Efficient Non-Parametric Spectral Density Estimation with Randomly Censored Data

Jiaju Wu* Sam Efromovich†

Abstract
Spectral Density estimation is a well known problem for a directly observed time series. The literature on spectral density estimation for a right-censored time series is next to none. A data-driven spectral density estimator for a right censored time series is suggested. This estimator adapts to unknown smoothness of the spectral density and unknown distribution of a censored random variable. Asymptotic upper bound of the mean integrated squared error (MISE) of the proposed estimator is obtained. The estimator is studied via simulated examples.

Key Words: Spectral Density; Time Series; Survival Analysis; Right Censoring; Non-parametric

1. Introduction

Spectral density provides a useful mathematical formulation to describe a time series in frequency domain. Spectral analysis is very useful to identify seasonal components or detect cyclical patterns. Assuming the variable of interest X_t is zero-mean and weekly stationery, the spectral density is denfined on the support $[-\pi, \pi]$ as

$$g(\lambda) := (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j \cos(\lambda j)$$
$$= (2\pi)^{-1}\gamma_0 + \pi^{-1} \sum_{j=1}^{\infty} \gamma_j \cos(\lambda j), \tag{1}$$

where $\gamma_j := \mathbb{E}[X_t X_{t+j}]$ is the auto-covariance function.

There is a vast literature on spectral density estimation when data is fully observable. Most nonparametric procedures use the smoothed periodogram. Early reference on s-moothing periodogram can be traced back to Parzen (1961). Common smoothing techniques can be categorized into two parts. The first approach is to directly smooth the data $(\lambda_j, I(\lambda_j))$ or the log-periodogram $(\lambda_j, \ln I(\lambda_j))$. For example, Wahba (1980) used a s-moothing spline with a smoothing or bandwidth parameter to fit the log-periodogram. The parameter and the degree of smoothing is chosen to minimize the MISE. Brillinger (2001) mentioned in chapter 5 a large selection of earlier references back to the 1940s. The second one is based on the quasi-Likelihood proposed by Whittle (1957)

$$L(g|X_1, ..., X_n) := \prod_{j=1}^{v} \frac{1}{g(\lambda_j)} e^{-U_j/g(\lambda_j)}, \tag{2}$$

where $(U_1, ..., U_v)$ forms the joint distribution and $v = \lfloor (n-1)/2 \rfloor$ due to the symmetry of the periodogram. It is often referred to as the Whittle Likelihood. This likelihood involves

*University of Texas at Dallas, 800 W Campbell Rd, Richardson, TX 75080
†University of Texas at Dallas, 800 W Campbell Rd, Richardson, TX 75080

the spectral density g directly, which is an advantage over using the true likelihood (Choudhuri et al., 2004). Pawitan and O'sullivan (1994) introduced a penalized maximum likelihood estimator (MLE), which maximizing the Whittle likelihood with a roughness penalty term. Fan and Kreutzberger (1998) suggested a local polynomial technique to fit Whittle likelihood. They also compared performance of the Whittle likelihood based approach and the directly smoothing approach, and claimed that the former is better due to smaller MISE. Besides, some bayesian approaches, based on the Whittle likelihood, have been developed and discussed. It can be traced back to Carter and Kohn (1997), who used the Whittle likelihood to obtain a pseudo-posterior distribution of spectral density. Choudhuri et al. (2004) proposed a Bernstein polynomials based prior and obtained a pseudo-posterior distribution by updating the prior through Whittle likelihood, with emphasizes on consistency results for the pseudo-posterior. Cadonna et al. (2017) developed another Bayesian approach based on a local Gaussian mixture approximation to the Whittle likelihood.

However, not much has been done in estimating the spectral density for right censored time series. Censoring creates extra complications in estimating spectral density because the underlying time series is no longer directly available. Furthermore, we devoted to random censoring where censoring effect is controlled by random variables, which is even more challenging. We will need additional procedure to recover the spectral density. The article is organized as follows. Section 2 explains the problem in detail and proposed a universal non-parametric estimator based on series expansion. Section 3 presented an asymptotic upper bound for estimator's MISE. Section 4 provides a simulation study of this new estimator.

2. Methodology

This section will focus on the methodology of non-parametric esitmation of a spectral density for a randomly right censored time series. First, let us describe the setup of our problem. Assume the variable of interest X_t is a stationary time series which is not directly observed. Instead, a pair $\{V_t, \Delta_t\}$ is observed. In this pair, $V_t := \min(X_t, C_t)$ is the observed time series and $\Delta_t := I(X_t \leq C_t)$ is an indicator function. $\Delta_t = 0$ if X_t is right censored and 1 otherwise. The censoring variable C_t is not observed but assumed to be identically independently distributed (i.i.d.) realizations of a random variable C imposing right censoring and independent with X_t. Our goal is to estimate the spectral density of unobserved X_t based on the observed pair $\{V_t, \Delta_t\}$.

2.1 E-estimation

The form of (1) makes it natural to use a series expansion approach. That equation is indeed a Fourier series expansion where $\gamma_j = \int_{-\pi}^{\pi} g(\lambda) \cos(\lambda j) dx$. Due to the form of the spectral density, $\cos(\lambda j)$ is naturally chosen to be the basis function and autocovariance functions γ_j also serve as Fourier coefficients of $g(\lambda)$.

The spectral density is an infinite sum but we can approximate it with a partial sum

$$g_J(\lambda) = (2\pi)^{-1}\gamma_0 + \pi^{-1} \sum_{j=1}^{J} \gamma_j \cos(\lambda j). \tag{3}$$

The main task is to estimate γ_j and choose a cutoff J, and then get an estimate of spectral density $\hat{g}(\lambda)$. The methodology, E-estimation, was proposed by Efromovich (2018) and can be adapted in context of spectral density estimate. It consists of three steps:

1. Use series expansion (1) and suggest a sample mean estimator $\hat{\gamma}_j$. Calculate a sample variance estimator $\hat{v}_{jn}$ of the variance $\mathbb{V}(\hat{\gamma}_j) = v_{jn}$.

2. The E-estimator is

$$\hat{g}(\lambda) = (2\pi)^{-1} \sum_{j=-\hat{J}}^{\hat{J}} \hat{\gamma}_j I(\hat{\gamma}_j^2 > c_{TH}\hat{v}_{jn}) \cos(\lambda j), \tag{4}$$

   where

$$\hat{J} = \text{argmin}_{0 \le J \le J_n} \sum_{j=0}^{J} \left(2\hat{v}_{jn} - \hat{\gamma}_j^2\right). \tag{5}$$

3. $g(\lambda)$ is a non-negative function, so we need to use a non-negative projection (3.1.15) of Efromovich (1999) on (4).

In that 3-step procedure, c_{TH} is a generic constant and J_n is the maximum cutoff considered which needs to be specified later. For almost all functions, some Fourier coefficients are very small or zero, so the indicator function used in (4) is used to filter out small coefficients, or in our context, the sample auto-covariances $\hat{\gamma}_j$. In equation (5), minimizing $\sum_{j=0}^{J}\left(2\hat{v}_{jn} - \hat{\gamma}_j^2\right)$ can be proved equivalent to minimizing the mean integrated squared error (MISE)

$$MISE(\hat{g}, g) := \mathbb{E}\left\{\int_{-\pi}^{\pi} [\hat{g}(\lambda) - g(\lambda)]^2\right\}. \tag{6}$$

Readers are referred to Efromovich (2018) for more thorough and complete explanation of this method.

Step 2 and step 3 are the same for all spectral density estimation. The only difference is in step 1: how to propose a sample mean estimator and derive its variance. This is also the key step in constructing such a non-parametric estimator. For regular time series without missing or censored observations, we can use the familiar sample auto-covariance $\hat{\gamma}_j = (n-j)^{-1}\sum_{l=1}^{n-j} X_l X_{l+j}$. With censored time series, however, we need some additional modifications. Under our setup of censoring, we suggest the following sample mean estimator

$$\hat{\gamma}_j^X = (n-j)^{-1} \sum_{l=1}^{n-j} \frac{V_l V_{l+j} \Delta_l \Delta_{l+j}}{G^C(V_l) G^C(V_{l+j})}. \tag{7}$$

If the survival function of censoring G^C is unknown, we can estimate it by

$$\hat{G}^C(v) = \exp(-\hat{H}^C(v)), \tag{8}$$

where $\hat{H}^C(v)$ is estimated cumulative hazard

$$\hat{H}^C(v) = n^{-1} \sum_{i=1}^{n} \frac{(1-\Delta_i) I(V_i \le v)}{\hat{G}^{V_l}(V_i)}, \tag{9}$$

and $\hat{G}^{V_l}(v) = n^{-1}\sum_{s=1}^{n} I(V_s \ge v)$. With equation (7), (8), and (9), we expressed $\hat{\gamma}_j^X$ only using information from the observed time series data.

After auto-covariance, we need to suggest a data-driven cutoff. Following Efromovich (2014), we set the maximum cutoff $J_n = \lceil b_n^{-1} \ln(n) \rceil$ where $b_n = 1/\ln(\ln(n))$ and introduce a statistic

$$\hat{F}(J) := \pi^{-1} \sum_{j=J+1}^{J_n} (\hat{\gamma}_j^2 - \hat{\gamma}_{j+J_n}^2), \tag{10}$$

which is a sufficiently accurate estimate of $F(J) := \pi^{-1} \sum_{j=J+1}^{J_n} \gamma_j^2$, and it is asymptotically equal to the integrated squared bias of the spectral density estimate. Next, we set the data-driven cutoff $\hat{J}$ to be the smallest integer satisfying $b_n \ln(n) \leq \hat{J} \leq J_n$ and $\hat{F}(\hat{J}) < b_n \ln(n) n^{-1}$.

3. Asymptotic Upper Bound

Before presenting asymptotic upper bound of MISE, the mixing theory is worth mentioning. This article deals with weekly dependent time series and many well known results for independent random variables do not hold anymore. Mixing theorem can help us explore the dependence properties. On this theorem, there is a vast literature which dates back to Rosenblatt (1956), who introduced the strong mixing coefficient or α-mixing coefficient. Using the notations from Dedecker et al. (2007) and Merlevède et al. (2009), for any two σ-algebra $\mathcal{F}$ and $\mathcal{G}$, the α-mixing coefficient is defined as

$$\alpha(\mathcal{F}, \mathcal{G}) := \sup_{F \in \mathcal{F}, G \in \mathcal{G}} |\, \mathbb{P}(F \cap G) - \mathbb{P}(F)\, \mathbb{P}(G)|. \tag{11}$$

Let $\{X_t, t \geq 1\}$ be a stationary time series and we can write the strong mixing coefficient corresponding to X_t as

$$\alpha^X(s) := \sup_{k \geq 1} \alpha(\sigma(X_t, t \leq k), \sigma(X_t, t \geq k+s)). \tag{12}$$

X_t is called strongly mixing or α-mixing if $\alpha^X(s) \to 0$ as $s \to \infty$. Besides Rosenblatt's α-mixing, a great variety of other mixing coefficients were proposed. The monograph by Dedecker et al. (2007) provides a huge collection of these mixing coefficients, as well as a rich amount of theoretical results and examples. Some mixing theorems in those references can be used to derive the upper bound.

Now we can prsent the results on estimation of the spectral density for randomly censored time series.

Assumption 1. *$\{X_t\}$ is Gaussian zero-mean and strictly stationary time series, and $\mathbb{E}[X_t^k] < \infty$ for $k = 8$.*

Assumption 2. $\mathbb{E}\left|\frac{X_1^8}{[G^C(X_1)]^7}\right|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Theorem 1. *Under assumption 1 and assumption 2, the spectral density estimate*

$$\hat{g}^X(\lambda, \hat{J}) := (2\pi)^{-1} \sum_{j=-\hat{J}}^{\hat{J}} \hat{\gamma}_j^X \cos(\lambda j) \tag{13}$$

has the following upper bound for its MISE:

$$MISE(\hat{g}^X(\lambda, \hat{J}), g^X(\lambda)) \leq (2\pi nr)^{-1} \ln(n) d^*(g^X)(1 + o_n(1)), \tag{14}$$

where

$$d^*(g^X) := \left\{ \mathbb{E}\left[\frac{V^2 \Delta}{(G^C(V))^2} \right] \right\}^2 + 2 \sum_{j=1}^{\infty} (\gamma_j^X)^2 \tag{15}$$

4. Simulation

This section carries out a simulation study of our proposed estimator. We will consider the following experiment. Sample sizes n are set to be 100, 300, and 500. The underlying time series X_t is ARMA(1,1) with MA to be 0.4 and AR to be 0.3, 0.5, and 0.7. Then we simulated the censoring variable C from N(2.1,1), N(0.8,1), and N(0,1) to impose light (6%), medium (25%) and heavy (50%) right censoring, respectively.

For each combination of model and sample size, we obtain 1000 replications of observed pairs $\{V_t, \Delta_t\}$ and try to recover the real spectral density. Then we compare our estimator with Naive and Oracle estimators. Naive estimator treats V_t as X_t, ignoring the censoring effect. Oracle estimator sees the hidden X_t and use X_t to estimate real spectral density which also serves as the best achivable one for our estimator. We presented here two graphs as examples.

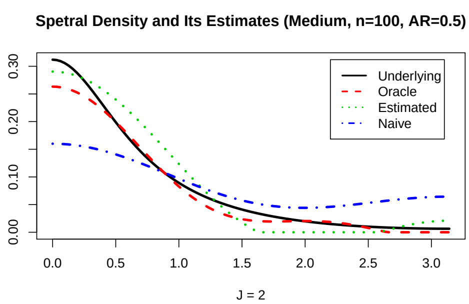

Figure 1

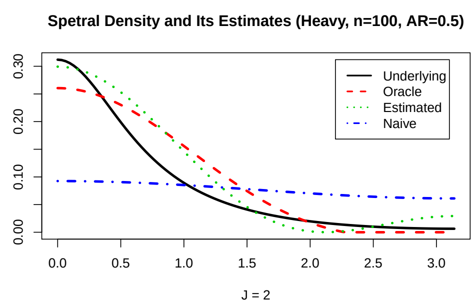

Figure 2

Both figures are generated with AR=0.5, MA=0.4, n=100. Figure 1 is under medium censoring and figure 2 is under heavy censoring. I do not show the case of light censoring since all three estimators perform almost the same. Under medium censoring, Naive estimator (blue) cannot keep up with the performance of our E-estimator (green). The curve of E-estimator is still smooth and close to the underlying spectral density. Even if we increase the censoring level to an average of 50%, the situation remains the same. The Oracle (red) and E-estimator are sufficiently accurate while the Naive one is almost a horizontal line.

Next we simulated data with random ARMA coefficients. We increased the replications of simulation to 2000. AR and MA coefficients will be chosen randomly in each replication. Three models were considered in our experiment: ARMA(1,1) under Normal censoring, ARMA(1,1) under Laplace censoring, and ARMA(3,3) under Normal censoring. This article takes the first model to illustrate. For ARMA(1,1) model, MA and AR coefficients are randomly chosen from the set (0.2, 0.4, 0.6, 0.8). MISEs are generated for different estimators to make comparisons.

Table 1 shows MISEs of different estimators under various scenarios with the exception of last line. The last line is the ratio of MISEs of E-estimator and Naive estimator.

	n=100			n=300			n=500		
	Light	Medium	Heavy	Light	Medium	Heavy	Light	Medium	Heavy
E-Est	0.0289	0.0385	0.0448	0.0164	0.0236	0.0326	0.0136	0.0206	0.0261
Naive	0.0326	0.0471	0.0650	0.0157	0.0334	0.0545	0.0139	0.0330	0.0502
Oracle	0.0278	0.0261	0.0261	0.0128	0.0123	0.0129	0.0105	0.0104	0.0087
E-Est/Nai	0.8974	0.8176	0.6891	1.0448	0.7083	0.5979	0.9746	0.6245	0.5203

Table 1: MISE, ARMA(1,1), Normal Censoring, Random AR and MA

Almost all such ratios are less than 1, meaning the E-estimator has smaller MISE and thus better than the Naive one. Thus, our estimator is superior to the Naive for ARMA(1,1) with Normal censoring. Besides, we also considered and tested other two models which was mentioned above. All simulation results confirm the conclusion and show that our estimator adapts to unknown smoothness of the spectral density and unknown distribution of a censored random variable.

References

Brillinger, D. R. (2001). *Time series: data analysis and theory*. SIAM.

Cadonna, A., Kottas, A., and Prado, R. (2017). Bayesian mixture modeling for spectral density estimation. *Statistics & Probability Letters*, 125:189–195.

Carter, C. K. and Kohn, R. (1997). Semiparametric bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):255–268.

Choudhuri, N., Ghosal, S., and Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059.

Dedecker, J., Doukhan, P., Lang, G., Rafael, L. R. J., Louhichi, S., and Prieur, C. (2007). Weak dependence. In *Weak dependence: With examples and applications*, pages 9–20. Springer.

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory, and applications*. New York: Springer.

Efromovich, S. (2014). Efficient non-parametric estimation of the spectral density in the presence of missing observations. *Journal of Time Series Analysis*, 35(5):407–427.

Efromovich, S. (2018). *Missing and Modified Data in Nonparametric Estimation: With R Examples*. Chapman and Hall/CRC.

Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scandinavian Journal of Statistics*, 25(2):359–369.

Merlevède, F., Peligrad, M., Rio, E., et al. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, pages 273–292. Institute of Mathematical Statistics.

Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics*, 3(2):167–190.

Pawitan, Y. and O'sullivan, F. (1994). Nonparametric spectral density estimation using penalized whittle likelihood. *Journal of the American Statistical Association*, 89(426):600–610.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75(369):122–132.

Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 19(1):38–47.