Statistica Sinica **23** (2013), 691-724
doi:http://dx.doi.org/10.5705/ss.2011.284

NONPARAMETRIC SPECTRUM ESTIMATION UNDER THE CONSTRAINT OF A MINIMUM INTER-SAMPLE SPACING

Radhendushka Srivastava and Debasis Sengupta

Cornell University and Indian Statistical Institute

Abstract: A minimum separation between successive samples is a practical constraint that often comes in the way of sampling of a continuous time stationary stochastic process for the purpose of spectrum estimation. It is known from a recent study that additive random sampling subject to the said constraint can be alias-free for bandlimited spectra with any specified support, but known estimation approaches do not work. In this paper, we propose a new spectrum estimator for this purpose and show that it can accurately and precisely estimate any power spectral density limited to an arbitrarily large but known support.

Key words and phrases: additive random sampling, bandwidth, smoothing, spectral density.

1. Introduction

Sampling of a continuous time process, and subsequent inference based on those samples, is carried out in many disciplines of science and engineering. The object of inference is typically an attribute of the underlying continuous time process, such as the mean function or the power spectral density (also referred to as the spectrum). In some applications, samples of the process are collected at uncontrolled observational epochs (Bronez (1988), Ishimaru and Chen (1965), Munson, O'Brien, and Jenkins (1983), Nobach, Müller, and Tropea (1998)). When one can choose a sampling scheme for the observational time points, one typically uses uniform sampling (i.e., sampling at regular intervals) or additive random sampling (i.e., sampling at the times of renewal of a renewal process with a known distribution of spacing between successive samples) (Costain and Çoruh (2004), Eldar et al. (1997), Roughan (2006)). In such situations, there is often a practical constraint on the minimum separation between successive samples, that can arise due to technological constraints or from economic considerations. For example, in paleoclimatic studies based on ice core data (Petit et al. (1999)), the age of a particular sample is determined from radioactive dating techniques. Samples from ice core slices at greater depths are regarded as older. Ice core slices cannot be arbitrarily thin, and this constraint induces a limit on the time separation between successive samples. In computer graphics, where the time parameter is replaced by a two-dimensional space parameter, photo-receptor arrays are designed subject to the restriction of a minimum distance between two receptors, depending on the physical dimensions and specifications of the receptor (Cook (1986), Dippé and Wold (1985)). The constraint arises due to similar other limitations in laser doppler velocimetry (Ouahabi et al. (1998)) as well as radar applications (Nohrden (1995)).

Assuming that the data arise from renewal process sampling of a stationary stochastic process, subject to the constraint of a minimum inter-sample spacing, we develop a method for the estimation of its power spectral density.

Suppose that the minimum separation between successive samples is d units of time. If one chooses to sample at regular intervals, the fastest possible sampling rate is $1/d$. Thus, only those spectra which are limited to the frequency band $[-\pi/d, \pi/d]$ can be consistently estimated from uniformly spaced samples (Kay (1999)). A spectrum having larger support (bandwidth) than $[-\pi/d, \pi/d]$ would not be distinguishable from a corresponding spectrum having a support contained in that interval, and consequently the bias cannot go to zero even for large sample size. This limitation is related to the problem of aliasing, as described by the Nyquist Theorem, in respect of reconstruction of a function from its uniformly spaced samples. For spectra having larger support than $[-\pi/d, \pi/d]$, one has to look for methods based on alternative sampling schemes for consistent estimation of spectra in the presence of the above constraint.

In the absence of any constraint on the minimum inter-sample spacing, Shapiro and Silverman (1960) showed that certain renewal process sampling schemes, including Poisson process sampling, are alias-free for the class of non-bandlimited processes. Subsequently, some methods of spectrum estimation based on the covariance sequence or other aspects of stochastic samples of the process were proposed (Bardet and Bertrand (2010), Buetler (1970), Dunsmuir (1983), Huang, Hsing, and Cressie (2011), Lii and Masry (1994), Matsuda and Yajima (2009), Moore, Visser, and Shirtcliffe (2008), Stein, Chi, and Welty (2004)). These methods typically have larger variance than methods based on regularly spaced samples (Cook (1986), Dippé and Wold (1985), Moore, Visser, and Shirtcliffe (2008), Roughan (2006)).

In the presence of a constraint on the minimum separation between successive samples many alias-free sampling schemes, including Poisson sampling, are rendered infeasible. For such problems, researchers have turned to minimum distance Poisson sampling (which discards a sampling time if it is too close to the previous one) (Cook (1986), Dippé and Wold (1985)) or periodic non-uniform sampling (Ouahabi et al. (1998), Qu and Tarczynski (2007)). The latter form of sampling, ensuring a minimum separation between successive samples, has also been applied to the problem of curve estimation (Marvasti (2001), Tarczynski and Allay (2004)).

In a recent work, Srivastava and Sengupta (2011) studied the effect of the inter-sample spacing constraint on spectrum estimation based on stochastically sampled data. They showed that under the said constraint, no point process sampling scheme is alias-free for the class of non-bandlimited processes. This result implies that two processes with different spectra cannot be distinguished from point process samples, and hence consistent spectrum estimation from these samples is not possible. They also showed that there are sampling schemes, that can identify spectra limited to any finite bandwidth.

The opportunity created by this theoretical result has so far not been exploited. The known approaches for construction of spectrum estimators through additive random sampling, such as that based on the estimated covariance sequence of the sampled data (Shapiro and Silverman (1960)) and that based on the sampling times together with the sampled values (Lii and Masry (1994)), do not work in the presence of the minimum inter-sample spacing constraint (Srivastava and Sengupta (2011)). The method developed here is consistent, under certain regularity conditions on the process, regardless of the constraint of a minimum separation between successive samples. Thus, even if there is a restriction on how closely one can sample a continuous time process, a large number of appropriately chosen samples permits accurate and precise estimation of the underlying spectrum, as long as it is confined to an arbitrarily large but known band.

In Section 2, we develop the spectrum estimator. We establish its consistency and rate of convergence in Section 3. In Section 4, we study the performance of this estimator through Monte Carlo simulations for different degrees of minimum inter-sample spacing. In Section 5, we provide some concluding remarks, including those on how the assumption of known bandwidth might possibly be relaxed. The proofs of the theoretical results are given in the Appendix.

2. Spectrum Estimation from Samples With Minimum Inter-Sample Spacing

Let $X = \{X(t),\ -\infty < t < \infty\}$ be a real, mean square continuous and wide sense stationary stochastic process with mean zero, covariance function $C(\cdot)$, power spectral density $\phi(\cdot)$, and spectral support $[-\lambda_0, \lambda_0]$, where λ_0 is known. Let $\tau = \{t_n,\ n = \ldots, -2, -1, 0, 1, 2, \ldots\}$ be sequence of real-valued sampling times which constitute a stationary renewal process. Assume that the renewal process τ is independent of X. The inter-sample spacing density is denoted by $f(\cdot)$. The minimum separation between successive samples is denoted by d, so

$f(u) = 0$ for $0 \le u < d$. We assume that

$$f(u) > 0 \quad \text{for} \quad u \ge d. \tag{2.1}$$

The results presented in this section are valid without (2.1), it is made only simplify proofs.

We first estimate the covariance function, and subsequently use it to estimate the power spectral density. We estimate $C(0)$ by

$$\widehat{C}(0) = \frac{1}{n}\sum_{i=1}^{n} X^2(t_i). \tag{2.2}$$

For $u > d$, we use the estimator

$$\widehat{C}(u) = \frac{1}{nH(u)}\sum_{i=1}^{n}\sum_{j=1}^{n} m_n W(m_n(u - t_i + t_j))X(t_i)X(t_j) \quad \text{for } u > d. \tag{2.3}$$

Here $W(\cdot)$ is a weight function, m_n is the smoothing parameter, and $H(\cdot)$ is the renewal density of τ,

$$H(u) = \sum_{n=1}^{\infty} f^{(n)}(u), \tag{2.4}$$

where $f^{(n)}(\cdot)$ is the n-fold convolution of the inter-sample spacing density $f(\cdot)$. Note that, in view of (2.1), $H(u) > 0$ for $u > d$. The estimator given by (2.3) is essentially a weighted average of products of pairs of samples separated by lag approximately equal to u, and m_n is a smoothing parameter.

The main difficulty arising from the constraint of minimum inter-sample spacing lies in the lack of pairs of samples separated by lags smaller than d. Nevertheless, it has been shown that $C(0)$ and $C(u)$ for $u > d$ contain complete information about $C(\cdot)$ over its entire domain (Srivastava and Sengupta (2011)). A possible way of reconstructing the function over the range $(0, d]$ is to use the representation of $C(\cdot)$ in terms of its values over a grid. Note that the covariance function $C(\cdot)$ has the representation (Oppenheim and Schafer (2009))

$$C(u) = \sum_{l=-\infty}^{\infty} C(lT)\mathrm{sinc}\left(\frac{\pi}{T}(u - lT)\right), \tag{2.5}$$

where $T = \pi/\lambda_0$ and

$$\mathrm{sinc}(x) = \begin{cases} \frac{\sin x}{x} & \text{if } x \ne 0, \\ 1 & \text{if } x = 0. \end{cases}$$

In view of (2.5) and the symmetry of $C(\cdot)$, one need only to specify the sequence $\{C(lT),\, l = 0, 1, 2, \ldots\}$ to specify the function $C(\cdot)$ completely.

We have estimated values of $C(lT)$ for $l = 0$ and $l > J$, where J is the integer part of d/T, from (2.2) and (2.3). The remaining values, i.e., $C(T), \ldots, C(JT)$, can be expressed in terms of the left side of (2.5) and the known terms of the right side. Note that the left side of (2.5) can also be estimated directly for any $u > d$. Thus, the missing values satisfy the linear equations

$$\sum_{l=1}^{J} x_l(u) C(lT) = y(u) \quad \text{for } u > d, \tag{2.6}$$

where

$$x_l(u) = \left\{\operatorname{sinc}\left(\frac{\pi}{T}(u - lT)\right) + \operatorname{sinc}\left(\frac{\pi}{T}(u + lT)\right)\right\}, \text{ for } l = 1, 2, \ldots, \tag{2.7}$$

$$y(u) = C(u) - \operatorname{sinc}(\frac{\pi u}{T}) C(0) - \sum_{l=J+1}^{\infty} x_l(u) C(lT), \tag{2.8}$$

for $u > d$. One can use these equations to reconstruct $C(T), \ldots, C(JT)$.

For indirect estimation of $C(T), \ldots, C(JT)$ from (2.6), we define for $u > d$,

$$y_n(u) = \widehat{C}(u) - \operatorname{sinc}(\frac{\pi u}{T}) \widehat{C}(0) - \sum_{l=J+1}^{L_n} x_l(u) \widehat{C}(lT), \tag{2.9}$$

where L_n is a finite integer. Note that $y_n(u)$ is an estimator of $y(u)$ defined in (2.8), with the infinite sum truncated at L_n. Substitution of this estimator on the right side of (2.6) gives a set of approximate equations in $C(T), \ldots, C(JT)$. This 'functional data' linear model (Ramsay, Hooker, and Graves (2009)) leads to the least squares estimator

$$\begin{pmatrix} \widehat{C}(T) \\ \widehat{C}(2T) \\ \vdots \\ \widehat{C}(JT) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1J} \\ a_{21} & a_{22} & \cdots & a_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{J1} & a_{J2} & \cdots & a_{JJ} \end{pmatrix}^{-1} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_J \end{pmatrix}, \tag{2.10}$$

where $a_{jl} = \int_{u_1}^{u_2} x_j(u) x_l(u) du$, $c_j = \int_{u_1}^{u_2} x_j(u) y_n(u) du$, for $l, j = 1, \ldots, J$ and (u_1, u_2) is a suitable sub-interval of $[d, \infty)$. The description of the covariance estimator is completed by defining it in the range $0 < u \le d$, $u \ne T, \ldots, JT$, as

$$\widehat{C}(u) = \widehat{C}(0) \operatorname{sinc}\left(\frac{\pi u}{T}\right) + \sum_{l=1}^{L_n} \widehat{C}(lT) \left\{\operatorname{sinc}\left(\frac{\pi}{T}(u - lT)\right) + \operatorname{sinc}\left(\frac{\pi}{T}(u + lT)\right)\right\}.$$

Once the function $C(\cdot)$ is estimated, we estimate the power spectral density by a commonly used lag window estimator

$$\widehat{\phi}_n(\lambda) = \left\{\frac{T}{2\pi} \widehat{C}(0) + \frac{T}{\pi} \sum_{l=1}^{[n\mu_f/T]} \widehat{C}(lT) K(b_n l) \cos(l\lambda T)\right\} \times 1_{[-\lambda_0, \lambda_0]}(\lambda), \tag{2.11}$$

where $K(\cdot)$ is a covariance averaging kernel, b_n is the kernel bandwidth, μ_f is the mean of the inter-sample spacing distribution, and $[n\mu_f/T]$ is the greatest integer smaller that $n\mu_f/T$. The upper limit here ensures that the largest covariance lag included in the sum is about the same as the expected span of the data.

The proposed method involves specification of parameters m_n, b_n, L_n, u_1 and u_2, and functions W and K. Our convergence results hold for a range of choices of them. However, performance of the method may depend on the choices made. Some guidelines for making these selections are as follows.

For the functional data linear model based on (2.6) and (2.9), u_1 and u_2 determine the 'sample size' for estimating its parameters. While determining the 'sample size', one must include all the values of u for which $y_n(u)$ can provide information about the model parameters $C(T), \ldots, C(JT)$. Since the envelope of the sinc function is the reciprocal function, it is easy to see from the left side of (2.6) that the magnitudes of coefficients of all the $C(lT)$ terms are bounded from above by $2T/[\pi(u - JT)]$. When this number is small, the corresponding $C(u)$ are unlikely to be affected much by the parameters of interest, and so $y_n(u)$ cannot provide much information about them. Therefore, u_2 may be chosen so that $\pi(u_2 - JT)/2T$ is a suitably chosen large number. On the other hand, u_1 should be greater than or equal to d. In the simulations reported in Section 4, we have used $u_2 = T[J + 30/\pi]$ (i.e., $\pi(u_2 - JT)/2T = 15$) and $u_1 = d$, and these choices appear to have worked well.

Selection of the kernel/window function K for the lag window spectrum estimator has been extensively dealt with in the literature. Tapered windows are generally found to produce less bias than the rectangular window. Harris (1978) has provided a comparison of many windows. The selection of the bandwidth b_n has been dealt with by Hurvich (1985) and Beltrão and Bloomfield (1985).

The issue of choosing the window function W and the corresponding window width m_n concerns the balance between bias and variance of the estimator. Ensuring non-negative definiteness of the estimator is an additional consideration. Guillot, Senoussi, and Monestiez (2001) have addressed the issue.

The parameter L_n determines how many terms of an infinite series expansion are actually used in indirect estimation of some covariance terms. It is seen from the right side of (2.9) that the contribution of $\widehat{C}(lT)$ is small when $x_l(u)$ is small. Since $|x_l(u)| \leq 2T/[\pi(lT - u)]$, $l > L_n$ implies $|x_l(u)| \leq 2T/[\pi(L_nT - u_2)]$. Thus one can choose L_n by setting $2T/[\pi(L_nT - u_2)]$ equal to a suitably small threshold. According to the large sample results obtained in Section 3, this threshold can decrease with n over a range of rates.

3. Consistency of the Spectrum Estimator

In order to establish consistency of the proposed spectrum estimator, we first show consistency of the corresponding estimator of the covariance function. For this, we choose the weight function $W(\cdot)$ of (2.3) so that it has the following properties.

W1. The function $W(\cdot)$ is compactly supported, even, continuous, and square integrable, with $\int_{-\infty}^{\infty} W(v)dv = 1$.

W2. For a specified r,

$$\int_{-\infty}^{\infty} v^k W(v)dv = \begin{cases} 0 & \text{for } k = 1, \ldots, r-1, \\ w_r \neq 0 & \text{for } k = r. \end{cases}$$

The number r is termed the order of the weight function.

We make some assumptions about the underlying process and the inter-sample spacing density.

Assumption 1. The covariance function $C(\cdot)$ is bounded by a decreasing and integrable function over $[0, \infty)$.

Assumption 1A. The function $|u|^q C(u)$, for some $q > 1$, is bounded by a decreasing and integrable function over $[0, \infty)$.

The parameter q of Assumption 1A signifies the degree of smoothness of $\phi(\cdot)$; in particular, if q is an integer, $\phi(\cdot)$ is q times differentiable. Assumption 1A is stronger than Assumption 1, and is used only to obtain rates of convergence.

Assumption 2. The inter-sample spacings density $f(\cdot)$ has a finite mean.

Theorem 1. *Under Assumptions* 1 *and* 2*, take* $W(\cdot)$ *in* (2.3) *to have the properties* W1 *and* W2*, and let the smoothing parameter* $m_n \to \infty$ *and the truncation parameter* $L_n \to \infty$ *as* $n \to \infty$.

(i) *The bias of the estimator* $\widehat{C}(\cdot)$ *converges to* 0 *pointwise as the sample size* n *goes to infinity.*

(ii) *Under Assumptions* 1A*, if* $m_n \log L_n / n \to 0$ *and* $\log L_n / m_n^r \to 0$ *as* $n \to \infty$*, we have*

$$E[\widehat{C}(u)] - C(u) = \begin{cases} 0 & \text{if } u = 0, \\ O\left(\frac{1}{m_n^r}\right) + O\left(\frac{m_n}{n}\right) & \text{if } |u| > d, \\ O\left(\frac{1}{L_n^q}\right) + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right) & \text{if } 0 < |u| \leq d, \end{cases}$$

where r *is the order of weight function and* $O(\cdot)$ *is uniform in* u.

Note that $L_n \to \infty$ together with $m_n \log L_n/n \to 0$ implies that $m_n/n \to 0$ as $n \to \infty$.

A further assumption on fourth order moments of the process is needed to establish convergence of the variance of the estimator $\widehat{C}(u)$.

Assumption 3. The fourth moment $E|X(t)|^4$ exists for every t, the fourth order cumulant Q of $X(t+t_1)$, $X(t+t_2)$, $X(t+t_3)$ and $X(t)$ does not depend on t, and

$$|Q(t_1,t_2,t_3)| \leq \prod_{j=1}^{3} g_j(t_j),$$

where $g_j(\cdot),\ j=1,2,3$, are continuous, even, nonnegative, and integrable functions that are non-increasing over $[0,\infty)$.

Assumption 3 holds trivially for a Gaussian process, since $Q(\cdot)=0$.

Theorem 2. *Under Assumptions* 1–3*,* $m_n \to \infty$*,* $L_n \to \infty$ *and* $m_n(\log L_n)^2/n \to 0$ *as* $n \to \infty$*, if* $W(\cdot)$ *satisfies W1 and W2,*

$$\mathrm{Var}\,(\widehat{C}(u)) = \begin{cases} O\left(\frac{1}{n}\right) & \textit{if } u=0,\\ O\left(\frac{m_n}{n}\right) & \textit{if } |u|>d,\\ O\left(\frac{m_n(\log L_n)^2}{n}\right) & \textit{if } 0<|u|\leq d. \end{cases}$$

Note that $L_n \to \infty$ and $m_n(\log L_n)^2/n \to 0$ implies that $m_n/n \to 0$ as $n \to \infty$.

Under the assumptions of Theorem 1(ii) and 2, we have, for $0<|u|\leq d$,

$$\begin{aligned} &E[\widehat{C}(u)-C(u)]^2\\ &\quad=(E[\widehat{C}(u)]-C(u))^2+\mathrm{Var}\,(\widehat{C}(u))\\ &\quad=O\left(\frac{1}{L_n^{2q}}\right)+O\left(\frac{(\log L_n)^2}{m_n^{2r}}\right)+O\left(\frac{m_n^2(\log L_n)^2}{n^2}\right)+O\left(\frac{m_n(\log L_n)^2}{n}\right). \end{aligned} \tag{3.1}$$

In view of the conditions of Theorem 2, the third term in (3.1) can be ignored. Note that, for fixed L_n, the second term is a decreasing function of m_n, while the fourth term is increasing in m_n. The fastest possible rate of convergence is achieved by equating these two rates, which yields $m_n = O\big(n^{\frac{1}{2r+1}}\big)$. Likewise, the first term is a decreasing function of L_n, while the fourth term is increasing in L_n. Again by equating the rates, we get the optimal rate of L_n satisfying the condition

$$(L_n^q \log L_n)^{(2r+1)/r} = O(n)$$

for fastest convergence of the mean squared error. Even though a closed form expression for the optimal rate of L_n is not available, use of the square root function in lieu of the log function leads to the nearly optimal rate $L_n = O(n^{2r/[(2r+1)(2q+1)]})$. Substitution of these rates of m_n and L_n in (3.1) shows that the optimal rate of convergence of the MSE is

$$E[\widehat{C}(u) - C(u)]^2 = o\Big(n^{-1/[(1+1/2r)(1+1/2q)]}\Big). \tag{3.2}$$

On the other hand, for $|u| > d$, the MSE of the direct estimate $\widehat{C}(u)$ is

$$E[\widehat{C}(u) - C(u)]^2 = O\left(\frac{1}{m_n^{2r}}\right) + O\left(\frac{m_n}{n}\right).$$

By using a similar argument as above, one can be see that the optimal rate of convergence of the MSE for $|u| > d$ is

$$E[\widehat{C}(u) - C(u)]^2 = O\Big(n^{-1/(1+1/2r)}\Big). \tag{3.3}$$

We now turn to the consistency of the spectrum estimator $\widehat{\phi}_n(\lambda)$ given in (2.11). For this purpose, we choose the covariance averaging kernel $K(\cdot)$ used in (2.11) so that it has the following properties.

K1. The function $K(\cdot)$ is an even, continuous and square integrable function with $K(0) = 1$, and is bounded by a nondecreasing function over $(0, \infty)$.

K2. The order of the kernel $K(\cdot)$ is q, where q is as in Assumption 1A.

Theorem 3. *Suppose Assumptions* 1–2 *hold, the kernel* $K(\cdot)$ *of* (2.11) *satisfies K1 and K2, and the weight function* $W(\cdot)$ *used in the covariance estimates has properties* W1 *and* W2. *Let* $m_n \to \infty$, $L_n \to \infty$, *and* $b_n \to 0$ *as* $n \to \infty$.

(i) *The bias of the estimator* $\widehat{\phi}_n(\cdot)$ *converges to* 0 *pointwise as the sample size* n *goes to infinity.*

(ii) *Under the conditions of Theorem* 1(ii) *and* $1/(b_n m_n^r) \to 0$, *we have*

$$\begin{aligned} &E[\widehat{\phi}_n(\lambda)] - \phi(\lambda) \\ &\quad = O(b_n^q) + O\left(\frac{1}{L_n^q}\right) + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right) + O\left(\frac{1}{b_n m_n^r}\right), \end{aligned}$$

where r *is the order of the weight function* $W(\cdot)$.

Theorem 4. *Under the conditions of Theorem* 2, $b_n \to 0$ *and* $m_n/nb_n^2 \to 0$ *as* $n \to \infty$, *and* $K(\cdot)$ *satisfying* K1 *and* K2,

$$\mathrm{Var}\,[\widehat{\phi}_n(\lambda)] = O\left(\frac{m_n(\log L_n)^2}{n}\right) + O\left(\frac{m_n}{nb_n^2}\right).$$

Under the conditions of Theorems 3 and 4, we have the rate of convergence of the MSE of the estimator $\widehat{\phi}_n(\cdot)$,

$$E[\widehat{\phi}_n(\lambda)-\phi(\lambda)]^2 = O\left(b_n^{2q}\right)+O\left(\frac{1}{L_n^{2q}}\right)+O\left(\frac{(\log L_n)^2}{m_n^{2r}}\right)+O\left(\frac{m_n^2(\log L_n)^2}{n^2}\right) +O\left(\frac{1}{b_n^2 m_n^{2r}}\right)+O\left(\frac{m_n(\log L_n)^2}{n}\right)+O\left(\frac{m_n}{nb_n^2}\right). \tag{3.4}$$

In view of the conditions of Theorem 2, the fourth term in (3.4) is smaller than the sixth term, and hence can be ignored. Note that for fixed b_n, the fifth term is a decreasing function of m_n, while the seventh term is increasing in m_n. The fastest rate of convergence is obtained by equating the two rates, which yields $m_n = O\big(n^{1/(1+2r)}\big)$. Now, for this optimal choice of m_n, the fifth term is a decreasing function of b_n, while the first term is increasing in b_n. Thus, by equating the two rates, we obtain the optimal rate of b_n as $O\big(n^{-\frac{r}{(1+2r)(1+q)}}\big)$. Substitution of these choices reduces (3.4) to

$$E[\widehat{\phi}_n(\lambda)-\phi(\lambda)] = O\Big(n^{-1/[(1+1/2r)(1+1/q)]}\Big)+O\left(\frac{1}{L_n^{2q}}\right)+O\left(\frac{(\log L_n)^2}{n^{2r/(1+2r)}}\right). \tag{3.5}$$

The requirement that the second and the third terms of the right side go to zero at least as fast as the first term, leads us to a range of optimal choices of L_n as

$$\log L_n = O\left(n^{1/[(1+1/2r)(2+2q)]}\right); \quad L_n^{-1} = O\left(n^{-1/[(1+1/2r)(2+2q)]}\right).$$

For any choice of L_n in this optimal range, and for $m_n = O\big(n^{1/(1+2r)}\big)$ and $b_n = O\big(n^{-r/[(1+2r)(1+q)]}\big)$, the MSE of $\widehat{\phi}_n(\lambda)$ achieves the fastest rate of convergence, $O\big(n^{-1/[(1+1/2r)(1+1/q)]}\big)$.

We conclude this section by looking into the special case $d \le \pi/\lambda_0$, where no indirect estimation of the covariance sequence is needed. In this situation, the parameter L_n is not required and the optimal rate of convergence of the MSE of $\widehat{\phi}_n(\cdot)$ is once again $O(n^{-1/[(1+1/2r)(1+1/q)]})$. Since the order r of the weight function $W(\cdot)$ can be chosen to be arbitrarily large, the optimal rate of convergence can approach $O(n^{-1/[1+1/q]})$. Note that when $d \le \pi/\lambda_0$, uniform sampling under the given constraint becomes alias-free, and the optimal rate of convergence of the MSE of the smoothed periodogram estimator based on uniform sampling is $O(n^{-1/[1+1/2q]})$ (Brillinger (1972)). The optimal rate of convergence of the MSE of $\widehat{\phi}_n(\cdot)$ is only marginally slower than this rate when q is large, indicating that the underlying power spectral density is very smooth.

4. Simulation

In this section, we study the finite sample performance of the estimator $\widehat{\phi}_n(\cdot)$ with different degrees of constraint on the inter-sample spacing. For comparison, we use an estimator proposed by Lii and Masry (1994) in the case of Poisson sampled data:

$$\hat{\psi}_n(\lambda) = \frac{1}{\pi\beta n}\sum_{j=1}^{n-1}\sum_{l=1}^{n-j} X(t_l)X(t_{j+l})K\left(b_n(t_j+l-t_l)\right)\cos(\lambda(t_j+l-t_l)), \quad (4.1)$$

Here, β is the average sampling rate, $K(\cdot)$ is a covariance averaging kernel, and b_n is the smoothing parameter. The estimator $\hat{\psi}_n(\cdot)$ is consistent in the absence of the constraint on the minimum inter-sample spacing, under conditions similar to those used for establishing consistency of $\widehat{\phi}_n(\cdot)$. However, in the presence of this constraint, it follows from an argument given by Srivastava and Sengupta (2011) that the estimator is not consistent.

We consider a continuous time stationary stochastic process X with mean 0 and covariance function

$$C(u) = \begin{cases} \sigma^2 \displaystyle\sum_{j=0}^{p-|l|} \xi_j \xi_{j+|l|} & \text{if } u = \frac{l\pi}{\lambda_0}, |l| = 0, 1, \ldots, p, \\ 0 & \text{if } u = \frac{l\pi}{\lambda_0}, |n| > p, p+1, \ldots, \\ \displaystyle\sum_{l=-\infty}^{\infty} C\left(\frac{l\pi}{\lambda_0}\right) \frac{\sin(\lambda_0 u - l\pi)}{(\lambda_0 u - l\pi)} & \text{otherwise.} \end{cases} \quad (4.2)$$

This covariance function corresponds to a process limited to the frequency band $[-\lambda_0, \lambda_0]$, and whose samples at regular intervals of length π/λ_0 constitute a discrete time MA(p) process with MA characteristic polynomial $\Xi(z) = \xi_0 + \xi_1 z + \xi_2 z^2 + \cdots + \xi_p z^p$ and innovation variance σ^2.

We use sampling with a stationary renewal process τ whose inter-sample spacing is distributed as $d + R$, where the random variable R has the exponential distribution with mean θ. Note that this 'minimum distance' Poisson sampling (Dippé and Wold (1985)) is obtained if one attempts to implement Poisson sampling by generating successive inter-sample spacings from the exponential distribution with mean θ, but is obliged to discard those inter-sample spacings which are smaller than d. We assume that n consecutive samples, denoted by $X(t_1), X(t_2), \ldots, X(t_n)$, are available for estimation. We chose the parameters as

$$\lambda_0 = 2\pi,$$
$$\Xi(z) = (1 + 1.2z)^8,$$

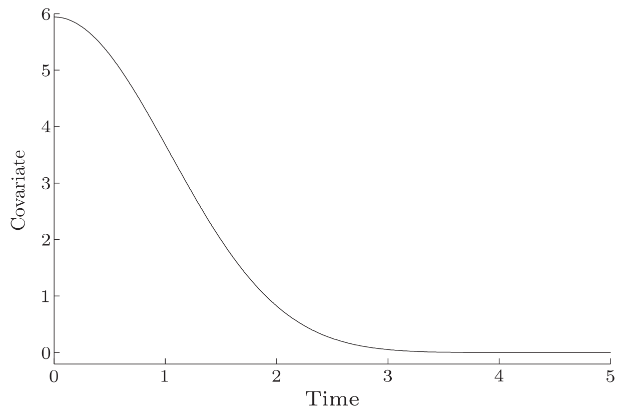

Figure 1. The covariance function corresponding to (4.2).

$$\sigma = \frac{1}{10^2},$$

$$W(x) = \begin{cases} \frac{1}{2}\{1+\cos(\pi x)\} & \text{if } -1 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{and } K(x) = \begin{cases} \frac{1}{2}\{1+\cos(\pi x)\} & \text{if } -1 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the common form chosen for the weight function W and the kernel function K is of the second order ($r = q = 2$). We chose two values of the minimum inter-sample spacing: $d = 0.75$ and $d = 1$. These thresholds are larger than the uniform inter-sample spacing needed for sampling at the Nyquist rate (which happens to be $T = 0.5$). It follows from the discussion of Section 2 that $J = 1$ for $d = 0.75$, and $J = 2$ for $d = 1$.

We chose $\theta = d/2$ and $\theta = 2d$ for each choice of d to illustrate the effect of θ. These two values correspond to the situations in which the inter-sample spacing is mostly dominated by the deterministic part (d) and the random part (R), respectively. The average sampling rate is $\beta = 1/(\theta + d)$, and this rate is used in computing $\widehat{\psi}_n(\cdot)$.

We ran simulations for sample sizes $n = 100$ and $n = 1,000$. For computing the estimator $\widehat{\phi}_n(\cdot)$, we chose $u_1 = d$ and $u_2 = T(J + 30/\pi)$. Going by the considerations mentioned in Section 2 and the rate calculations of Section 3, we chose L_n by setting $2T/[\pi(L_nT - u_2)] = \left(5n^{1/[(1+1/2r)(2+2q)]}\right)^{-1}$. For both $\widehat{\phi}_n(\cdot)$ and $\widehat{\psi}_n(\cdot)$ and both sample sizes, we used $b_n = 0.1$. Finally, we used $m_n = 5$ for $n = 100$ and $m_n = 8$ for $n = 1,000$.

The left column of Figure 2 shows the average of the estimates of $\widehat{\phi}_n(\lambda)$ and $\widehat{\psi}_n(\lambda)$ computed from 500 Monte Carlo runs, along with the true density.

The plots show that when the sample size increased, the average value of $\widehat{\psi}_n(\cdot)$ approached a wrong function, not even positive over its entire range. On the other hand, the average of the estimator $\widehat{\phi}_n(\lambda)$ approached the true power spectral density for larger n, for both choices of the parameter θ. The middle column of Figure 2 shows the variance of both the estimators. Though the variance of $\widehat{\psi}_n$ is smaller than that of $\widehat{\phi}_N$, the right column of Figure 2 indicates that the mean squared error of $\widehat{\phi}_n(\lambda)$ is smaller. Further, the mean squared error of $\widehat{\phi}_n(\lambda)$ decreased with sample size, while that of $\widehat{\psi}_n(\lambda)$ saturated to a non-zero level because of the bias component. In all cases, the smaller value θ corresponded to larger bias and smaller variance of $\widehat{\phi}_n(\lambda)$. This is expected, as a small value of the θ generally brings the inter-sample spacing closer to d, and sampling at uniform intervals of size d would lead to aliasing in this case. For a large value of θ, the sampling scheme moves away from uniform sampling, and the proposed estimator has a better chance to rectify the bias at the cost of increased variance.

The more challenging case has $d = 1$. The simulations were run with the parameters chosen as described in the previous case. Figures 3 depicts the empirical average, the variance, and the mean squared error for the two estimators, computed from 500 simulation runs. The plots show a similar pattern as in the case of $d = 0.75$, though the performance of both the estimators was worse, as expected.

The MSE of both estimators appears to be large for $\theta = 2d$ in comparison with $\theta = d/2$ in all cases. The value θ controls the amount of randomness in the sampling scheme: very small value of θ leads to the sampling scheme close to uniform sampling, and estimation suffers from the aliasing in the small sample situation; the larger MSE for $\theta = 2d$ indicates that too much randomness in the sampling scheme leads to higher MSE in finite samples. The MSE plot shows that for larger sample size the MSE of $\widehat{\phi}_n$ decreases for both choices of θ. It shows that the estimator performs well for any choice of θ provided one has enough sample size.

The MSE of $\widehat{\phi}_n(\lambda)$ is seen to be larger in the case $d = 1$ than in the case $d = 0.75$. This finding can be explained by the fact that, in the former case, the estimator $\widehat{\phi}_n(\cdot)$ involves indirect estimation of *two* covariance parameters, $C(T)$ and $C(2T)$, as opposed to indirect estimation of $C(T)$ only when $d = 0.75$. Reduction of the set of lags suitable for direct estimation is another reason why the estimator has poorer performance for larger values of d. In any case, such difficulties are made up by large sample size, as is evident from the MSE of $\widehat{\phi}_n(\lambda)$ for $n = 1,000$.

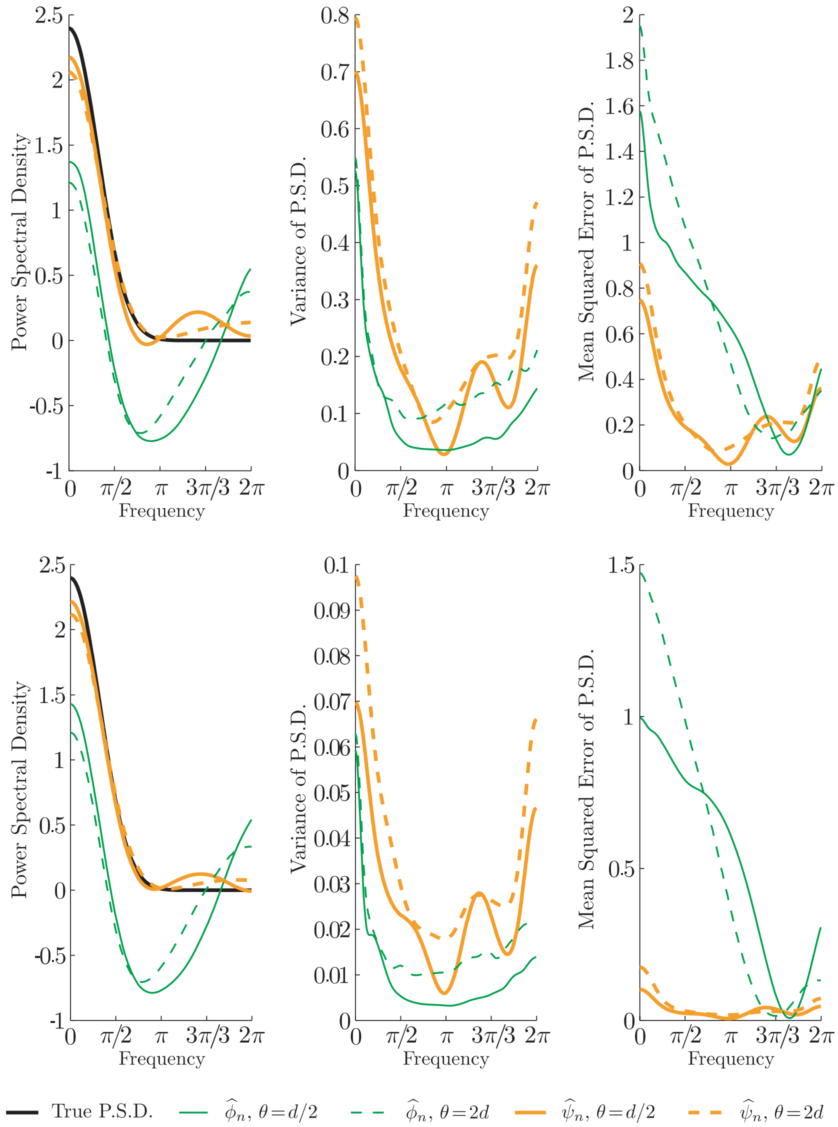

Figure 2. Averages (left column), variances (middle column) and mean squared errors (right column) of the estimates of $\widehat{\phi}_n(\lambda)$ and $\widehat{\psi}_n(\lambda)$ for sample sizes $n = 100$ (top row) and 1,000 (bottom row) computed from 500 Monte Carlo simulation runs for minimum separation $d = 0.75$.

5. Concluding Remarks

This paper provides a method of consistent estimation of an arbitrarily bandlimited power spectral density of a continuous time stationary stochastic process under the constraint that there is at least a specified amount of sep-

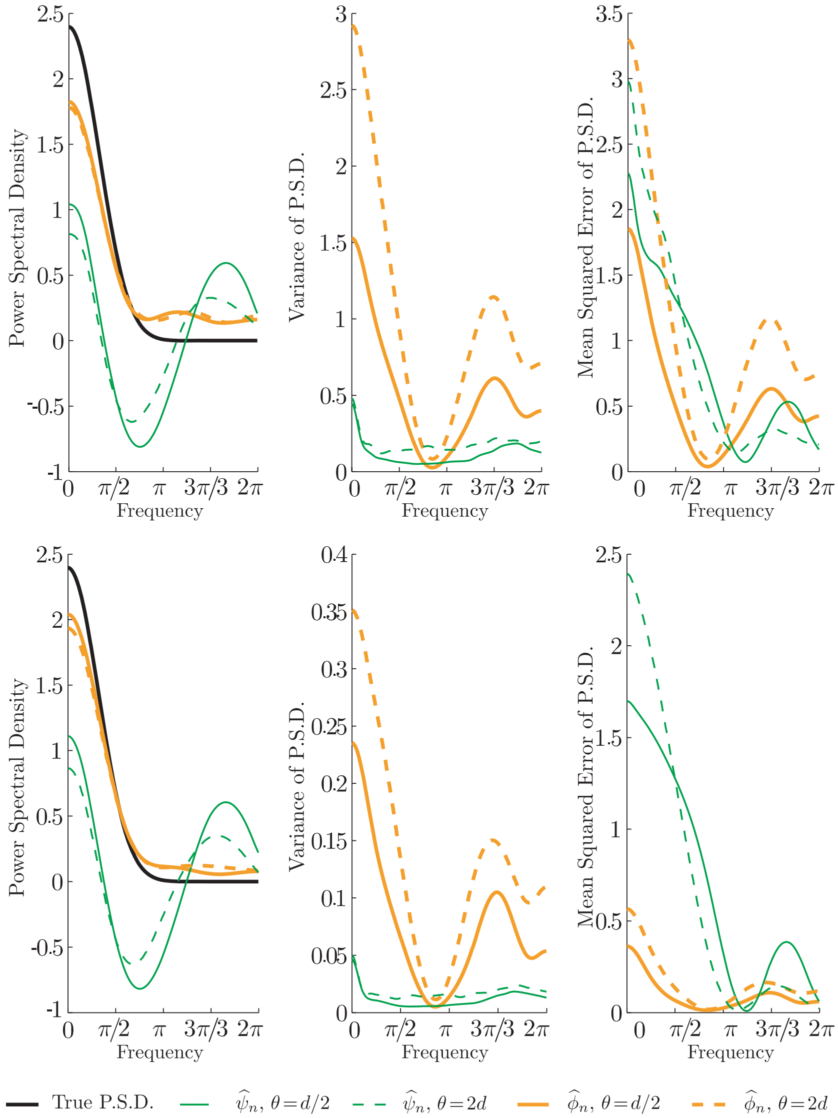

Figure 3. Averages (left column), variances (middle column) and mean squared errors (right column) of the estimates of $\widehat{\phi}_n(\lambda)$ and $\widehat{\psi}_n(\lambda)$ for sample sizes $n = 100$ (top row) and 1,000 (bottom row) computed from 500 Monte Carlo simulation runs for minimum separation $d = 1$.

aration between successive samples. The proposed nonparametric estimator is based on additive random sampling of the underlying process subject to this constraint. The estimator is the first of its kind, as its known competitors based on stochastic sampling are inconsistent in the presence of the constraint, and the known competitors based on uniform sampling are consistent only when the bandwidth of the underlying process is within the limit implied by the Nyquist theorem. Consistency of the proposed estimator demonstrates that large samples collected through appropriate additive random sampling can be used to surpass the Nyquist limit for the bandwidth of a spectrum that can be consistently estimated through uniform sampling.

The constraint of minimum separation between successive samples makes it impossible to estimate autocovariances at small lags directly from the data. The proposed method circumvents this difficulty by expressing these autocovariances in terms of directly estimable autocovariances through the representation (2.5). This indirect method of estimation leads to larger variance than in the case of direct estimation. The greater the minimum separation, the larger the need for indirect estimation and the larger is the variance of the resulting spectrum estimator. The simulation results reported in Section 4 confirm this fact. Thus, while it is possible to make up for the deficiency of sampling resolution through sample size, the sample size needs to be large when the resolution is not good.

A crucial assumption for the estimation problem considered in this article is that the underlying power spectral density has a known and finite bandwidth. This is indeed an undesirable assumption. For a proper perspective on this issue, consider the problem of consistent spectrum estimation based on data sampled with a minimum inter-sample spacing d, under a set of progressively weaker assumptions: (a) the underlying spectrum has bandwidth less than π/d, (b) the underlying spectrum has a finite and known bandwidth (possibly larger than π/d), (c) the underlying spectrum has a finite bandwidth that may be unknown, (d) the underlying spectrum has possibly infinite bandwidth. Srivastava and Sengupta (2011) have shown that the problem, under (d), has no solution based on point process sampling. Estimators based on uniform sampling can be consistent only under (a). The method based on additive random sampling, proposed in this article, is consistent under (b). As of now, there is no solution under (c). However, if one does not have a constraint on sample size or computing power, one may develop a procedure on the basis of the proposed estimator. For instance, one can assume successively higher values of the maximum bandwidth, then determine the value after which the corresponding estimators exhibit no substantial change. Development of a suitable estimator along these lines would require further research.

Some researchers have promoted point process sampling as a means of consistent estimation of non-bandlimited power spectral densities, for which uniformly spaced samples are said to be inadequate no matter how large the sampling rate (Shapiro and Silverman (1960), Lii and Masry (1994)). This apparent deficiency of uniform sampling disappears if one uses shrinking asymptotics to establish consistency (Srivastava and Sengupta (2010)). However, this asymptotic approach becomes inappropriate when inter-sample spacings are constrained to be larger than a threshold value. On the other hand, under this constraint, consistent estimation of a non-bandlimited power spectral density through point process sampling is also not possible (Srivastava and Sengupta (2011)). Thus, as far as estimation of a *non-bandlimited* power spectral density is concerned, uniform sampling and point process sampling have similar strengths and limitations. The spectrum estimator introduced in this paper underscores an exclusive advantage of point process sampling in the area of estimation of a *bandlimited* power spectral density.

Control over the sampling mechanism has been assumed in the present work. In particular, knowledge of the inter-sample spacing distribution has been used to determine $H(u)$ in the estimator (2.3), and in choosing J for (2.6) and (2.10). When there is no control over the sampling mechanism, one may still wish to model it as additive random sampling and use the proposed method for spectrum estimation. In such a case, one has to test whether the inter-sample spacings are independent and then estimate the spacings distribution, along with its support, from the observed data of sampling times. These are standard inferential problems. While the use of consistent estimators for $H(u)$ and J in (2.3) and (2.10) may not affect consistency of the estimators of covariance and power spectral density, there may be an impact on the rates of convergence. This matter would require further study.

Acknowledgement

The research work of Radhendushka Srivastava was partially supported by the Indian Statistical Institute Kolkata and the Office of Naval Research (Grant YIP-N000140910911 of Prof. Ping Li). The authors thank two referees and an associate editor for comments that sharpened many technical points and improved the general readability of the paper.

Appendix

We denote the domain of the function $W(\cdot)$ by $[-a, a]$, the suprema of the functions $|W|(\cdot)$ and $H(\cdot)$ by M_1 and M_2, respectively, and the infimum of $H(\cdot)$ by M_3. The function $g(\cdot)$ is a bounding function as in Assumption 1.

Proof of Theorem 1. Part (i). Observe that

$$E[\widehat{C}(0)] = \frac{1}{n}\sum_{i=1}^{n} E[X^2(t_i)] = C(0). \tag{A.1}$$

From (2.3), we have for $|u| > d$

$$E[\widehat{C}(u)]=\frac{1}{nH(u)}\sum_{i=1}^{n}\sum_{j=1}^{n} m_n E[E\{W(m_n(u-t_i+t_j))X(t_i)X(t_j)|t_i,\ i=1,\ldots,n\}]$$

$$= \frac{1}{nH(u)} \sum_{i=1}^{n} \sum_{j=1}^{n} m_n E[W(m_n(u - t_i + t_j))C(t_i - t_j)]$$

By considering the case $i = j$ separately, and combining the other terms, we have

$$\begin{aligned} E[\widehat{C}(u)] &= \frac{m_n W(m_n u)C(0)}{H(u)} + \frac{1}{H(u)} \int_0^\infty m_n \bigg[\{W(m_n(u+v)) + W(m_n(u-v))\} \\ &\qquad \times C(v) \bigg\{ \frac{1}{n} \sum_{1 \le i < j \le n} f^{(j-i)}(v) \bigg\} \bigg] dv \\ &= \frac{m_n W(m_n u)C(0)}{H(u)} + \frac{1}{H(u)} \int_0^\infty m_n \bigg[\{W(m_n(u+v)) + W(m_n(u-v))\} \\ &\qquad \times C(v) H_n(v) \bigg] dv, \end{aligned} \tag{A.2}$$

where

$$H_n(u) = \sum_{l=1}^{n-1} \left(1 - \frac{l}{n}\right) f^{(l)}(u). \tag{A.3}$$

By making a transformation of the variable of integration and using the symmetry of the covariance function $C(\cdot)$, we have

$$\begin{aligned} E[\widehat{C}(u)] &= \frac{m_n W(m_n u)C(0)}{H(u)} \\ &\quad + \frac{1}{H(u)} \int_{-\infty}^{\infty} W(v) C\left(u - \frac{v}{m_n}\right) \left[H_n\left(u - \frac{v}{m_n}\right) + H_n\left(-u + \frac{v}{m_n}\right)\right] dv. \end{aligned} \tag{A.4}$$

For sufficiently large n, we have $m_n > a/d$, and consequently $W(m_n u) = 0$ for all $|u| > d$. This implies that the first term is identically zero for large n. Thus, we have for large n,

$$E[\widehat{C}(u)] = \frac{1}{H(u)} \int_{-\infty}^{\infty} W(v) C\left(u - \frac{v}{m_n}\right) \left[H_n\left(u - \frac{v}{m_n}\right) + H_n\left(-u + \frac{v}{m_n}\right)\right] dv. \tag{A.5}$$

Further, by using Assumptions 1 and 2, we have the dominance

$$\begin{aligned} &\left| W(v) C\left(u - \frac{v}{m_n}\right) \left[H_n\left(u - \frac{v}{m_n}\right) + H_n\left(-u + \frac{v}{m_n}\right)\right] \right| \\ &\quad \le |W(v)|\, g(0) \left[H_n\left(u - \frac{v}{m_n}\right) + H_n\left(-u + \frac{v}{m_n}\right)\right] \\ &\quad \le |W(v)|\, g(0) 2M_2, \end{aligned}$$

and Property W1 of the weight function $W(\cdot)$ ensures that the bounding function is integrable. Since $m_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n\to\infty} W(v)C\left(u-\frac{v}{m_n}\right)\left[H_n\left(u-\frac{v}{m_n}\right)+H_n\left(-u+\frac{v}{m_n}\right)\right]=W(v)C(u)H(u).$$

By applying the Dominated Convergence Theorem (DCT), we have

$$\lim_{n\to\infty} E[\widehat{C}(u)] = C(u). \tag{A.6}$$

In order to compute the expectation of the indirect estimators $\widehat{C}(T), \ldots, \widehat{C}(JT)$ given in (2.10), we first compute $E[\int_{u_1}^{u_2} x_j(u)y_n(u)du]$. Note that for $j = 1, 2, \ldots, J$,

$$E\left[\int_{u_1}^{u_2} x_j(u)y_n(u)du\right] = \int_{u_1}^{u_2} x_j(u)E[y_n(u)]du,$$

where interchange of the integrals is justified by the finiteness of the double integral that follows from arguments similar to those given below to establish convergence. We compute

$$\begin{aligned} E[y_n(u)] &= E[\widehat{C}(u)] - \mathrm{sinc}(\frac{\pi u}{T})E[\widehat{C}(0)] \\ &\quad - \sum_{l=J+1}^{L_n} E[\hat{C}(lT)]\left\{\mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big)+\mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}. \end{aligned} \tag{A.7}$$

It is seen from (A.1) and (A.6) that $E[\widehat{C}(u)] \to C(u)$ as $n \to \infty$ for $u > d$ and $E[\widehat{C}(0)] = C(0)$. By using (A.5) for large n, the third term of the right side of (A.7) simplifies to

$$\begin{aligned} \int_{-\infty}^{\infty} W(v)\Bigg[\sum_{l=J+1}^{L_n} \frac{1}{H(lT)}C\left(lT-\frac{v}{m_n}\right)\left\{H_n\Big(lT-\frac{v}{m_n}\Big)+H_n\Big(-lT+\frac{v}{m_n}\Big)\right\} \\ \times\left\{\mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big)+\mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}\Bigg]dv. \end{aligned} \tag{A.8}$$

By using Assumptions 1 and 2, we have the dominance

$$\begin{aligned} &\Bigg|W(v)\sum_{l=J+1}^{L_n} \frac{1}{H(lT)}C\left(lT-\frac{v}{m_n}\right)\left\{H_n\left(lT-\frac{v}{m_n}\right)+H_n\left(-lT+\frac{v}{m_n}\right)\right\} \\ &\qquad\times\left\{\mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big)+\mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}\Bigg| \\ &\le |W(v)|\frac{2M_2}{M_3}\sum_{l=J+1}^{L_n} g\Big(lT-\frac{v}{m_n}\Big) \le |W(v)|\frac{2M_2}{M_3}\times 2\sum_{l=1}^{\infty} g(lT), \end{aligned}$$

and the integrability of the bound is guaranteed by Property W1 of $W(\cdot)$. Since $m_n \to \infty$ and $L_n \to \infty$ as $n \to \infty$, the integrand of (A.8) converges pointwise,

$$\lim_{n\to\infty} W(v) \sum_{l=J+1}^{L_n} E[\widehat{C}(lT)] \Big\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\Big\}$$
$$= W(v) \sum_{l=J+1}^{\infty} C\,(lT) \Big\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\Big\}.$$

By using the representation (2.5) and the DCT, we have

$$\lim_{n\to\infty} E[y_n(u)] = C(u) - \mathrm{sinc}(\frac{\pi u}{T})C(0) - \sum_{l=J+1}^{\infty} C(lT) \Big\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big)$$
$$+ \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\Big\} = y(u).$$

Thus, for $j = 1, \ldots, J$, using (2.6) we have

$$\lim_{n\to\infty} E\left[\int_{u_1}^{u_2} x_j(u) y_n(u) du\right] = \int_{u_1}^{u_2} x_j(u) \left\{\sum_{l=1}^{J} x_l(u) C(lT)\right\} du. \quad \text{(A.9)}$$

Then (A.9) and (2.10), for $l = 1, \ldots, J$, imply

$$\lim_{n\to\infty} E[\widehat{C}(lT)] = C(lT). \quad \text{(A.10)}$$

Uitilizing a similar argument as above, one can establish that for $u \in (0, d]$ and $u \neq T, \ldots, JT$, $\lim_{n\to\infty} E[\widehat{C}(u)] = C(u)$, completing the proof of Part (i).

Part (ii). With (A.2), (A.3), and Assumption 2, we have, for sufficiently large n such that $m_n > a/d$,

$$E[\widehat{C}(u)] = \int_0^{\infty} m_n \big[\{W(m_n(u+v)) + W(m_n(u-v))\}C(v)\big] dv$$
$$- \frac{1}{H(u)} \int_0^{\infty} m_n \left[\{W(m_n(u+v)) + W(m_n(u-v))\}C(v) \sum_{l=n}^{\infty} f^{(l)}(v)\right] dv$$
$$- \frac{1}{nH(u)} \int_0^{\infty} m_n \left[\{W(m_n(u+v)) + W(m_n(u-v))\}C(v) \sum_{l=1}^{n-1} l f^{(l)}(v)\right] dv.$$
(A.11)

Observe that

$$\sum_{l=1}^{\infty} l f^{(l)}(u) = \sum_{l=1}^{\infty} l \frac{d}{du} F^{(l)}(u) = \frac{d}{du} \sum_{l=1}^{\infty} l P(N(u) \geq l)$$

$$
\begin{aligned}
&= \frac{d}{du}\sum_{l=1}^{\infty} l \sum_{j=l}^{\infty} P(N(u)=j) = \frac{d}{du}\sum_{j=1}^{\infty} P(N(u)=j) \sum_{l=1}^{j} l \\
&= \frac{1}{2}\frac{d}{du}\sum_{j=1}^{\infty} j(j-1)P(N(u)=j),
\end{aligned}
$$

where $F^{(l)}(\cdot)$ is the distribution function of $f^{(l)}(\cdot)$, and $N(\cdot)$ is the counting process associated with the sampling process. From Daley (1971),

$$
\sum_{j=1}^{\infty} j(j-1)P(N(u)=j) = \int_0^u \left\{1 + 2\int_0^y H(v)dv\right\} dy - \int_0^u H(v)dv,
$$

and so

$$
\sum_{l=1}^{\infty} l f^{(l)}(u) = \frac{1}{2} + \int_0^u H(v)dv - \frac{1}{2}H(u). \tag{A.12}
$$

In view of (A.12), the sum of the second and third terms of the right side of (A.11) is absolutely bounded from above by

$$
\begin{aligned}
&2\frac{M_1}{M_3}\frac{m_n}{n}\int_0^{\infty} |C(v)| \sum_{l=1}^{\infty} l f^{(l)}(v)dv \\
&\quad \le 2\frac{M_1}{M_3}\frac{m_n}{n}\int_0^{\infty} |C(v)| \left[\frac{1}{2} + \int_0^v H(t)dt - \frac{1}{2}H(v)\right] dv \\
&\quad \le 6\frac{M_1 M_2}{M_3}\frac{m_n}{n}\int_0^{\infty} v\,|C(v)|\, dv.
\end{aligned}
$$

Thus, for sufficiently large n, we have

$$
E[\widehat{C}(u)] = \int_0^{\infty} m_n \big[\{W(m_n(u+v)) + W(m_n(u-v))\}C(v)\big] dv + O\left(\frac{m_n}{n}\right), \tag{A.13}
$$

where $O(\cdot)$ is uniform. After making a suitable transformation, we have

$$
E[\widehat{C}(u)] = \int_{-\infty}^{\infty} W(v) C\left(u - \frac{v}{m_n}\right) dv + O\left(\frac{m_n}{n}\right). \tag{A.14}
$$

Note that the covariance function of a bandlimited process is infinitely differentiable. Therefore,

$$
\begin{aligned}
&C\left(u - \frac{v}{m_n}\right) \\
&= C(u) + \frac{v}{m_n}C_{(1)}(u) + \cdots + \frac{v^{r-1}}{m_n^{r-1}(r-1)!}C_{(r-1)}(u) + \frac{v^r}{r!\, m_n^r}C_{(r)}\left(u - \frac{\alpha v}{m_n}\right),
\end{aligned} \tag{A.15}
$$

where $C_{(l)}(\cdot)$ is the lth order derivative of $C(\cdot)$ and α is an appropriate real number in the interval $(0, 1)$. By (A.15) and Property W2 of $W(\cdot)$, we have for large n,

$$E[\widehat{C}(u)] = C(u) + \frac{1}{m_n^r r!} \int_{-\infty}^{\infty} C_{(r)} \left(u - \frac{\alpha v}{m_n}\right) v^r W(v) dv + O\left(\frac{m_n}{n}\right).$$

By the DCT and the fact that $m_n \to \infty$ as $n \to \infty$,

$$E[\widehat{C}(u)] = C(u) + O\left(\frac{1}{m_n^r}\right) + O\left(\frac{m_n}{n}\right), \tag{A.16}$$

where $O(\cdot)$ is uniform. For indirect estimation of $C(lT)$ for $l = 1 \ldots, J$, we have, by using (A.1) and (A.16) in (A.7),

$$E[y_n(u)] = C(u) - \text{sinc}(\frac{\pi u}{T})C(0) - \sum_{l=J+1}^{L_n} E[\hat{C}(lT)]$$
$$\times \left\{\text{sinc}\left(\frac{\pi}{T}(u-lT)\right) + \text{sinc}\left(\frac{\pi}{T}(u+lT)\right)\right\} + O\left(\frac{1}{m_n^r}\right) + O\left(\frac{m_n}{n}\right). \tag{A.17}$$

With (A.16) and the fact that

$$\sum_{l=J+1}^{L_n} \left| \left\{\text{sinc}\left(\frac{\pi}{T}(u-lT)\right) + \text{sinc}\left(\frac{\pi}{T}(u+lT)\right)\right\} \right| \le O(\log L_n),$$

the third term of the right side of (A.17) simplifies to

$$\sum_{l=J+1}^{L_n} C(lT)\left\{\text{sinc}\left(\frac{\pi}{T}(u-lT)\right) + \text{sinc}\left(\frac{\pi}{T}(u+lT)\right)\right\} + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right).$$

Further, we have

$$E(y_n(u)) = y(u) + \sum_{l=L_n+1}^{\infty} C(lT) \left\{\text{sinc}\left(\frac{\pi}{T}(u-lT)\right) + \text{sinc}\left(\frac{\pi}{T}(u+lT)\right)\right\}$$
$$+ O\left(\frac{1}{m_n^r}\right) + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right).$$

In view of Assumption 1A,

$$\left| \sum_{l=L_n+1}^{\infty} C(lT)\left\{\text{sinc}\left(\frac{\pi}{T}(u-lT)\right) + \text{sinc}\left(\frac{\pi}{T}(u+lT)\right)\right\} \right| \le \sum_{l=L_n+1}^{\infty} |C(lT)| = O\left(\frac{1}{L_n^q}\right).$$

Thus,

$$E(y_n(u)) = y(u) + O\left(\frac{1}{L_n^q}\right) + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right), \tag{A.18}$$

where $O(\cdot)$ is uniform in u. By (2.6) and (A.18), we have for $j = 1, \dots, J$,

$$E\left[\int_{u_1}^{u_2} x_j(u) y_n(u) du\right]$$
$$= \int_{u_1}^{u_2} x_j(u) \Big\{ \sum_{l=1}^{J} x_l(u) C(lT) \Big\} du + O\Big(\frac{1}{L_n^q}\Big) + O\Big(\frac{\log L_n}{m_n^r}\Big) + O\Big(\frac{m_n \log L_n}{n}\Big). \quad \text{(A.19)}$$

By (A.19) and (2.10), for $l = 1, \dots, J$, we have

$$E[\widehat{C}(lT)] = C(lT) + O\left(\frac{1}{L_n^q}\right) + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right). \quad \text{(A.20)}$$

Using a similar argument as above, for the range $0 < u \leq d$ but $u \neq T, \dots, JT$, one can establish that

$$E[\widehat{C}(u)] = C(u) + O\left(\frac{1}{L_n^q}\right) + O\left(\frac{\log L_n}{m_n^r}\right) + O\left(\frac{m_n \log L_n}{n}\right).$$

This completes the proof.

Proof of Theorem 2. Observe that

$$\begin{aligned}\mathrm{Var}\,[\widehat{C}(0)] &= \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} E[X(t_{i_1}) X(t_{i_2})] \\ &= \frac{C(0)}{n} + \frac{1}{n^2} \int_0^\infty C(v) \sum_{1 \leq i_1 < i_2 \leq n} f^{(i_2 - i_1)}(v) dv,\end{aligned}$$

and, from Assumption 2,

$$\mathrm{Var}\,[\widehat{C}(0)] = O\Big(\frac{1}{n}\Big). \quad \text{(A.21)}$$

With Assumption 3, and writing joint moments in terms of cumulant as in Leonov and Shiryayev (1959) and Parzen (1957), we have for $u > d$,

$$\begin{aligned}E[\widehat{C}^2(u)] = &\frac{m_n^2}{n^2 H^2(u)} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} E\Big[E\Big\{W\left(m_n(u - t_{i_1} + t_{i_2})\right) \\ &\times W\left(m_n(u - t_{i_3} + t_{i_4})\right) X(t_{i_1}) X(t_{i_2}) X(t_{i_3}) X(t_{i_4}) \big| t_i, i = 1, \dots, n\Big\}\Big] \\ = &\frac{m_n^2}{n^2 H^2(u)} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} E\Big[W\left(m_n(u - t_{i_1} + t_{i_2})\right) W\left(m_n(u - t_{i_3} + t_{i_4})\right) \\ &\times \Big\{C(t_{i_1} - t_{i_2}) C(t_{i_3} - t_{i_4}) + C(t_{i_1} - t_{i_3}) C(t_{i_2} - t_{i_4}) \\ &+ C(t_{i_1} - t_{i_4}) C(t_{i_2} - t_{i_3}) + Q(t_{i_1} - t_{i_4}, t_{i_2} - t_{i_4}, t_{i_3} - t_{i_4})\Big\}\Big].\end{aligned}$$

Thus, for $u > d$,

$$\begin{aligned}\mathrm{Var}\,[\widehat{C}(u)] =& \frac{m_n^2}{n^2H^2(u)}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n} E\Big[W\left(m_n(u-t_{i_1}+t_{i_2})\right)\\ &\times W\left(m_n(u-t_{i_3}+t_{i_4})\right)\Big\{C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4}) + C(t_{i_1}-t_{i_4})C(t_{i_2}-t_{i_3})\\ &+Q(t_{i_1}-t_{i_4}, t_{i_2}-t_{i_4}, t_{i_3}-t_{i_4})\Big\}\Big]\\ =& I_1(u) + I_2(u) + I_3(u), \end{aligned} \tag{A.22}$$

where $I_1(u)$, $I_2(u)$ and $I_3(u)$ are appropriate summations. Observe that the terms $I_1(u)$ and $I_2(u)$ are bounded from above by

$$\begin{aligned}I(u) = M_1\frac{m_n}{n}\frac{1}{nH^2(u)}&\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\\ &E\Big[\left|m_nW\left(m_n(u-t_{i_1}+t_{i_2})\right)C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})\right|\Big]. \end{aligned} \tag{A.23}$$

It follows from Lemma 1 below that $I(u) = O(m_n/n)$ uniformly in u, while Lemma 2 has $I_3(u) = O(\frac{m_n}{n})$ uniformly in u. Therefore,

$$\mathrm{Var}\,(\widehat{C}(u)) = O\left(\frac{m_n}{n}\right) \quad \text{for } d < u < \infty\ . \tag{A.24}$$

Consider the indirect estimators $\widehat{C}(T), \ldots, \widehat{C}(JT)$. From (2.10), it is enough to consider the convergence of the variance covariance matrix of the vector

$$\begin{pmatrix} \int_{u_1}^{u_2} x_1(u)y_n(u)du \\ \int_{u_1}^{u_2} x_2(u)y_n(u)du \\ \vdots \\ \int_{u_1}^{u_2} x_J(u)y_n(u)du \end{pmatrix}.$$

For $j, j' \in \{1, 2, \ldots J\}$, we compute

$$\begin{aligned}&Cov\left(\int_{u_1}^{u_2} x_j(u)y_n(u)du, \int_{u_1}^{u_2} x_{j'}(u)y_n(u)du\right)\\ &\qquad = \int_{u_1}^{u_2}\int_{u_1}^{u_2} x_j(u)x_{j'}(v)Cov(y_n(u), y_n(v))dudv.\end{aligned}$$

The interchange of the integrals is justified by the finiteness of the double integral, which follows from arguments similar to those given below to establish convergence. Note that

$$\mathrm{Var}\,[y_n(u)] = \mathrm{Var}\,\Big[\widehat{C}(u) - \mathrm{sinc}(\frac{\pi u}{T})\widehat{C}(0)$$

$$
\begin{aligned}
&\quad - \sum_{l=J+1}^{L_n} \widehat{C}(lT) \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}\Bigg] \\
&\leq 3 \times \mathrm{Var}\,[\widehat{C}(u)] + 3 \times \mathrm{Var}\,[\mathrm{sinc}(\frac{\pi u}{T})\widehat{C}(0)] \\
&\quad +3 \times \mathrm{Var}\left[\sum_{l=J+1}^{L_n} \widehat{C}(lT) \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}\right] \\
&= 3 \times \mathrm{Var}\left[\sum_{l=J+1}^{L_n} \widehat{C}(lT) \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}\right] \\
&\quad +O\left(\frac{m_n}{n}\right) + O\left(\frac{1}{n}\right).
\end{aligned}
$$

The last inequality follows from (A.24) and (A.21). We now consider the first term. From (A.24), we have

$$
\begin{aligned}
&\mathrm{Var}\left[\sum_{l=J+1}^{L_n} \widehat{C}(lT) \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\}\right] \\
&= \sum_{l=J+1}^{L_n} \sum_{l'=J+1}^{L_n} Cov(\widehat{C}(lT), \widehat{C}(l'T)) \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+lT)\Big)\right\} \\
&\qquad \times \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-l'T)\Big) + \mathrm{sinc}\Big(\frac{\pi}{T}(u+l'T)\Big)\right\} \\
&\leq\ O\left(\frac{m_n}{n}\right) \left[\sum_{l=J+1}^{L_n} \left| \left\{ \mathrm{sinc}\Big(\frac{\pi}{T}(u-lT)\Big) + \mathrm{sinc}\left(\frac{\pi}{T}(u+lT)\right)\right\} \right| \right]^2 \\
&=\ O\left(\frac{m_n(\log L_n)^2}{n}\right)
\end{aligned}
$$

uniformly in u. Thus, for $j, j' \in \{1, 2, \ldots, J\}$,

$$
\begin{aligned}
&Cov\left(\int_{u_1}^{u_2} x_j(u) y_n(u) du, \int_{u_1}^{u_2} x_{j'}(u) y_n(u) du\right) \\
&\qquad =\ O\left(\frac{1}{n}\right) + O\left(\frac{m_n}{n}\right) + O\left(\frac{m_n(\log L_n)^2}{n}\right).
\end{aligned}
$$

This completes the proof.

Lemma 1. *Under the Assumptions of Theorem* 2, $I(u) = O\left(m_n/n\right)$ *uniformly in* u.

Proof of Lemma 1. We partition the range of summation as

$$\{(i_1,i_2,i_3,i_4) : 1 \le i_1,i_2,i_3,i_4 \le n\} = \bigcup_{j=1}^{24} S_{1,j} \bigcup_{j=1}^{36} S_{2,j} \bigcup_{j=1}^{8} S_{3,j} \bigcup S_4,$$

where $S_{1,j}$, $j=1,\ldots,24$, are sets of quadruples of indices having different types of strict order among themselves (thus $i_1 < i_2 < i_3 < i_4$, $i_1 < i_2 < i_4 < i_3$, and 22 other permutations), $S_{2,j}$, $j=1,\ldots,36$, are sets of quadruples of indices exactly two of which are equal and are in strict order with the other two indices ($i_1 < i_2 < i_3 = i_4$, $i_1 < i_3 = i_4 < i_2$, and 34 other arrangements), $S_{3,j}$, $j=1,\ldots,8$, are sets of quadruples of indices exactly three of which are equal and are in strict order with the fourth ($i_1 < i_2 = i_3 = i_4$, $i_2 = i_3 = i_4 < i_1$, and 6 other arrangements), and S_4 is the set $\{(i_1,i_2,i_3,i_4) : 1 \le i_1 = i_2 = i_3 = i_4 \le n\}$.

Consider $S_{1,1} = \{(i_1,i_2,i_3,i_4) : 1 \le i_1 < i_2 < i_3 < i_4 \le n\}$. By using the transformation $t_{i_2} - t_{i_1} = \vartheta_{i_2-i_1}$, $t_{i_3} - t_{i_2} = \vartheta_{i_3-i_2}$ and $t_{i_4} - t_{i_3} = \vartheta_{i_4-i_3}$, and by the independence of the transformed random variables, we have

$$\begin{aligned}
&\frac{M_1}{nH^2(u)}\sum_{S_{1,1}} E\Big[|m_n W(m_n(u-t_{i_1}+t_{i_2}))\, C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})|\Big] \\
&= \frac{M_1}{nH^2(u)}\sum_{S_{1,1}} E\Big[|m_n W\,(m_n(u-\vartheta_{i_2-i_1}))\, C(\vartheta_{i_3-i_2}+\vartheta_{i_2-i_1})C(\vartheta_{i_4-i_3}+\vartheta_{i_3-i_2})|\Big] \\
&= \frac{M_1}{nH^2(u)}\sum_{S_{1,1}} \int_0^\infty\int_0^\infty\int_0^\infty |m_n W\,(m_n(u-v_1))|\,|C(v_2+v_1)C(v_3+v_2)| \\
&\qquad\qquad \times f^{(i_2-i_1)}(v_1) f^{(i_3-i_2)}(v_2) f^{(i_4-i_3)}(v_3)\, dv_1 dv_2 dv_3 \\
&= \frac{M_1}{H^2(u)} \int_0^\infty\int_0^\infty\int_0^\infty |m_n W(m_n(u-v_1))\, C(v_2+v_1)C(v_3+v_2)| \\
&\qquad\qquad \times\left\{\frac{1}{n}\sum_{S_{1,1}} f^{(i_2-i_1)}(v_1) f^{(i_3-i_2)}(v_2) f^{(i_4-i_3)}(v_3)\right\} dv_1 dv_2 dv_3.
\end{aligned}$$

Now note that

$$\frac{1}{n}\sum_{S_{1,1}} f^{(i_2-i_1)}(v_1) f^{(i_3-i_2)}(v_2) f^{(i_4-i_3)}(v_3) \le H(v_1)H(v_2)H(v_3) \;\le\; M_2^3.$$

Thus,

$$\begin{aligned}
&\frac{M_1}{nH^2(u)}\sum_{S_{1,1}} E\Big[\,|m_n W(m_n(u-t_{i_1}+t_{i_2}))\, C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})|\,\Big] \\
&\le \frac{M_1 M_2^3}{M_3^2} \int_0^\infty\int_0^\infty\int_0^\infty |m_n W\,(m_n(u-v_1))\, C(v_2+v_1)C(v_3+v_2)|\, dv_1 dv_2 dv_3.
\end{aligned}$$

With similar arguments, one can establish the boundedness of

$$\frac{M_1}{nH^2(u)}\sum_{S_{1,j}} E\left[|m_n W(m_n(u-t_{i_1}+t_{i_2}))\, C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})| \right]$$

for the partitions $S_{1,j}$, $2 \le j \le 22$. A slightly different argument is needed for $S_{1,23} = \{(i_1, i_2, i_3, i_4) : 1 \le i_1 < i_4 < i_3 < i_2 \le n\}$ and $S_{1,24} = \{(i_1, i_2, i_3, i_4) : 1 \le i_2 < i_3 < i_4 < i_1 \le n\}$. Consider the case of $S_{1,24}$. With $\vartheta_{i_4-i_1} = t_{i_4} - t_{i_1}$, $\vartheta_{i_3-i_4} = t_{i_3} - t_{i_4}$, and $\vartheta_{i_2-i_3} = t_{i_2} - t_{i_3}$, and using the fact that $\vartheta_{i_4-i_1}$, $\vartheta_{i_3-i_4}$ and $\vartheta_{i_2-i_3}$ are independent random variables, we have

$$\begin{aligned}
&\frac{M_1}{nH^2(u)}\sum_{S_{1,23}} E\left[|m_n W\left(m_n(u-t_{i_1}+t_{i_2})\right) C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})| \right] \\
&\quad= \frac{M_1}{nH^2(u)}\sum_{S_{1,23}} E\Big[|m_n W\left(m_n(u+\vartheta_{i_4-i_1}+\vartheta_{i_3-i_4}+\vartheta_{i_2-i_3})\right)| \\
&\qquad\qquad\qquad \times |C(\vartheta_{i_4-i_1}+\vartheta_{i_3-i_4})C(\vartheta_{i_3-i_4}+\vartheta_{i_2-i_3})| \Big] \\
&\quad= \frac{M_1}{H^2(u)}\int_0^\infty\int_0^\infty\int_0^\infty |m_n W\left(m_n(u+v_1+v_2+v_3)\right)|\, |C(v_1+v_2)| \\
&\qquad \times |C(v_2+v_3)| \left\{ \frac{1}{n}\sum_{S_{1,23}} f^{(i_4-i_1)}(v_1) f^{(i_3-i_4)}(v_2) f^{(i_2-i_3)}(v_3) \right\} dv_1 dv_2 dv_3 \\
&\quad\le \frac{M_1 M_2^3}{M_3^2}\int_0^\infty\int_0^\infty\int_0^\infty |m_n W\left(m_n(u+v_1+v_2+v_3)\right)| \\
&\qquad\qquad\qquad \times |C(v_1+v_2)C(v_2+v_3)|\, dv_1 dv_2 dv_3.
\end{aligned}$$

The boundedness of the sum over $S_{1,23}$ can be established in a similar manner.

In subsets $S_{2,j}$ for $j = 1, \ldots, 36$, the summation runs over only three indices. Consider $S_{2,1} = \{(i_1, i_2, i_3, i_4) : i_1 < i_2 < i_3 = i_4\}$, and the transformation $\vartheta_{i_2-i_1} = t_{i_2} - t_{i_1}$ and $\vartheta_{i_3-i_2} = t_{i_3} - t_{i_2}$. Then

$$\begin{aligned}
&\frac{M_1}{nH^2(u)}\sum_{S_{2,1}} E\left[|m_n W(m_n(u-t_{i_1}+t_{i_2}))\, C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})| \right] \\
&\quad= \frac{M_1}{nH^2(u)}\sum_{S_{2,1}} E\left[|m_n W\left(m_n(u+\vartheta_{i_2-i_1})\right)|\, |C(\vartheta_{i_2-i_1}+\vartheta_{i_3-i_2})C(\vartheta_{i_3-i_2})| \right] \\
&\quad= \frac{M_1}{H^2(u)}\int_0^\infty\int_0^\infty |m_n W\left(m_n(u+v_1)\right) C(v_1+v_2)C(v_2)| \\
&\qquad\qquad\qquad \times \left\{ \frac{1}{n}\sum_{S_{2,1}} f^{(i_2-i_1)}(v_1) f^{(i_3-i_2)}(v_2) \right\} dv_1 dv_2
\end{aligned}$$

$$\leq \frac{M_1}{H^2(u)}\int_0^\infty\int_0^\infty |m_nW\left(m_n(u+v_1)\right)C(v_1+v_2)C(v_2)|\,H(v_1)H(v_2)dv_1dv_2$$
$$\leq \frac{M_1M_2^2C(0)}{M_3^2}\int_0^\infty m_n|W\left(m_n(u+v_1)\right)|\int_0^\infty |C(v_2)|dv_2dv_1.$$

A similar argument can be used to establish the boundedness of the sums over 29 other sets of quadruples of indices with $i_1 \neq i_2$. A slightly different argument is needed in the cases of the six sets with $i_1 = i_2$. We show the calculations for $S_{2,31} = \{(i_1,i_2,i_3,i_4) : i_1 = i_2 < i_3 < i_4\}$, as representative. By using the transformation $\vartheta_{i_3-i_2} = t_{i_3} - t_{i_2}$ and $\vartheta_{i_4-i_3} = t_{i_4} - t_{i_3}$, we have

$$\frac{M_1}{nH^2(u)}\sum_{S_{2,31}}E\Big[|m_nW(m_n(u-t_{i_1}+t_{i_2}))\,C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})|\Big]$$
$$= \frac{M_1}{nH^2(u)}\sum_{S_{2,31}}E\Big[\,|m_nW\left(m_nu\right)C(\vartheta_{i_3-i_2})C(\vartheta_{i_3-i_2}+\vartheta_{i_4-i_3})|\,\Big]$$
$$= \frac{M_1}{H^2(u)}\cdot m_n|W\left(m_nu\right)|\int_0^\infty\int_0^\infty |C(v_1)C(v_1+v_2)|$$
$$\times\left\{\frac{1}{n}\sum_{S_{2,31}}f^{(i_3-i_2)}(v_1)f^{(i_4-i_3)}(v_2)\right\}dv_1dv_2$$
$$\leq \frac{M_1M_2^2}{M_3^2}\cdot m_n|W\left(m_nu\right)|\int_0^\infty\int_0^\infty |C(v_1)C(v_1+v_2)|dv_1dv_2.$$

For sufficiently large n (such that $m_n > a/d$), the last expression is identically zero. Since the threshold a/d does not depend on u, this term is identically zero for large n, uniformly for all u.

Now consider the double sums over the subsets $S_{3,1},\ldots,S_{3,8}$. We show the boundedness of the sums over $S_{3,1} = \{(i_1,i_2,i_3,i_4) : 1 \leq i_1 = i_2 = i_3 < i_4 \leq n\}$ and $S_{3,2} = \{(i_1,i_2,i_3,i_4) : 1 \leq i_1 < i_2 = i_3 = i_4 \leq n\}$, each case being representative of the calculations needed. For $S_{3,1}$,

$$\frac{M_1}{nH^2(u)}\sum_{S_{3,1}}E\Big[\,|m_nW(m_n(u-t_{i_1}+t_{i_2}))\,C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})|\Big]$$
$$= \frac{M_1}{H^2(u)}\int_0^\infty |m_nW\left(m_nu\right)C(0)C(v)|\left\{\frac{1}{n}\sum_{S_{3,1}}f^{(i_4-i_3)}(v)\right\}dv$$
$$\leq \frac{M_1M_2}{M_3^2}m_n|W\left(m_nu\right)|C(0)\int_0^\infty |C(v)|\,dv.$$

For sufficiently large n (such that $m_n > a/d$), the last expression is identically zero. As the threshold a/d does not depend on u, this term is also identically zero for large n, uniformly for all u. On the other hand,

$$\begin{aligned}
&\frac{M_1}{nH^2(u)}\sum_{S_{3,2}}E\Big[|m_nW(m_n(u-t_{i_1}+t_{i_2}))\,C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})|\Big]\\
&=\frac{M_1}{H^2(u)}\int_0^\infty |m_nW\left(m_n(u+v)\right)C(v)C(0)|\left\{\frac{1}{n}\sum_{S_{3,2}}f^{(i_2-i_1)}(v)\right\}dv\\
&\le\frac{M_1M_2}{M_3^2}C^2(0)\int_0^\infty |W(v)|\,dv.
\end{aligned}$$

The sum over S_4 does not involve any random variables, and is bounded as

$$\begin{aligned}
&\frac{M_1}{nH^2(u)}\sum_{S_4}E\Big[\,|m_nW(m_n(u-t_{i_1}+t_{i_2}))\,C(t_{i_1}-t_{i_3})C(t_{i_2}-t_{i_4})|\,\Big]\\
&\quad\le\frac{M_1}{nM_3^2}m_nW\left(m_nu\right)C^2(0).
\end{aligned}$$

Again for large n such that $m_n > a/d$, the upper bound is identically zero. This completes the proof.

Lemma 2. Under the Assumptions of Theorem 2, $I_3(u) = O\left(m_n/n\right)$, where $O(\cdot)$ is uniform in u.

Proof of Lemma 2. From Assumption 3, $I_3(u)$ is bounded as

$$\begin{aligned}
&\frac{n}{m_n}|I_3(u)|\\
&\le\frac{M_1}{nM_3^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}E\big[m_n|W\left(m_n(u-t_{i_1}+t_{i_2})\right)|\\
&\qquad\qquad\times g_1(t_{i_1}-t_{i_4})g_2(t_{i_2}-t_{i_4})g_3(t_{i_3}-t_{i_4})\big]\\
&\le\frac{M_1g_1(0)}{nM_3^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}E\big[m_n|W\left(m_n(u-t_{i_1}+t_{i_2})\right)|g_2(t_{i_2}-t_{i_4})g_3(t_{i_3}-t_{i_4})\big].
\end{aligned}$$

The proof follows from an argument similar to the one used in the proof of Lemma 1.

Proof of Theorem 3. Part (i). From (2.11), for $\lambda\in[-\lambda_0,\lambda_0]$, we have

$$\begin{aligned}
E[\widehat{\phi}_n(\lambda)] &= \frac{T}{2\pi}E[\widehat{C}(0)]+\frac{T}{\pi}\sum_{l=1}^{J}E[\widehat{C}(lT)]K(b_nl)\cos(l\lambda T)\\
&\quad+\frac{T}{\pi}\sum_{l=J+1}^{[n\mu_f/T]}E[\widehat{C}(lT)]K(b_nl)\cos(l\lambda T).
\end{aligned}\tag{A.25}$$

Note that the second term on the right side of (A.25) is a finite sum. By using Theorem 1 Part (i), Property K1 of the kernel $K(\cdot)$, and the fact $b_n \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} E[\widehat{\phi}_n(\lambda)] = \frac{T}{2\pi}C(0) + \frac{T}{\pi}\sum_{l=1}^{J} C(lT)\cos(l\lambda T) + \lim_{n\to\infty}\frac{T}{\pi}\sum_{l=J+1}^{[n\mu_f/T]} E[\widehat{C}(lT)]K(b_n l)\cos(l\lambda T). \tag{A.26}$$

With the exact expectation of $E(\widehat{C}(lT))$ for $l > J$ from (A.5), for large n we have

$$\frac{T}{\pi}\sum_{l=J+1}^{[n\mu_f/T]} E[\widehat{C}(lT)]K(b_n l)\cos(l\lambda T) = \frac{T}{\pi}\int_{-\infty}^{\infty} W(v)\Bigg\{\sum_{l=J+1}^{[n\mu_f/T]} K(b_n l)\cos(l\lambda T)\frac{1}{H(lT)}C\Big(lT-\frac{v}{m_n}\Big) \times\left[H_n\left(lT-\frac{v}{m_n}\right)+H_n\left(-lT+\frac{v}{m_n}\right)\right]\Bigg\}dv.$$

From Assumptions 1, 2 and Property K1 of $K(\cdot)$, the integrand is dominated as

$$|W(v)|\sup|K(\cdot)|\frac{2M_2}{M_3}\sum_{l=J+1}^{[n\mu_f/T]} g\left(lT-\frac{v}{m_n}\right) \le |W(v)|\sup K(\cdot)\frac{2M_2}{M_3}\times 2\sum_{l=0}^{\infty} g\,(lT)\,.$$

Integrability of the bounding function is ensured by Property W1 of $W(\cdot)$. As $m_n \to \infty$ and $b_n \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} W(v)\sum_{l=J+1}^{[n\mu_f/T]} K(b_n l)\cos(l\lambda T)\frac{1}{H(lT)}C\Big(lT-\frac{v}{m_n}\Big) \times\Big[H_n\Big(lT-\frac{v}{m_n}\Big)+H_n\Big(-lT+\frac{v}{m_n}\Big)\Big] = W(v)\sum_{l=J+1}^{\infty}\cos(l\lambda T)C(lT).$$

Applying the DCT, we have

$$\lim_{n\to\infty}\frac{T}{\pi}\sum_{l=J+1}^{[n\mu_f/T]} E[\widehat{C}(lT)]K(b_n l)\cos(l\lambda T) = \frac{T}{\pi}\sum_{l=J+1}^{\infty} C(lT)\cos(l\lambda T),$$

and from (A.26),

$$\lim_{n\to\infty} E[\widehat{\phi}_n(\lambda)] = \frac{T}{2\pi}C(0) + \frac{T}{\pi}\sum_{l=1}^{\infty} C(lT)\cos(l\lambda T) = \phi(\lambda).$$

This completes the proof of Part (i).

Part (ii). From Theorem 1 (ii), we have

$$E[\widehat{\phi}_n(\lambda)] = \frac{T}{2\pi}C(0)+\frac{T}{\pi}\sum_{l=1}^{[n\mu_f/T]} C(lT)K(b_nl)\cos(l\lambda T)$$
$$+\Big\{\sum_{l=1}^{J} K(b_nl)\cos(l\lambda T)\Big\} \times \Big[O\Big(\frac{1}{L_n^q}\Big)+O\left(\frac{\log L_n}{m_n^r}\right)+O\left(\frac{m_n \log L_n}{n}\right)\Big]$$
$$+\Big\{\sum_{l=J+1}^{[n\mu_f/T]} K(b_nl)\cos(l\lambda T)\Big\} \times O\Big(\frac{1}{m_n^r}\Big).$$

Note that

$$\sum_{l=J}^{[n\mu_f/T]} K(b_nl)\cos(l\lambda T) = O\left(\frac{1}{b_n}\right).$$

Since

$$\phi(\lambda) = \frac{T}{2\pi}C(0)+\frac{T}{\pi}\sum_{l=1}^{\infty} C(lT)\cos(l\lambda T),$$

we have

$$E[\widehat{\phi}_n(\lambda)]-\phi(\lambda)=-\frac{b_n^qT}{\pi}\sum_{l=1}^{[n\mu_f/T]}\frac{(1-K(b_nl))}{(b_nl)^q}l^qC(lT)\cos(l\lambda T)-\frac{T}{\pi}\sum_{l=n}^{\infty}C(lT)\cos(l\lambda T)$$
$$+O\Big(\frac{1}{L_n^q}\Big)+O\Big(\frac{\log L_n}{m_n^r}\Big)+O\Big(\frac{m_n\log L_n}{n}\Big)+O\Big(\frac{1}{b_nm_n^r}\Big).$$

By Assumption 1A,

$$\Big|\sum_{l=n}^{\infty} C(lT)\cos(l\lambda T)\Big| \le \sum_{l=n}^{\infty}|C(lT)| = O\left(\frac{1}{n^q}\right).$$

Using Assumption 1A, Property K2 of $K(\cdot)$, and the DCT, we have

$$\lim_{n\to\infty}\sum_{l=1}^{[n\mu_f/T]}\frac{(1-K(b_nl))}{(b_nl)^q}l^qC(lT)\cos(l\lambda T) = \frac{k_q}{T^q}\sum_{l=1}^{\infty}(lT)^qC(lT)\cos(l\lambda T),$$

where $k_q = \lim_{n\to 0}(1-K(x))/|x|^q$, which is non-zero for a qth order kernel $K(\cdot)$. Thus, we have

$$E[\widehat{\phi}_n(\lambda)] - \phi(\lambda) = O(b_n^q) + O\Big(\frac{1}{n^q}\Big) + O\Big(\frac{1}{L_n^q}\Big) + O\Big(\frac{\log L_n}{m_n^r}\Big)$$

$$+O\Big(\frac{m_n \log L_n}{n}\Big) + O\Big(\frac{1}{b_n m_n^r}\Big).$$

The proof is completed by observing that, in view of Assumptions 1A and 3 and the fact $L_n \to \infty$ as $n \to \infty$, the second term on the right side can be ignored in the presence of the fifth term.

Proof of Theorem 4. From (2.11), we have

$$\begin{aligned}\mathrm{Var}\,[\widehat{\phi}_n(\lambda)] \le \frac{3T^2}{(2\pi)^2}\mathrm{Var}\,[\widehat{C}(0)] + \frac{3T^2}{\pi^2}\mathrm{Var}\Bigg[\sum_{l=1}^{J}\widehat{C}(lT)K(b_n l)\cos(l\lambda T)\Bigg] \\ +\frac{3T^2}{\pi^2}\mathrm{Var}\Bigg[\sum_{l=J+1}^{[n\mu_f/T]}\widehat{C}(lT)K(b_n l)\cos(l\lambda T)\Bigg]. \qquad \text{(A.27)}\end{aligned}$$

From (A.21), the first term on the right side is $O\,(1/n)$. The second term of right side of (A.27) is a finite sum, and we have from Theorem 2,

$$\mathrm{Var}\Bigg[\sum_{l=1}^{J}\widehat{C}(lT)K(b_n l)\cos(l\lambda T)\Bigg] = O\left(\frac{m_n(\log L_n)^2}{n}\right).$$

Now consider the third term. From Theorem 2,

$$\begin{aligned}&\mathrm{Var}\Bigg[\sum_{l=J+1}^{[n\mu_f/T]}\widehat{C}(lT)K(b_n l)\cos(l\lambda T)\Bigg] \\ &= \sum_{l=J+1}^{[n\mu_f/T]}\sum_{l'=J+1}^{[n\mu_f/T]} Cov(\widehat{C}(lT),\widehat{C}(l'T))K(b_n l)\cos(l\lambda T)K(b_n l')\cos(l'\lambda T) \\ &\le \Bigg\{\sum_{l=J+1}^{[n\mu_f/T]}|K(b_n l)|\Bigg\}^2 \times O\left(\frac{m_n}{n}\right).\end{aligned}$$

By Property K1 of $K(\cdot)$, we have

$$\sum_{l=J+1}^{[n\mu_f/T]}|K(b_n l)| = O\left(\frac{1}{b_n}\right).$$

Thus, since $m_n \to \infty$ and $b_n \to 0$ as $n \to \infty$,

$$\mathrm{Var}\Bigg[\sum_{l=J+1}^{[n\mu_f/T]}\widehat{C}(lT)K(b_n l)\cos(l\lambda T)\Bigg] = O\left(\frac{m_n}{nb_n^2}\right).$$

This completes the proof.

References

Bardet, J. M. and Bertrand, P. R. (2010). A non-parametric estimator of the spectral density of a continuous-time Gaussian process observed at random times. *Scand. J. Statist.* **37**, 458-476.

Beltrão, K. I. and Bloomfield, P. (1985). Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* **8**, 21-38.

Brillinger, D. R. (1972). The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Prob. Statist.* 483-513.

Bronez, T. P. (1988). Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences. *IEEE Trans. Acoustics Speech Signal Process.* **36**, 1862-1873.

Buetler, F. J. (1970). Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inform. Theory* **16**, 147-152.

Cook, R. L. (1986). Stochastic sampling in computer graphics. *ACM Trans. Graphics* **5**, 51-72.

Costain, J. K. and Çoruh, C. (2004). *Basic Theory of Exploration Seismology.* Elsevier, Amsterdam.

Daley, D. J. (1971). Weakly stationary point processes and random measures. *J. Roy. Statist. Soc. Ser. B* **33**, 406-428.

Dippé, M. A. Z. and Wold, E. H. (1985). Antialiasing through stochastic sampling. *ACM Siggraph* **19**, 69-78.

Dunsmuir, W. (1983). A central limit theorem for estimation in Gaussian stationary time series observed at unequally spaced times. *Stochastic Process. Appl.* **14**, 279-295.

Eldar, Y., Lindenbaum, M., Porat, M. and Zeevi, Y. Y. (1997). The farthest point strategy for progressive image sampling. *IEEE Trans. Image Processing* **6**, 1305-1315.

Guillot, G., Senoussi, R. and Monestiez, P. (2001). A positive definite estimator of the nonstationary covariance of random fields. *In geoENV-III – Geostatistics for Environmental Applications* (P. Moneztiez, D. Allard and R. Froidevauxeds., eds.), 333-344. Kluwer Academic Publishers.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **66**, 51-83.

Huang, C., Hsing, T. and Cressie, N. (2011). Spectral density estimation through a regularized inverse problem. *Statist. Sinica* **21**, 1115-1144.

Hurvich, C. M. (1985). Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods. *J. Amer. Statist. Assoc.* **80**, 933-940.

Ishimaru, A. and Chen, Y. (1965). Thinning and broadbanding antenna arrays by unequal spacings. *IEEE Trans. Antenna Propagation* **AP-13**, 208-215.

Kay, S. M. (1999). *Modern Spectral Estimation: Theory and Application.* Prentice Hall, Englewood Cliffs, New Jersey.

Leonov, V. P. and Shiryayev, A. N. (1959). On a method of calculation of semi-invariants, *Theory Probab. Appl.* **4**, 319-329.

Lii, K. S. and Masry, E. (1994). Spectral estimation of continuous-time stationary processes from random sampling. *Stochastic Process. Appl.* **52**, 39-64.

Marvasti, F. A. (2001). *Nonuniform Sampling.* Kluwer Plenum, New York.

Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Statist. Soc. Ser. B* **71**, 191-217.

Moore, M. I., Visser, A. W. and Shirtcliffe, T. L. G. (2008). Experiences with the Brillinger spectral estimator applied to simulated irregularly observed process. *Time Ser. Anal.* **8**, 433-442.

Munson, D. C., Jr., O'Brien, J. D. and Jenkins, W. K. (1983). A tomographic formulation of spotlight-mode synthetic aperture radar. *Proc. IEEE* **71**, 917-925.

Nobach, H., Müller, E. and Tropea, C. (1998). Efficient estimation of power spectral density from laser Doppler anemometer data. *Experiments in Fluids* **24**, 499-509.

Nohrden, J. M. (1995). Spectral Estimation Using Nonuniform Sampling. Master's Thesis, Department of Electrical and Computer Engineering, Massachusetts Institute of Technology, Boston.

Oppenheim, A. V. and Schafer, R. W. (2009). *Discrete-Time Signal Processing.* 3rd edition. Pentice-Hall, Upper Saddle River, N. J.

Ouahabi, A., Depollier, C., Simon, L. and Kouamé, D. (1998). Spectrum estimation from randomly sampled velocity data. *IEEE Trans. Instru. Measur.* **47**, 1005-1012.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series, *Ann. Math. Statist.* **28**, 329–348.

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J. M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., épin, L., Ritz, C., Saltzman, E. and Stievenard, M. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429-435.

Qu, D. and Tarczynski, A. (2007). A novel spectral estimation method by using periodic nonuniform sampling. *Forty-First Asilomar Conference on ACSSC*, 1134-1138.

Ramsay, J. O., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and Matlab.* New York, Spriger.

Roughan, M. (2006). A comparison of Poisson and Uniform sampling for active measurements. *IEEE Trans. Selected Areas Communication* **24**, 2299-2312.

Shapiro, H. S. and Silverman, R. A. (1960). Alias-free sampling of random noise. *J. Soc. Indust. Appl. Math* **8**, 225-248.

Srivastava, R. and Sengupta, D. (2010). Consistent estimation of non-bandlimited spectral density from uniformly spaced samples. *IEEE Trans. Inform. Theory* **56**, 3642-3659.

Srivastava, R. and Sengupta, D. (2011). Effect of inter-sample spacing constraint on spectrum estimation with irregular sampling. *IEEE Trans. Inform. Theory* **57**, 4709-4719.

Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. Roy. Statist. Soc. Ser. B* **66**, 275-296.

Tarczynski, A. and Allay, N. (2004). Spectral Analysis of Randomly Sampled Signals: Suppression of Aliasing and Sampler Jitter. *IEEE Trans. Signal Process.* **52**, 3324-3334.

Department of Statistical Science, Cornell University, USA.

E-mail: rs899@cornell.edu

Applied Statistics Unit, Indian Statistical Institute, Kolkata, India.

E-mail: sdebasis@isical.ac.in

(Received November 2011; accepted July 2012)