Statistical Methodology 27 (2015) 12–19

Contents lists available at ScienceDirect

Statistical Methodology

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

# Multivariate wavelet-based density estimation with size-biased data

Esmaeil Shirazi [a,*], Hassan Doosti [b]

[a] Department of Statistics, Faculty of Science, Gonbad Kavous University, Gonbad Kavous 4971799151, Iran

[b] Department of Biostatistics and Epidemiology, Mashhad University of Medical Sciences, Mashhad, Iran

ARTICLE INFO

*Article history:*
Received 14 May 2014
Received in revised form
1 May 2015
Accepted 7 May 2015
Available online 19 May 2015

*MSC:*
primary 62G07
secondary 62G20

*Keywords:*
Besov spaces
Density estimation
Size-biased data
Wavelet

ABSTRACT

In this paper, we employ wavelet method to propose a multivariate density estimator based on a biased sample. We investigate the asymptotic rate of convergence of the proposed estimator over a large class of densities in the Besov space, $B^s_{pq}$. Moreover, we prove the consistency of our estimator when the expectation of *weight function* is unknown. This paper is an extension of results in Ramirez and Vidakovic (2010) and Chesneau et al. (2012) to the multivariate case.



## 1. Introduction

Weighted distribution is widely employed in variety of disciplines such as reliability, biometry, survival analysis, forestry, ecology and wildlife studies. We may equally refer to the survey in [19] on several practical examples of weighted distributions.

Suppose that $\widetilde{X}$ is a $d$-dimensional nonnegative random vector with distribution function (df) $F$ and probability density function (pdf) $f$. Usually $\widetilde{X} = (X_1, \ldots, X_d)$ is not observable but we observe a biased sample from $\widetilde{Y} = (Y_1, \ldots, Y_d)$ with df $G$ and pdf $g$ related to $f$ as follows:

$$g(\widetilde{y}) = w(\widetilde{y}) f(\widetilde{y})/\mu,$$

* Corresponding author.
*E-mail addresses:* shirazi.esmaeel@gmail.com (E. Shirazi), h.doosti@tmu.ac.ir (H. Doosti).

http://dx.doi.org/10.1016/j.stamet.2015.05.002
1572-3127/

where $w(\widetilde{y})$ is the weight function, $w(\widetilde{y}) \geq 0$ and $\mu = \mathbb{E}(w(\widetilde{X})) < \infty$. As $f(\widetilde{x})$ is unknown, the parameter $\mu$ is also unknown. The problem is indirect since one observes $\widetilde{Y}$ and wants to estimate the density of an unobserved $\widetilde{X}$.

Currently, the most popular nonparametric weighted distribution estimation is kernel density estimation. In univariate case we may refer to Jones [15] and Wu and Mao [25]. Barmi and Simonoff [4] propose a simple transformation-based approach to estimate $f$. Vardi [23] and Alvarez and Rodriguez [2] suggest the nonparametric maximum likelihood estimators. The orthogonal series estimators discussed by Efromovich [10,11] based on Fourier series. Patil et al. [20], Arnold and Nagaraja [3], Ahmad [1] and Jain and Nanda [14] extend the previous results to the multivariate case.

Recently, Chesneau [7], Ramirez and Vidakovic [21], Chesneau et al. [8] and Chaubey and Shirazi [6] considered an estimator of the density function based on wavelets for a sample from weighted distributions and studied their properties in the univariate case.

In this paper, we generalize the results of Ramirez and Vidakovic [21] and Chesneau et al. [8] to the multivariate case. There are many situations where the multivariate weighted density estimation is required. For example, Jain and Nanda [14] considered multivariate weighted density estimation in a reliability system which is a generalization of the example given in [16] for bivariate distribution.

We propose a linear wavelet density estimator and study its asymptotic convergence rate over a large class of densities in the Besov space $B^s_{pq}$. We find an upper bound on $L_2$-loss for introduced estimator.

The organization of the paper is as follows. In Section 2, we list basic properties of wavelets and multiresolution analysis. Besov spaces on $\mathbb{R}^d$ are presented in this section. The proposed estimator and our main results are discussed in Section 3 while the proofs are postponed to Section 4.

## 2. Model and estimators

This section contains a brief preliminary about wavelets which will be used in the sequel. The main part of this section is adopted from [17] and references therein. By using $d$ univariate orthogonal multiresolution analysis,

$$\cdots V_{-2,(i)} \subset V_{-1,(i)} \subset V_{0,(i)} \subset V_{1,(i)} \subset V_{2,(i)} \cdots \subset L_2(\mathbb{R}), \quad i = 1, 2, \ldots, d,$$

one can define $d$-dimensional multiresolution analysis

$$\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots \subset L_2(\mathbb{R}^d)$$

such that

$$V_j = \bigotimes_{i=1}^{d} V_{j,(i)} \subset L_2(\mathbb{R}^d).$$

There exists a scale function $\phi \in L_2(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} \phi(x)dx = 1$ such that for any integer $j$ and any $k \in \mathbb{Z}^d$, $\phi_{jk}(x) = 2^{jd/2}\phi(2^j x - k)$ is the elements of orthonormal basis for $V_j$.

**Definition 2.1.** The multiresolution analysis is said $r$-regular if $\phi \in \mathbf{C}^r$ and all its partial derivatives up to total order $r$ are bounded continuous and rapidly decreasing that is for any integer $m \geq 1$ there exists a constant $c_m$ such that

$$\left|(D^\beta \phi)(x)\right| \leq \frac{c_m}{(1 + \|x\|)^m} \tag{2.1}$$

for all $|\beta| \leq r$, where

$$(D^\beta \phi)(x) = \frac{\partial \phi(x)}{\partial x_1^{\beta_1}, \ldots, \partial x_d^{\beta_d}} \tag{2.2}$$