Journal of the Korean Statistical Society 45 (2016) 51–63

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Multiplicative bias correction for generalized Birnbaum–Saunders kernel density estimators and application to nonnegative heavy tailed data

Nabil Zougab [a,b,*], Smail Adjabi [b]

[a] University of Tizi-Ouzou, Algeria

[b] LAMOS, Laboratory of Modeling and Optimization of Systems, University of Béjaïa, Algeria

ARTICLE INFO

Article history:
Received 18 December 2014
Accepted 1 July 2015
Available online 17 July 2015

AMS 2000 subject classifications:
62G05
62G07

Keywords:
Bandwidth
Multiplicative bias correction
Generalized Birnbaum–Saunders kernel
Heavy tailed data

ABSTRACT

In this paper, we show that the multiplicative bias correction (MBC) techniques can be applied for generalized Birnbaum–Saunders (GBS) kernel density estimators. First, some properties of the MBC-GBS kernel density estimators (bias, variance and mean integrated squared error) are shown. Second, the choice of bandwidth is investigated by adopting the popular cross-validation technique. Finally, the performances of the MBC estimators based on GBS kernels are illustrated by a simulation study, followed by a real application for nonnegative heavy tailed (HT) data. In general, in terms of integrated squared bias (ISB) and integrated squared error (ISE), the proposed estimators outperform the standard GBS kernel estimators.

Contents

1. Introduction 52
2. A short review on GBS kernels 52
   2.1. GBS distribution 52
   2.2. GBS kernel estimator 53
3. MBC for GBS density estimators 54
   3.1. Estimators 54
   3.2. Asymptotic properties 54
   3.3. Global property 55
   3.4. Choice of bandwidth for MBC-GBS kernel estimators 56
4. Simulation study 56
5. Application to air pollution data 57
6. Proofs 58
   6.1. Sketch of the proof of Theorem 1 59
      6.1.1. Bias 59
      6.1.2. Variance 61

* Correspondence to: LAMOS, route de Targa-Ouzemmour, 06000 Béjaïa, Algeria.
E-mail addresses: nabilzougab@yahoo.fr (N. Zougab), adjabi@hotmail.com (S. Adjabi).

http://dx.doi.org/10.1016/j.jkss.2015.07.001

6.2. Sketch of the proof of Theorem 2 61
6.2.1. Bias........ 61
6.2.2. Variance........ 62
7. Conclusion 62
Acknowledgments 62
References........ 62

1. Introduction

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) continuous random variables with an unknown probability density function (pdf) f on the support $\mathbb{T}$ ($\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = [0, \infty)$). A continuous symmetric or asymmetric kernel estimator $\hat{f}_h(x)$ of $f(x)$ can be defined as follows:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i) \quad (1)$$

where $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters (bandwidths) and $K_{x,h}$ is the continuous symmetric or asymmetric kernel with the target x and the bandwidth h. Note that in symmetric case, we have $K_{x,h}(\cdot) = (1/h)K\{(x-\cdot)/h\}$, where $K(\cdot)$ is the kernel function which is a symmetric pdf independent of x and h; see, e.g., Parzen (1962), Rosenblatt (1956) and Silverman (1986). It is well known that the symmetric kernel estimator is inappropriate for estimating densities with support $\mathbb{T} = [0, \infty)$, because it causes boundary bias. Thus, the asymmetric kernels have been proposed as a good solution for avoiding these boundary effects. This simple idea is due to Chen (2000) (gamma and modified gamma kernels), Scaillet (2004) (inverse and reciprocal inverse Gaussian kernels), Jin and Kawczak (2003) (log-normal and Birnbaum–Saunders (BS) kernels) and Marchant, Bertin, Leiva, and Saulo (2013) (generalized Birnbaum–Saunders (GBS) kernels); see also Chen (1999) when the support is $\mathbb{T} = [0, 1]$ and Kokonendji and Senga Kiessé (2011) for discrete case ($\mathbb{T} = \mathbb{N}$).

Assuming that the true density f is twice continuously differentiable, then the bias of (1) is $O(h^2)$ with symmetric kernels and $O(h)$ with asymmetric kernels as $h \to 0$. Jones, Linton, and Nielsen (1995) and Terrell and Scott (1980) proposed the so-called multiplicative bias correction (MBC) techniques, which improve bias from $O(h^2)$ as the bandwidth $h \to 0$, to $O(h^4)$ for kernel density using symmetric second-order kernels, see also Jones and Foster (1993) for the same context. Recently, Hirukawa (2010) and Hirukawa and Sakudo (2014) have demonstrated that these two classes of MBC approaches can be applied to kernel density estimation on the unit interval using beta and modified beta kernels and for density estimation using asymmetric kernels (gamma, modified gamma, inverse Gaussian, reciprocal inverse Gaussian, log-normal and BS kernels), respectively. The authors have shown that the order of magnitude in bias is improved from $O(h)$ to $O(h^2)$.

The main goal of this paper is to extend the application of MBC approaches for GBS kernel density estimation as in Hirukawa (2010) and Hirukawa and Sakudo (2014). These previous studies are motivations of this paper. Our study is also motivated by several points. First, the family of GBS kernels introduced recently by Marchant et al. (2013) has a large number of particular cases such as the BS-classical, BS-power-exponential (BS-PE) and BS-Student-t (BS-t) kernels. Second, the GBS kernels are more appropriate for estimating densities of nonnegative HT data, because of their flexibility and properties; see Jin and Kawczak (2003) and Marchant et al. (2013). As third motivation, some applications of GBS kernel methods can be found in various domains such as in economics, finance, reliability, actuarial and also environmental sciences.

This paper is organized as follows. Section 2 briefly recalls the GBS distribution and standard GBS kernel density estimators. In Section 3 we develop asymptotic properties of MBC-GBS kernel density estimators and adopt the unbiased-cross validation (UCV) procedure of choosing the bandwidth for proposed estimators. Section 4 conducts Monte Carlo simulations to compare sample finite performance of standard GBS and proposed MBC-GBS kernel estimators. Section 5 provide an application on real environmental data and all proofs are given in Section 6. Finally, Section 7 concludes the paper.

2. A short review on GBS kernels

In this section, we present a brief recall on GBS distribution and GBS kernel density estimators.

2.1. GBS distribution

Consider a GBS random variable $T \sim GBS(\alpha, \beta; g)$, where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter and g is a real function that generates the density of random variable $Z = \left(\sqrt{T/\beta} - \sqrt{\beta/T}\right)/\alpha$ with standard symmetric distributions; see Marchant et al. (2013) for more details. The probability density function (pdf) of T is given by

$$f_T(t; \alpha, \beta; g) = cg\left(\frac{1}{\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right) \frac{1}{2\alpha}\left(\frac{1}{\sqrt{\beta t}} + \sqrt{\frac{\beta}{t^3}}\right), \quad t > 0; \alpha > 0,\ \beta > 0, \quad (2)$$