0$ such that for all $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $k \geq 2$*

$$|\mathrm{cum}\{\tilde{d}^{(T)}(\lambda_1), \ldots, \tilde{d}^{(T)}(\lambda_k)\} - (2\pi)^{k-1} H_k^{(T)}(\lambda_1 + \cdots + \lambda_k) f_k(\lambda_1, \ldots, \lambda_{k-1})| \leq c^k.$$

Proof. The lemma is an immediate consequence of relation (A.2) and Assumption (A1). □

Note that because of Assumption (A1) we can choose the constant c such that also $|f_k(\lambda_1, \ldots, \lambda_{k-1})| \leq c^k$ for all $k \geq 2$.

Proof of Theorem 2.3. We start by proving that there exists a constant c_0 such that

$$|\mathrm{cum}_k\{E_T^{(w)}(g)\}| \leq (2k)! c_0^k \rho_w(g)^k \tag{A.4}$$

for all $K \in \mathbb{N}$ where $\mathrm{cum}_k\{E_T^{(w)}(g)\}$ denotes the k-th cumulant of $E_T^{(w)}(g)$. Using Lemma A.2 we obtain

$$\begin{aligned} |\mathrm{cum}_1\{E_T^{(w)}(g)\}| &\leq \sqrt{T} \int_{\mathbb{R}} |g(\lambda)| w(\lambda) |\mathbb{E} I^{(T)}(\lambda) - f_2(\lambda)| d\lambda \\ &\leq c^2 \{2\pi H_2 \sqrt{T}\}^{-1} \rho_w(g) \rho_w(1). \end{aligned} \tag{A.5}$$

For $k \geq 2$, we find

$$\begin{aligned} &|\mathrm{cum}_k\{E_T^{(w)}(g)\}| \\ &\quad \leq [2\pi H_2^{(T)}(0)]^{-k} T^{k/2} \int_{\mathbb{R}^k} |g(\lambda_1)| w(\lambda_1) \cdots |g(\lambda_k)| w(\lambda_k) \\ &\qquad \times |\mathrm{cum}\{\tilde{d}^{(T)}(\lambda_1)\tilde{d}^{(T)}(-\lambda_1), \ldots, \tilde{d}^{(T)}(\lambda_k)\tilde{d}^{(T)}(-\lambda_k)\}| d\lambda_1 \cdots d\lambda_k. \end{aligned} \tag{A.6}$$

In order to apply the product theorem for cumulants [cf. Brillinger (1981), Theorem 2.3.2], let $\sum_{i.p.}$ denote the sum over all indecomposable partitions $P_1, \ldots, P_m$ of the table

$$\begin{array}{cc} \lambda_1 & -\lambda_1 \\ \vdots & \vdots \\ \lambda_k & -\lambda_k \end{array}$$

with $p_j = |P_j| \geq 2$ as $\mathbb{E}\{\tilde{d}^{(T)}(\lambda)\} = 0$. Further if $P_j = \{\beta_{j,1}, \ldots, \beta_{j,p_j}\}$, we write $\bar{\beta}_j = \beta_{j,1} + \ldots + \beta_{j,p_j}$. Then using Lemma A.2, we find for the cumulant in (A.6)

$$
\begin{aligned}
&|\mathrm{cum}\{\tilde{d}^{(T)}(\lambda_1)\tilde{d}^{(T)}(-\lambda_1), \ldots, \tilde{d}^{(T)}(\lambda_k)\tilde{d}^{(T)}(-\lambda_k)\}| \\
&\quad \leq \sum_{i.p.} \prod_{j=1}^{m} \Big\{(2\pi)^{p_j-1}|H_{p_j}^{(T)}(\bar{\beta}_j)||f_{p_j}(\beta_{j,1}, \ldots, \beta_{j,p_j-1})| + c^{p_j}\Big\},
\end{aligned}
$$

which, by using $f_k(\lambda_1, \ldots, \lambda_{k-1}) \leq c^k$ and (A.3), is less than

$$
\sum_{i.p.} (2\pi)^{2k} c^{2k} K^{2k} \sum_{J \subseteq M} \prod_{j \in J} L^{(T)}(\bar{\beta}_j) \tag{A.7}
$$

where $M = \{1, \ldots, m\}$. Substituting (A.7) in (A.6), we obtain as an upper bound

$$
\sum_{i.p.} \sum_{J \subseteq M} (2\pi c^2 K^2 H_2^{-1})^k \, T^{-k/2} \int_{\mathbb{R}^k} \prod_{j=1}^{k} |g(\lambda_j)| w(\lambda_j) \prod_{j \in J} L^{(T)}(\bar{\beta}_j) d\lambda_1 \cdots d\lambda_k.
$$

For $J = \emptyset$, the integral is equal to

$$
\Big(\int_{\mathbb{R}^k} |g(\lambda)| w(\lambda) d\lambda\Big)^k \leq \rho_w(g)^k \rho_w(1)^k.
$$

Similarly, if $m = 1$ and $J = \{1\}$, the integral is bounded by $T\rho_w(g)^k\rho_w(1)^k$ since $L^{(T)}(\bar{\beta}_1) = T$. For $J \neq \emptyset$ and $m > 1$, we split $\{1, \ldots, k\}$ into disjoint sets I and I^C and J into disjoint sets J_0 and J_0^C to be selected later. Using the Cauchy-Schwarz inequality, the integral now is less than

$$
\begin{aligned}
&\Big(\int_{\mathbb{R}^k} \prod_{j \in I} |g(\lambda_j)|^2 w(\lambda_j) \prod_{j \in I^C} w(\lambda_j) \prod_{j \in J_0^C} L^{(T)}(\bar{\beta}_j)^2 d\lambda_1 \cdots d\lambda_k\Big)^{1/2} \\
&\quad \times \Big(\int_{\mathbb{R}^k} \prod_{j \in I^C} |g(\lambda_j)|^2 w(\lambda_j) \prod_{j \in I} w(\lambda_j) \prod_{j \in J_0} L^{(T)}(\bar{\beta}_j)^2 d\lambda_1 \cdots d\lambda_k\Big)^{1/2}.
\end{aligned} \tag{A.8}
$$

Now we have to make a suitable choice for I, I^C, J_0 and J_0^C. Since J is not empty, we can define $J_0 = \{j_0\}$ for some arbitrary $j_0 \in J$. Then there exists i_0 such that λ_{i_0} or $-\lambda_{i_0}$ is in P_{j_0}, and we can set $I = \{i_0\}$. We obtain with Lemma A.1 (vi) and (vii) for the first integral in (A.8)

$$
\begin{aligned}
&\int_{\mathbb{R}} |g(\lambda_{i_0})|^2 w(\lambda_{i_0}) \Big\{\int_{\mathbb{R}^{k-1}} \prod_{i \in I^C} w(\lambda_i) \prod_{j \in J_0^C} L^{(T)}(\bar{\beta}_j)^2 \prod_{i \in I^C} d\lambda_i\Big\} d\lambda_{i_0} \\
&\quad \leq \int_{\mathbb{R}} |g(\lambda_{i_0})|^2 w(\lambda_{i_0}) \Big\{\|w\|_\infty^{|J|-1} \rho_w(1)^{2(k-|J|)} (KT)^{|J|-1}\Big\} d\lambda_{i_0} \\
&\quad = \rho_w(g)^2 \Big\{\|w\|_\infty^{|J|-1} \rho_w(1)^{2(k-|J|)} (KT)^{|J|-1}\Big\}
\end{aligned}
$$

where we have used the indecomposability of the partition. Similarly, the second integral in (A.8) is equal to

$$\int_{\mathbb{R}^{k-1}} \prod_{j\in I^C} |g(\lambda_j)|^2 w(\lambda_j)\Big\{ \int_{\mathbb{R}} w(\lambda_{i_0}) L^{(T)}(\bar{\beta}_{j_0})^2 d\lambda_{i_0}\Big\} \prod_{i\in I^C} d\lambda_i$$
$$\leq \; \rho_w(g)^{2(k-1)}\Big\{\|w\|_\infty^2 K\, T\Big\}.$$

Thus, we obtain as an upper bound for $\mathrm{cum}_k\{E_T^{(w)}\}$

$$\sum_{i.p.}\sum_{J\subseteq M} (2\pi c^2 K^2 H_2^{-1})^k T^{(|J|-k)/2}\rho_w(g)^k \rho_w(1)^k \|w\|_\infty^{k/2} K^{|J|/2}$$
$$\leq \; \rho_w(g)^k \sum_{i.p.}\sum_{J\subseteq M} (c_0/4)^k,$$

which implies (A.4) since the sums include at most 2^k subsets of M and $(2k)!\,2^k$ indecomposable partitions.

Analogously to the proof of the stochastic equicontinuity in Dahlhaus (1988), relation (A.4) leads to an uniform bound for the moments of $E_T^{(w)}$, from which we obtain the exponential inequality

$$\mathbb{P}\{|E_T^{(w)}(g-h)| > \eta\rho_w(g-h)\} \leq 96\exp(-\sqrt{\eta/D})$$

with a constant D for all $g,h\in\mathcal{F}$ and $\eta>0$. Application of a chaining argument now yields the assertion of the theorem.

In the general case where p is estimated by $\hat{p}^{(T)}$, we have

$$d^{(T)}(\lambda) = \tilde{d}^{(T)}(\lambda) - H_1^{(T)}(\lambda) H_1^{(T)}(0)^{-1}\tilde{d}^{(T)}(0). \tag{A.9}$$

Therefore, we obtain by using Lemma A.2

$$\begin{aligned} &\mathrm{cum}\{d^{(T)}(\lambda_1),\ldots,d^{(T)}(\lambda_k)\} \\ &= \sum_{S\subseteq\{1,\ldots,k\}} (-1)^{|S|} H_1^{(T)}(0)^{-|S|}\prod_{j\in S} H_1^{(T)}(\lambda_j) \\ &\qquad \times (2\pi)^{k-1} H_k^{(T)}(\lambda_{S,1}+\ldots+\lambda_{S,k}) f_k(\lambda_{S,1},\ldots,\lambda_{S,k-1}) + R, \end{aligned}$$

where $|R|\leq (2c)^k$ uniformly in $\lambda_1,\ldots,\lambda_k$ and $\lambda_{S,j}=\lambda_j\, 1_{S^C}(j)$. Because of Lemma A.1 (ii) and (iii), the sum over all subsets S of $\{1,\ldots,k\}$ is bounded by

$$(2\pi)^{k-1} H_1^{-k}(4c)^k K^{2k} L^{(T)}(\lambda_1+\ldots+\lambda_k).$$

Thus, the cumulant in (A.6) with $d^{(T)}$ substituted for $\tilde{d}^{(T)}$ has again an upper bound of the form (A.7) with different constants. The case $k=1$ is treated similarly. □

Proof of Theorem 2.4. We prove the convergence of the cumulants of first, second and higher order to the corresponding cumulants of the limit distribution.

For the first cumulant, $\mathbb{E}\{E_T^{(w)}(g)\} = o(1)$ is an immediate consequence of (A.5) and Lemma A.1 since (A.9) yields

$$\mathrm{cum}\{d^{(T)}(\lambda), d^{(T)}(\mu)\} = \mathrm{cum}\{\tilde{d}^{(T)}(\lambda), \tilde{d}^{(T)}(\mu)\} + O(T^{-1}L^{(T)}(\lambda)L^{(T)}(\mu)).$$

A similar equation with remainder of order $O(L^{(T)}(\lambda)+L^{(T)}(\mu))$ holds for the cumulant $\mathrm{cum}\{d^{(T)}(\lambda), d^{(T)}(-\lambda), d^{(T)}(\mu), d^{(T)}(-\mu)\}$. Thus we find by using the product theorem for cumulants and Lemma A.2

$$\begin{aligned}
&\mathrm{cov}\{E_T^{(w)}(g_i), E_T^{(w)}(g_j)\} \\
&\quad = \frac{2\pi H_4}{H_2^2}\int_{\mathbb{R}^2} g_i(\lambda)w(\lambda)\overline{g_j(\mu)}w(\mu)\Big[f_4(\lambda,-\lambda,\mu) \\
&\qquad + \Big(\Phi_2^{(T)}(\lambda+\mu)+\Phi_2^{(T)}(\lambda-\mu)\Big)f_2(\lambda)f_2(\mu)\Big]d\lambda d\mu + o(1), \quad \text{(A.10)}
\end{aligned}$$

where $\Phi_2^{(T)}(\lambda) = H_2^{(T)}(\lambda)H_2^{(T)}(-\lambda)/[2\pi H_4^{(T)}(0)]$. Now, it follows from the convolution properties of $H_k^{(T)}$ and Lemma A.1 (vi) that $\Phi_2^{(T)}$ is an approximate identity and thus we have for $f, g \in \mathcal{L}_w^2(\mathbb{R})$

$$\int_{\mathbb{R}^2} f(\lambda)w(\lambda)g(\mu)w(\mu)\Phi_2^{(T)}(\lambda-\mu)d\lambda d\mu = \int_{\mathbb{R}} f(\lambda)g(\lambda)w(\lambda)^2 d\lambda + o(1).$$

Therefore, (A.10) converges to (2.1).

For the cumulants of higher order, we obtain by using (A.7), which holds also for $d^{(T)}$ as noted before,

$$\begin{aligned}
&|\mathrm{cum}\{E_T^{(w)}(g_1),\ldots,E_T^{(w)}(g_k)\}| \\
&\quad \leq \sum_{i.p.}\sum_{J\subseteq M} O(T^{-k/2})\int_{\mathbb{R}^k}\prod_{j=1}^{k}|g_j(\lambda_j)|w(\lambda_j)\prod_{j\in J}L^{(T)}(\bar{\beta}_j)d\lambda_1\cdots d\lambda_k.
\end{aligned}$$

With Lemma A.1 (iv) and (v), the integral now is of order $O(T\log(T)^{|J|-1})$ if $|J| = m$ and of order $O(\log(T)^{|J|})$ if $|J| < m$. Thus, the cumulant converges to zero. $\square$

Acknowledgements

This paper is part of the author's Diplom-thesis at the University of Heidelberg. The author would like to thank his advisor Professor R. Dahlhaus for the problem and his support.

References

Brillinger, D.R. (1972), The spectral analysis of stationary interval functions, in *Proc. 6th Berkeley Symp.* Vol. 1, pp. 483-513, Berkeley, California.

Brillinger, D.R. (1975), Statistical inference for stationary point processes, in *Stochastic Processes and Related Topics* (M.L. Puri, ed.) Vol. 1, pp. 55-99, Academic Press, New York.

Brillinger, D.R. (1978), Comparative aspects of the study of ordinary time series and of point processes, in *Developments in Statistics* (P.R. Krishnaiah, ed.) Vol. 1, pp. 33-133, Academic Press, New York.

Brillinger, D.R. (1981), *Time Series: Data Analysis and Theory*, Mc Graw Hill, New York.

Dahlhaus, R. (1983), Spectral analysis with tapered data, *J. Time Ser. Anal.* **4**, 163-174.

Dahlhaus, R. (1988), Empirical spectral processes and their applications to time series analysis, *Stoch. Proc. Appl.* **30**, 69-83.

Dahlhaus, R. (1990), Nonparametric High Resolution Spectral Estimation, *Probab. Th. Rel. Fields* **85**, 147-180.

Daley, D.J. and Vere-Jones, D. (1988), *An Introduction to the Theory of Point Processes*, Springer, New York.

Gänssler, P. (1983), *Empirical Processes*, Institute of Mathematical Statistics Lecture Notes - Monograph Series 3.

Hawkes, A.G. and Adamopoulos, L. (1973), Cluster models for earthquakes - regional comparisons, *Bull. Inst Internat. Statist.* **45**, 454-461.

Kolmogorov, A.N. und Tikhomirov, V.M. (1961), ε-entropy and ε-capacity of sets in function spaces, *Amer. Math. Soc. Transl. (2)* **17**, 277-364.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer, New York.

Pollard, D. (1990), *Empirical Processes: Theory and Application*, NSF-CBMS Regional Conference Series in Probability and Statistics Vol. 2, SIAM/IMS, Philadelphia.

Taniguchi, M. (1982), On estimation of the integrals of the forth order cumulant spectral density, *Biometrika* **69**, 117-122.

Tuan, P.D. (1981), Estimation of the spectral parameters of a stationary point process, *Ann. Statist.* **9**, 615-627.

Whittle, P. (1953), Estimation and information in stationary time series, *Ark. Mat.* **2**, 423-434.