Statistics and Probability Letters 94 (2014) 204–213

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Aggregation of spectral density estimators

Christopher Chang, Dimitris Politis *

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA

ARTICLE INFO

Article history:
Received 13 December 2013
Received in revised form 15 July 2014
Accepted 15 July 2014
Available online 22 July 2014

Keywords:
Estimation
Spectral analysis
Time series analysis

ABSTRACT

Given stationary time series data, we study the problem of finding the best linear combination of a set of lag window spectral density estimators with respect to the mean squared risk. We present an aggregation procedure and prove a sharp oracle inequality for its risk. We also provide simulations demonstrating the performance of our aggregation procedure, given Bartlett and other estimators of varying bandwidths as input. This extends work by P. Rigollet and A. Tsybakov on aggregation of density estimators.

0. Introduction

Consider stationary time series data $X_1, \ldots, X_n$ having mean zero and spectral density

$$p(\lambda) := \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-i\lambda j}, \quad \lambda \in [-\pi, \pi) \tag{1}$$

where $\gamma(k)$ is the autocovariance at lag k. For an estimator $\hat{p}(X_1, \ldots, X_n)$ of p, define the L_2-risk

$$R_n(\hat{p}, p) = E\left[\int_{-\pi}^{\pi} (\hat{p}(x) - p(x))^2 \, dx\right]. \tag{2}$$

Let $\hat{p}_1, \ldots, \hat{p}_J$ be a collection of lag window (i.e. kernel) spectral density estimators of p—see Eq. (5) for a precise definition. We investigate the construction of a new estimator $\hat{p}_n^L$ which is asymptotically as good, in terms of L_2-risk, as using the best possible linear combination of $\hat{p}_1, \ldots, \hat{p}_J$; more precisely, $\hat{p}_n^L$ satisfies the oracle inequality

$$R_n(\hat{p}_n^L, p) \leq \inf_{\lambda \in \mathbb{R}^J} R_n\left(\sum_{j=1}^{J} \lambda_j \hat{p}_j, p\right) + \Delta_{n,J} \tag{3}$$

where $\Delta_{n,J}$ is a small remainder term independent of p.

Such an estimator may find a variety of applications. For instance, one may try to bypass the difficult issue of bandwidth (or model) selection by setting the $\hat{p}$s to cover a wide spread of possibly reasonable bandwidths/models. Alternatively, when a linear combination of kernels outperforms all the individual inputs, e.g. when the $\hat{p}$s are Bartlett windows as in Politis and Romano (1995), our aggregating estimator is capable of discovering it.

* Corresponding author.
E-mail addresses: chrchang@alumni.caltech.edu (C. Chang), politis@math.ucsd.edu, dpolitis@ucsd.edu (D. Politis).

http://dx.doi.org/10.1016/j.spl.2014.07.017
0167-7152/

Kernel density estimation dates back to Rosenblatt (1956) and Parzen (1962); Priestley (1981) and Brillinger (1981) discuss its application to spectral densities. More recently, Yang (2000) and Rigollet and Tsybakov (2007) analyzed aggregation of probability density estimators, while Wang et al. (2011) studied the related problem of linear aggregation in nonparametric regression. We extend Rigollet and Tsybakov (2007)'s work to spectral estimation.

To perform aggregation, we use a sample splitting scheme. The time series data is divided into a training set, a buffer zone, and a validation set; with an exponential mixing rate, the buffer zone need not be more than logarithmic in the size of the other sets to ensure approximate independence between the training and validation sets.

The estimator, and theoretical results concerning its performance, are presented in Section 1. Simulation studies are conducted in Section 2, and our conclusions are stated in Section 3.

1. Theoretical results

1.1. Aggregation procedure

Split the time series $X_1, \ldots, X_n$ into a training set $X_1, \ldots, X_{n_t}$, a buffer zone $X_{n_t+1}, \ldots, X_{n_t+n_b}$, and a validation set $X_{n_t+n_b+1}, \ldots, X_{n_t+n_b+n_v}$, where the first and third sets can be treated as independent. We investigate appropriate choices of n_t, n_b, and n_v at the end of this section.

With the training set, we produce an initial estimate

$$\hat{\gamma}_1(k) := \frac{1}{n_t} \sum_{j=1}^{n_t-k} X_{j+k} X_j \tag{4}$$

of the autocovariance function, after centering the data. (In practice, the data will be centered to the sample mean rather than the true mean, but the resulting discrepancy is asymptotically negligible w.r.t. autocovariance and spectral density estimation. So, for simplicity of presentation, we center at the true mean above.)

We then propose the following candidate estimators:

$$p_j(\lambda) := \frac{1}{\sqrt{2\pi}} \sum_{k=-b_j}^{b_j} \hat{\gamma}_1(k) \cdot w_j\left(\frac{k}{b_j}\right) \frac{e^{ik\lambda}}{\sqrt{2\pi}} \tag{5}$$

where the b_js ($j = 1, \ldots, J$) are candidate bandwidths arrived at via some selection procedure, and the w_js ($j = 1, \ldots, J$) are lag windows with $w_j(0) = 1$, $w_j(x) \le 1$ for $x \in (-1, 1)$, and $w_j(x) = 0$ for $|x| \ge 1$ for all j. The p_js have some linear span $\mathcal{L}$ in L_2 whose dimension is denoted by M where $M \le J$. Now construct an orthonormal basis $\{\phi_j\}$ ($j = 1, \ldots, M$), and note that the ϕ_js are – by necessity – trigonometric polynomials of degree at most $b := \max_j b_j$, i.e.,

$$\phi_j(\lambda) = \sum_{k=-b}^{b} a_{j,k} \frac{e^{ik\lambda}}{\sqrt{2\pi}} \tag{6}$$

so the coefficient $a_{j,k}$ is the inner product of ϕ_j and $\frac{e^{ik\lambda}}{\sqrt{2\pi}}$ in L_2.

Then, based on our validation set, we produce a different estimate of the autocovariance function, namely

$$\hat{\gamma}_2(k) := \frac{1}{n_v} \sum_{j=1}^{n_v-k} X_{n_t+n_b+j+k} X_{n_t+n_b+j} \tag{7}$$

and compute the coefficients

$$\hat{K}_j := \frac{1}{\sqrt{2\pi}} \sum_{k=-b}^{b} \hat{\gamma}_2(k) a_{j,k}. \tag{8}$$

Finally, our proposed aggregate estimator of the spectral density is given by

$$\hat{p}(\lambda) := \sum_{j}^{M} \hat{K}_j \phi_j(\lambda). \tag{9}$$

1.2. Performance bounds

We start with the simplest mixing assumption, m-dependence (i.e., that for all positive integers j and k where $k \ge m$, X_j and X_{j+k} are independent).