ARTICLE IN PRESS
Journal of the Korean Statistical Society ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect
Journal of the Korean Statistical Society
journal homepage: www.elsevier.com/locate/jkss

Central limit theorem for the variable bandwidth kernel density estimators

Janet Nakarmi [a], Hailin Sang [b,*]

[a] Department of Mathematics, University of Central Arkansas, Conway, AR 72035, USA
[b] Department of Mathematics, The University of Mississippi, University, MS 38677, USA

ARTICLE INFO

Article history:
Received 12 September 2016
Accepted 23 January 2018
Available online xxxx

AMS 2000 subject classifications:
primary 62G07
secondary 62E20

Keywords:
Central limit theorem
Variable bandwidth kernel density estimation

ABSTRACT

In this paper we study the ideal variable bandwidth kernel density estimator introduced by McKay (1993a, b) and Jones et al. (1994) and the plug-in practical version of the variable bandwidth kernel estimator with two sequences of bandwidths as in Giné and Sang (2013). Based on the bias and variance analysis of the ideal and plug-in variable bandwidth kernel density estimators, we study the central limit theorems for each of them. The simulation study confirms the central limit theorem and demonstrates the advantage of the plug-in variable bandwidth kernel method over the classical kernel method.

1. Introduction

Suppose that X_i, $i \in \mathbb{N}$, are independent identically distributed (i.i.d.) observations with density function $f(t)$, $t \in \mathbb{R}^d$. Let K to be a symmetric probability kernel satisfying some differentiability properties. The classical kernel density estimator

$$\hat{f}(t; h_n) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n}\right), \tag{1}$$

where h_n is the bandwidth sequence with $h_n \to 0$, $nh_n^d \to \infty$, and its properties have been well studied in the literature. The variance of (1) has order $O((nh_n^d)^{-1})$ and the bias has order $O(h_n^2)$ if $f(t)$ has bounded second order partial derivatives. See Silverman (1986) and Wand and Jones (1995) for the literature on kernel density estimation. For $k = (k_1, \ldots, k_d) \in (\mathbb{N} \cup \{0\})^d$, set $|k| = \sum_{i=1}^{d} k_i$, one may obtain bias with order $O(h_n^4)$ for the estimator (1) if the fourth order kernel function $K(x)$ is allowed: $\int_{\mathbb{R}^d} K(x)dx = 1$ and $\int_{\mathbb{R}^d} x_1^{k_1} \cdots x_d^{k_d} K(x)dx = 0$ for $|k| = 1, 2, 3$. Nevertheless, $\hat{f}(t; h_n)$ in (1) may take negative values and therefore not a true density function in this case since $K(x)$ may take negative values. For example, see Marron (1994). In this paper we study the following multi-dimensional version of the variable bandwidth kernel density estimator proposed by McKay (1993a, b):

$$\bar{f}(t; h_n) = \frac{1}{nh_n^d} \sum_{i=1}^{n} \alpha^d(f(X_i))K(h_n^{-1}\alpha(f(X_i))(t - X_i)), \tag{2}$$

* Corresponding author.
E-mail addresses: janetn@uca.edu (J. Nakarmi), sang@olemiss.edu (H. Sang).

https://doi.org/10.1016/j.jkss.2018.01.001
1226-3192/

Please cite this article in press as: Nakarmi, J., Sang, H., Central limit theorem for the variable bandwidth kernel density estimators. Journal of the Korean Statistical Society (2018), https://doi.org/10.1016/j.jkss.2018.01.001.

where $\alpha(s)$ is a smooth function of the form

$$\alpha(s) := cp^{1/2}(s/c^2). \tag{3}$$

The function p has at least fourth order derivative and satisfies $p(x) \geq 1$ for all x and $p(x) = x$ for all $x \geq t_0$ for some $1 \leq t_0 < \infty$, and a fixed number c, where $0 < c < \infty$. Eq. (2) is a variable bandwidth kernel density estimator since the bandwidth has form $h_n/\alpha(f(X_i))$ if we rewrite (2) in the form of the classical one, (1).

The study of variable bandwidth kernel density estimation goes back to Abramson (1982). Abramson proposed the following estimator

$$f_A(t; h_n) = \frac{1}{nh_n^d}\sum_{i=1}^{n}\gamma^d(t, X_i)K(h_n^{-1}\gamma(t, X_i)(t - X_i)), \tag{4}$$

where $\gamma(t, s) = (f(s) \vee f(t)/10)^{1/2}$. The bandwidth $h_n/\gamma(t, X_i)$ at each observation X_i is inversely proportional to $f^{1/2}(X_i)$ if $f(X_i) \geq f(t)/10$. Notice that (2) also has the square root law since $\alpha(f(X_i)) = f^{1/2}(X_i)$ if $f(X_i) \geq t_0c^2$ by the definition of the function $p(x)$. The estimator (2) or (4) has clipping procedure in (3) or $\gamma(t, s)$ since they make the true bandwidth $h_n/\alpha(f(X_i)) \geq h_n/c$ or $h_n/\gamma(t, X_i) \geq 10^{1/2}h_n/f(t)^{1/2}$. The clipping procedures prevent too much contribution to the density estimation at t if the observation X_i is too far away from t. Abramson showed that this square root law and the clipping procedure improve the bias from the order of h_n^2 to the order of h_n^4 for the estimator (4) while at the same time keep the variance at the order of $(nh_n^d)^{-1}$ if $f(t) \neq 0$ and $f(x)$ has fourth order continuous derivatives at t. So, one has a *non-negative* estimator of the density that performs asymptotically as a kernel estimator based on a fourth order (hence, partly negative) kernel. However, this variable bandwidth estimator (4) is not a density function of a true probability measure since the integral of $f_A(t; h_n)$ over t is not 1.

McKay (1993b) and Terrell and Scott (1992) showed that the following modification of the Abramson estimator without the 'clipping filter' $(f(t)/10)^{1/2}$ on $f^{1/2}(X_i)$ studied in Hall and Marron (1988),

$$f_{HM}(t; h_n) = \frac{1}{nh_n^d}\sum_{i=1}^{n}f^{d/2}(X_i)K(h_n^{-1}f^{1/2}(X_i)(t - X_i)), \tag{5}$$

which has integral 1 and thus is a true probability density, may have bias of order much larger than h_n^4. Therefore, the clipping is necessary for such bias reduction. In the case $d = 1$, Hall, Hu, and Marron (1995) proposed the estimator

$$f_{HHM}(t; h_n) = \frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{t - X_i}{h_n}f^{1/2}(X_i)\right)f^{1/2}(X_i)I(|t - X_i| < h_nB) \tag{6}$$

where B is a fixed constant; see also Novak (1999) for a similar estimator. This estimator is non-negative and achieves the desired bias reduction but, like Abramson's, it does not integrate to 1.

In conclusion, it seems that the estimator (2) has all the advantages: it is a true density function with square root law and smooth clipping procedure. However, notice that this estimator and all the other variable bandwidth kernel density estimators are not applicable in practice since they all include the studied density function f. Therefore, we call them ideal estimators in the literature. Hall and Marron (1988) studied a true density estimator

$$\hat{f}_{HM}(t; h_{1,n}, h_{2,n}) = \frac{1}{nh_{2,n}^d}\sum_{i=1}^{n}K\left(\frac{t - X_i}{h_{2,n}}\hat{f}^{1/2}(X_i; h_{1,n})\right)\hat{f}^{d/2}(X_i; h_{1,n}),$$

by plugging in a pilot estimator, the classical estimator (1), into the estimator (5). Here the bandwidth sequence $h_{2,n}$ is the h_n as in (5) and the bandwidth sequence $h_{1,n}$ is applied in the classical kernel density estimator (1), i.e.,

$$\hat{f}(t; h_{1,n}) = \frac{1}{nh_{1,n}^d}\sum_{i=1}^{n}K\left(\frac{t - X_i}{h_{1,n}}\right).$$

They took the Taylor expansion of $K\left(\frac{t-X_i}{h_{2,n}}\hat{f}^{1/2}(X_i; h_{1,n})\right)$ at $K\left(\frac{t-X_i}{h_{2,n}}f^{1/2}(X_i)\right)$ and then proved that the discrepancy between the plug-in estimator $\hat{f}_{HM}(t; h_{1,n}, h_{2,n})$ and the ideal version (5) has asymptotic convergence rate $O_P(n^{-4/(8+d)})$ pointwise. By applying this Taylor decomposition, McKay (1993b) studied convergence of plug-in estimator of (2) in probability and point-wise. Giné and Sang (2010, 2013) studied plug-in estimators of (2) and (6) for one and d-dimensional observations. They proved that the discrepancy between the plug-in estimator and the true value converges uniformly over a data adaptive region at a rate of $O_{a.s.}((\log n/n)^{4/(8+d)})$ by applying empirical process techniques. The plug-in estimator in Giné and Sang (2013) has the following form

$$\hat{f}(t; h_{1,n}, h_{2,n}) = \frac{1}{nh_{2,n}^d}\sum_{i=1}^{n}K\left(\frac{t - X_i}{h_{2,n}}\alpha(\hat{f}(X_i; h_{1,n}))\right)\alpha^d(\hat{f}(X_i; h_{1,n})). \tag{7}$$

In this paper, we concentrate on the study of central limit theorem of the plug-in estimator (7).