NONPARAMETRIC SPECTRAL DENSITY ESTIMATION WITH MISSING OBSERVATIONS

Thomas C. M. Lee[*] Senior Member, IEEE

Chinese University of Hong Kong
Department of Statistics
Shatin, Hong Kong

Zhengyuan Zhu[†]

University of North Carolina at Chapel Hill
Department of Statistics & Operations Research
Chapel Hill, NC 27599-3260

ABSTRACT

Self-consistency is a fundamental principle in statistics for retaining maximum amount of information in the data. In this paper this principle is applied to develop a new method for nonparametric spectrum estimation with missing data. One major advantage of the proposed method is that it can be coupled with any complete data nonparametric spectrum estimation procedure, including kernel smoothing, wavelet and spline estimators. The practical performance of the method is illustrated by a simulation study.

Index Terms— missing data, nonparametric spectrum estimation, periodogram smoothing, self-consistency

1. INTRODUCTION

Spectral density estimation is an important problem that arises in many different application areas, including astrophysics, communications and geology, just to name a few. Very often, data collected from these applications contain missing values. This forbids the use of many powerful periodogram-based nonparametric spectral analysis techniques, as the discrete Fourier transform cannot be directly applied to calculate the periodogram when part of the data is missing.

To overcome this issue, different methods for spectral density estimation with missing data have been proposed. In the astrophysics literature, a popular nonparametric approach for spectral estimation with missing data is the Lomb-Scargle method [11, 16]. This method "re-constructs" the complete data set by imputing the missing values with the mean of the observed data, so that the discrete Fourier transform can be applied. A potential shortcoming of this time-domain mean-imputation is that high frequency information in the data may be destroyed. See also [1] and [4] for some theoretical investigations. More recently, a nonparametric missing data spectral estimation procedure that uses the EM-Algorithm was developed by [19]. However, as noted by [2], although this new nonparametric EM-Algorithm performs well in terms of detect strong spectral peaks, it performs poorly at regions where the spectrum is of low energy.

In this paper we propose a new method for nonparametric spectral density estimation with missing data. In particular, we adopt the self-consistency principle to optimally impute the missing data, so that an "imputed periodogram" can be obtained by applying the discrete Fourier transform to the imputed complete data. As illustrated by [18], self-consistency is a fundamental concept in statistics, and is a powerful and widely applicable statistical principle for retaining as much as possible the information in the data. It has been applied to construct the most efficient estimation procedures in many different contexts, including maximum likelihood estimation via the EM-Algorithm in the parametric setting, and the Kaplan-Meier estimator [8] in the nonparametric setting. Therefore it is not unexpected that our proposed method, as demonstrated below, also possesses excellent statistical properties.

In addition, another advantage of our approach is that it is extremely flexible, in the sense that it can be coupled with any complete data nonparametric spectrum estimation techniques, such as kernel smoothing, wavelet regression, or spline estimators. It can also be straightforwardly extended to higher dimensional settings, or to parametric or semiparametric modeling. For simplicity, we shall illustrate our approach with kernel estimators.

The rest of this paper is organized as follows. Background material that is required from the rest of the paper is first reviewed in Section 2. Then the proposed method for spectral density estimation with missing data is presented in Section 3. The empirical performance of the proposed method is illustrated via numerical experiments in Section 4. Lastly, concluding remarks are offered in Section 5.

2. BACKGROUND

2.1. Spectral Density Estimation with Complete Data

We first review the basic ingredients for spectrum estimation by smoothing the periodogram when there is no missing data. Let $Y_0, \ldots, Y_{2n-1}$ be a finite-sized realization of a

[*]Corresponding author. Supported in part by CUHK Direct Grant, HKRGC under CERG 401507, and NSF under Grant 0707037. Also affiliated with the Department of Statistics, Colorado State University.

[†]Supported in part by NSF under DMS Grant 0605434.

real-valued, zero mean stationary process $\{Y_t\}$ with unknown spectrum f. To estimate f nonparametrically, one typically starts by calculating the periodogram, defined as

$$I(\omega) = \frac{1}{2\pi \times 2n} \left| \sum_{t=0}^{2n-1} Y_t \exp(-\mathrm{i}\omega t) \right|^2, \quad \omega \in [0, 2\pi).$$

To simplify notation, write $\omega_j = \pi j/n$, $f_j = f(\omega_j)$ and $I_j = I(\omega_j)$. Since the spectrum f is symmetric about $\omega = \pi$, the rest of this article shall focus on f_j for $j = 0, \ldots, n-1$. Also, as f is periodic with period 2π, we have $f_{-j} = f_j$ and $I_{-j} = I_j$ for $j = 1, \ldots, n-1$.

In this paper we adopt the following model for I_j:

$$I_j = f_j \epsilon_j, \qquad j = 0, \ldots, n-1, \tag{1}$$

where the ϵ_j's are independent standard exponential random variables. This model has been used by many previous authors [e.g., 9, 14, 15] Under model (1), it is straightforward to see $E(I_j) = f_j$ and $\mathrm{Var}(I_j) = f_j^2$. Given its large variance, I_j is seldom used as an estimate of f_j. One approach to reduce the variance and obtain a consistent estimate for f is to apply kernel smoothing to the periodogram, as follows. Let $K(\cdot)$ be a univariate kernel function and h be a nonnegative smoothing parameter that controls the amount of smoothing. The kernel $K(\cdot)$ is often taken as a symmetric density function. Write $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$. Then the kernel estimator $\hat{f}_j$ for f_j is given by

$$\hat{f}_j = \sum_{m=-n}^{2n-1} K_h(\omega_m - \omega_j) I_m \Big/ \sum_{l=-n}^{2n-1} K_h(\omega_l - \omega_j) \tag{2}$$

for $j = 0, \ldots, n-1$. It is widely known that the choice of h plays an important role in the quality of $\hat{f}_j$. Automatic methods for selecting h can be found for examples in [6, 15] and references given therein. We note that $\hat{f}_j$ is a function of h, but for simplicity this dependence is suppressed in its notation. In many other kernel smoothing problems the summation limits in both the numerator and denominator in (2) are 0 and $n-1$. However, since for the present problem boundary effects can be handled by periodic smoothing, the limits are changed from 0 and $(n-1)$ to $(-n)$ and $(2n-1)$, respectively.

Lastly, we note that if the autocovariance function of $\{Y_t\}$ is denoted as $\gamma(k) = \mathrm{Cov}(Y_t, Y_{t+k})$, then we have

$$\gamma(k) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega. \tag{3}$$

2.2. The Self-Consistency Principle

One important task in signal processing is to obtain a concise summary of data while at the same time retain as much information as possible. Suppose we wish to approximate a target random vector $\mathbf{X}$ by a second random vector $\mathbf{Z}$ that is simpler in structure. If we quantify the approximation error with the measure $E\|\mathbf{X} - \mathbf{Z}\|^2$, then the approximation can always be improved by using $E(\mathbf{X}|\mathbf{Z})$ instead of $\mathbf{Z}$. It is because, for any function $g(\cdot)$ including the identity, we have $E\|\mathbf{X} - E(\mathbf{X}|\mathbf{Z})\|^2 \le E\|\mathbf{X} - g(\mathbf{Z})\|^2$. This leads to the following definition for self-consistency provided by [18]: For two jointly distributed random vectors $\mathbf{X}$ and $\mathbf{Z}$, $\mathbf{Z}$ is *self-consistent* for $\mathbf{X}$ if

$$E(\mathbf{X}|\mathbf{Z}) = \mathbf{Z} \quad \text{almost surely.}$$

In fact, such notion of self-consistency can be dated back at least as early as [5], and has been applied successfully to construct many efficient estimators, and to define various linear and nonlinear data summaries [e.g., 7, 17]. More recently, it has also been applied to derive wavelet image denoising with missing data [10]. Of course, some most fruitful and influential outcomes originated from the applications of the self-consistency principle are the EM-Algorithm and its various generalizations [e.g., 12, 13], for which when the principle is applied to the complete data score function, it frequently leads to the incomplete data maximum likelihood estimator. Given these successful examples, in below we will apply this self-consistency principle to construct our nonparametric spectral density estimator for missing data.

3. THE PROPOSED METHOD

First we fix the notation. In sequel we write the complete data as $\mathbf{Y}_{\mathrm{com}} = (Y_0, \ldots, Y_{2n-1})$, and denote the observed and missing portions of $\mathbf{Y}_{\mathrm{com}}$ as $\mathbf{Y}_{\mathrm{obs}}$ and $\mathbf{Y}_{\mathrm{mis}}$, respectively. With this notation our goal is to estimate f when only $\mathbf{Y}_{\mathrm{obs}}$ is available.

Our proposal begins with the assumption that an efficient complete data spectrum estimation method is available for estimating f if $\mathbf{Y}_{\mathrm{com}}$ is given. In our framework no structural assumption is made about this complete data estimation method: it can be parametric, nonparametric, or even semiparametric, although our focus will be on nonparametric. Denote the corresponding estimate for f obtained by this complete data method as $\hat{f}_{\mathrm{com}}$. Then, when only $\mathbf{Y}_{\mathrm{obs}}$ is available, we define our missing data estimate $\hat{f}$ for f as the solution to the following self-consistent equation:

$$E\left\{ \hat{f}_{\mathrm{com}}(\cdot) \middle| \mathbf{Y}_{\mathrm{obs}}, f = \hat{f} \right\} = \hat{f}(\cdot). \tag{4}$$

The solution of this self-consistent equation, as demonstrated next, can be approximated numerically via an iterative process, as long as the missing portion of $\mathbf{Y}_{\mathrm{com}}$ can be (conditionally) simulated given $\mathbf{Y}_{\mathrm{obs}}$ and an estimate of f.

To solve (4), one could iterate the between following two steps. The first step is to, given a current estimate $\hat{f}$, calculate the conditional expectation on the left-hand side, while the second step is to set the next iterative estimate for f as this calculated conditional expectation. Unfortunately, for the present problem the conditional expectation cannot always be

calculated analytically, and one could use Monte Carlo approximation to overcome this issue. This leads to our proposal, the MCSC-Algorithm, for nonparametric spectrum estimation with missing data.

Algorithm 1 *Monte-Carlo Self-Consistent (MCSC) Algorithm for Nonparametric Spectral Density Estimation with Missing Data*

Begin with an initial estimate $\hat{f}^{(0)}$ of f, iterate until convergence the following steps for $t = 1, 2, \ldots$:

1. From $\hat{f}^{(t-1)}$ obtain $\hat{\gamma}^{(t-1)}$ through (3).
2. For $m = 1, \ldots, M$,
   - 2a. simulate the missing data $\mathbf{Y}_{\mathrm{mis},m}$ from the distribution $\mathbf{Y}_{\mathrm{mis},m}|\mathbf{Y}_{\mathrm{obs}}, \gamma = \hat{\gamma}^{(t-1)}$;
   - 2b. from the imputed complete data set $\{\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_{\mathrm{mis},m}\}$, calculate the m-th imputed periodogram $I_{\mathrm{imp},m}$, and
   - 2c. apply the complete data spectrum estimation procedure to $I_{\mathrm{imp},m}$ and obtain $\hat{f}_{\mathrm{imp},m}$.
3. Compute the t-th iterative estimate $\hat{f}^{(t)}$ of f as the average of all $\hat{f}_{\mathrm{imp},m}$; i.e.,

$$\hat{f}^{(t)}(\omega) = \sum_{m=1}^{M} \hat{f}_{\mathrm{imp},m}(\omega).$$

Declare convergence when the distance between $\hat{f}^{(t-1)}$ and $\hat{f}^{(t)}$ is sufficiently small.

Notice that a major advantage of the MCSC-Algorithm is its great flexibility, in the sense that in Step 2c one can apply any nonparametric estimator to the imputed periodogram $I_{\mathrm{imp},m}$ to obtain $\hat{f}_{\mathrm{imp},m}$. For examples, one could use wavelet estimators, or spline estimators. In all our numerical work to be reported below, we used the kernel estimator (2) with bandwidth chosen by the method of [6].

One could also further generalize Step 2c to other spectral estimation techniques, such at those based on log-periodogram smoothing, or even to those that impose a parametric form on the spectrum. Even further, one could also easily modify the MCSC-Algorithm to handle higher dimensional problems.

Lastly, we remark that in our implementation the initial estimate $\hat{f}^{(0)}$ is obtained by smoothing the periodogram computed from the Lomb-Scargle method; i.e., by time-domain mean-imputation. Also, details for carrying out the conditional simulation of $\mathbf{Y}_{\mathrm{mis},m}$ in Step 2a can be found in many standard references, such as [3].

4. SIMULATION STUDY

A simulation study was conducted to evaluate the practical performance of the MCSC-Algorithm. Two spectral densities from the ARMA(α, β) model

$$Y_t + a_1 Y_{t-1} + \cdots + a_\alpha Y_{t-\alpha} = e_t + b_1 e_{t-1} + \cdots + b_\beta e_{t-\beta}$$

with e_t as iid $N(0,1)$ were used to simulate $\{Y_t\}$. Spectrum 1 corresponds to an AR(3) process with $(a_1, a_2, a_3) = (0.9, 0.8, 0.6)$, while Spectrum 2 corresponds to a MA(3) process with $(b_1, b_2, b_3) = (0.9, 0.8, 0.6)$. These two spectra have been used by previous authors [9, 15] and are displayed in Figure 1.

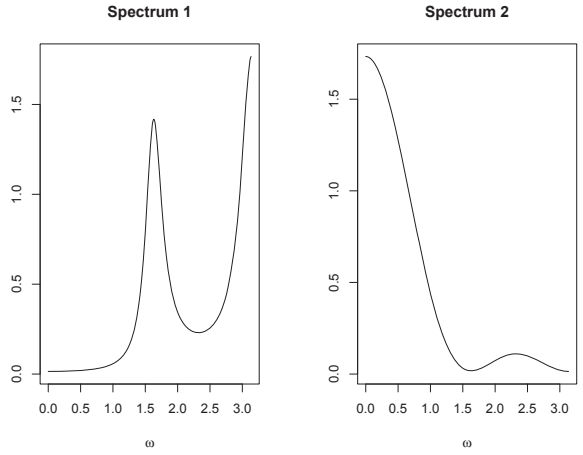

Fig. 1. The two spectral densities used in the simulation study.

For each spectrum, 500 realizations $\{Y_t\}_{t=0}^{2n-1}$ were simulated with $2n = 512$. Then for each realization, missing data were introduced with the following four mechanisms:

1. randomly remove 10% of the observations;
2. remove the middle 10% observations, i.e., leaving a gap of 10% missing values in the middle;
3. randomly remove 30% of the observations; and
4. remove total 30% of the observations, in the form of 3 equi-length gaps, 10% each, starting at $t = n/5, 3n/5$, and n.

Lastly, both the Lomb-Scargle method and the MCSC-Algorithm were applied to each realization with missing data to estimate the corresponding spectrum. Their corresponding mean-squared-errors MSE $= \frac{1}{n}\sum_{j=0}^{n-1}(\hat{f}_j - f_j)^2$ were calculated. For benchmark comparison, we also computed the MSEs of the spectrum estimates obtained by using the complete data. The means of $\sqrt{\mathrm{MSE}}$, together with their standard errors, are listed in Table 1. One could see that the proposed MCSC-Algorithm is capable of achieving a substantial reduction of $\sqrt{\mathrm{MSE}}$, especially for higher missing data percentages.

method	spectrum 1	spectrum 2
10% missing at random		
Lomb-Scargle	0.176 (0.002)	0.180 (0.004)
MCSC	0.162 (0.002)	0.136 (0.004)
complete data	0.156 (0.002)	0.134 (0.004)
10% missing, gap in the middle		
Lomb-Scargle	0.165 (0.002)	0.165 (0.004)
MCSC	0.159 (0.002)	0.154 (0.003)
complete data	0.151 (0.002)	0.133 (0.003)
30% missing at random		
Lomb-Scargle	0.282 (0.002)	0.349 (0.003)
MCSC	0.248 (0.002)	0.147 (0.003)
complete data	0.154 (0.003)	0.132 (0.003)
30% missing, with three gaps of length 10%		
Lomb-Scargle	0.231 (0.002)	0.274 (0.004)
MCSC	0.201 (0.002)	0.228 (0.003)
complete data	0.154 (0.003)	0.136 (0.004)

Table 1. Averages of $\sqrt{\text{MSE}}$ of the estimated spectral density using different methods. Numbers in parentheses are the corresponding standard errors.

5. CONCLUDING REMARKS

In this paper the self-consistency principle is applied to develop a new method for nonparametric spectrum estimation with missing data. The resulting proposal, termed the MCSC-Algorithm, is extremely flexible and performed well in a simulation study. Additional numerical results, together with some theoretical properties of the MCSC-Algorithm, will be reported elsewhere.

References

[1] P. Bloomfield. Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society*, 32:369–380, 1970.

[2] P. Broersen. Automatic spectral analysis with missing data. *Digital Signal Processing*, 16:754–766, 2006.

[3] J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley-Interscience, 1999.

[4] R. Dahlhaus. Nonparametric spectral analysis with missing observations. *Sankhya, Series A*, 3:347–367, 1987.

[5] B. Efron. The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume IV, pages 831–851. Berkeley: University of California Press, 1967.

[6] J. Hannig and T. C. M. Lee. Kernel smoothing of periodograms under Kullback-Leibler discrepancy. *Signal Processing*, 84:1255–1266, 2004.

[7] T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.

[8] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[9] T. C. M. Lee. A stabilized bandwidth selection method for kernel smoothing of the periodogram. *Signal Processing*, 81:419–430, 2001.

[10] T. C. M. Lee and X.-L. Meng. A self-consistent wavelet method for denoising images with missing pixels. In *Proceedings of the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II: 41–44, 2005.

[11] N. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, 1976.

[12] X.-L. Meng and D. B. Rubin. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80:267–278, 1993.

[13] X.-L. Meng and D. A. van Dyk. The EM algorithm – an old folk song sung to a fast new tune (with discussion). *Journal of the Royal Statistical Society Series B*, 59:511–567, 1997.

[14] P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing*, 42:3126–3136, 1994.

[15] H. C. Ombao, J. A. Raz, R. L. Strawderman, and R. von Sachs. A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88:1186–1192, 2001.

[16] J. Scargle. Studies in astronomical time series analysis. II- Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, 263:835–853, 1982.

[17] T. Tarpey. Parallel Principal Axes. *Journal of Multivariate Analysis*, 75:295–313, 2000.

[18] T. Tarpey and B. Flury. Self–Consistency: A fundamental concept in statistics. *Statistical Science*, 11:229–243, 1996.

[19] Y. Wang, P. Stoica, J. Li, and T. Marzetta. Nonparametric spectral analysis with missing data via the EM algorithm. *Digital Signal Processing*, 15:191–206, 2005.