Journal of the Korean Statistical Society (2000), **29**: 3, pp 307–313

A study on bandwith selection based on ASE for nonparametric density estimators

Tae Yoon Kim[1]

Abstract

Suppose we have a set of data $X_1, \cdots, X_n$ and employ kernel density estimator to estimate the marginal density of X. In this article bandwidth selection problem for kernel density estimator is examined closely. In particular the Kullback-Leibler method (a bandwidth selection method based on average square error (ASE)) is considered.

Keywords: Bandwidth selection, Kullback-Leibler method, Kernel estimator.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent identically distributed real valued random variables. Consider the problem of estimating the marginal density of X in which the well practiced kernel density estimator is employed;

$$\hat{f}_h(x) = (nh)^{-1} \sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

where $K : R \to R$ is a kernel function, $h = h(n) \in R^+$ is the bandwidth (i.e, smoothing parameter). One of the very decisive points in applying $\hat{f}_h$ is the choice of the bandwidth h. Up to now most of work has been focused on searching for optimal bandwidth minimizing integrated square error (ISE). In this paper the Kullback-Leibler method which is based on averaged square error (ASE) is investigated as a possible alternative bandwidth selector for $\hat{f}_h$. The Kullback-Leibler method is designed to estimate ASE given by

$$d_A(h) = n^{-1} \sum_{j=1}^{n} [\hat{f}_h(X_j) - f(X_j)]^2 f^{-1}(X_j) w(X_j)$$

[1]Department of Statistics, Keimyung University, Taegu 704-701, Korea.

where w denotes a weight function. Then its minimizer $\hat{h}_0$ is often estimated by finding $\hat{h}_K$, the minimizer of

$$-2n^{-1}\log KL(h) = -2n^{-1}\sum_{j=1}^{n}\log\left[\hat{f}_{j,h}(X_j)^{u(X_j)}e^{-\hat{p}(h)}\right]$$

where $\hat{f}_{j,h}$ is "leave one out" version of $\hat{f}_h$; that is, the observation X_j is left out in constructing $\hat{f}_{j,h}$. Here $\hat{p}(h) = \int \hat{f}_h(x)u(x)dx$ and $u(x) = w(x)f(x)$. Marron (1987) investigated the Kullback-Leibler method to establish its asymptotic optimality with regard to ASE. Note that the above Kullback-Leibler method is more natural in the important applications of density estimation to discrimination and to minimum Hellinger distance estimation.

2. THEORETICAL RESULTS

Since Marron (1987), little has been done about the Kullback-Leibler method, though their result surely needs to be stuided further for their practical use. Behind this is that analyzing ASE based selector (e.g.,the Kullback-Leibler method) usually requires getting some difficult features resolved from analytical point of view. For example, behaviour of ASE should be addressed, which has been relatively less explored compared to integrated square error (ISE) (see Hall (1984) and Kim (1997) for possible references). In this article, we study the convergence rate for the Kullback-Leibler method in density estimation. Recall

$$d_A(h) = n^{-1}\sum_{j=1}^{n}[\hat{f}_h(X_j) - f(X_j)]^2 f^{-1}(X_j)w(X_j)$$

and $d_M(h) = Ed_A(h)$. Then the optimal bandwidth is $\hat{h}_0$, the minimizer of the average squared error $d_A(h)$ and let h_0 be the minimizer of $d_M(h)$. If f'' is uniformly continuous, then $d_A(h)$ and $d_M(h)$ are approximately

$$d_m(h) = n^{-1}h^{-1}c_1 + h^4 c_2 \tag{2.1}$$

with $c_1 = \int K^2$ and $c_2 = [\int z^2K(z)dz/2]^2 \int (f'')^2$, in the sense that

$$\sup_{h\in H_n}\left(\left|\frac{d_A(h)-d_M(h)}{d_M(h)}\right| + \left|\frac{d_M(h)-d_m(h)}{d_m(h)}\right|\right) \to 0 \tag{2.2}$$

in probability as $n \to \infty$, where $H_n = [n^{-1+\delta}, n^{-\delta}]$, for arbitrary small $\delta > 0$ (see Marron and Härdle (1986)). Thus one may see that $\hat{h}_0$ and h_0 are each roughly equal to the unique minimizer of d_m, $h_m = c_0 n^{-1/5}$ where $c_0 = (c_1/4c_2)^{1/5}$, that is,

$$\hat{h}_0/h_m, h_0/h_m \to 1 \tag{2.3}$$

in probability. In addition it has been proved by Marron (1987)

$$\hat{h}_K/h_m \to 1 \tag{2.4}$$

in probability. Now our objective is to study how fast the above convergence occurs. For this assume that (i) K is a compactly supported, symmetric function on R with Hölder continuous derivative K', and satisfies $\int K = 1$, $\int z^2 K(z)dz = 2k \neq 0$. (ii) f is bounded and twice differentiable, f' and f'' are bounded and integrable, and f'' is uniformly continuous. Let $c_3 = 2c_1 c_0^{-3} + 12 c_2 c_0^2$. Set $L(z) = -zK'(z)$ and

$$\sigma_c^2 = (2/c_0)^3 (\int f^2)(\int L^2) + (4kc_0)^2 \{\int (f'')^2 f - (\int f'' f)^2\}.$$

Theorem 1. *Under the preceding assumptions*

$$n^{3/10}(\hat{h}_K - \hat{h}_0) \to N(0, \sigma_c^2 c_3^{-2})$$

in distribution.

Hall and Marron (1987) established convergence rates for the least square cross validation method which is ISE based method. Our result shows that the same rate holds for ASE based Kullback-Leibler method for one dimension. From our proof below one may see that extension to higher dimension of our result is not straightforward. Detailed work is left for the upcoming research.

3. PROOFS

Set

$$R = n^{-1}\sum_{j=1}^{n} f(X_j)w(X_j) - E[f(X_j)w(X_j)],$$

$$S = 2n^{-1}\sum_{j=1}^{n}[u(X_j)(1 - \log f(X_j)) - R].$$

And it is convenient to define, for $j = 1, \ldots, n$

$$\Delta_j = \left[\frac{\hat{f}_{j,h}(X_j) - f(X_j)}{f(X_j)}\right] 1_S(X_j)$$

where S is the support of w. Then it follows that

$$-2n^{-1}\log KL(h) = S - 2n^{-1}\sum_{j=1}^{n}[u(X_j)(1+\log(1+\Delta_j)) - \hat{p}(h) - R]$$

$$= S - 2n^{-1}\sum_{j=1}^{n}[u(X_j)(1+\Delta_j) - \hat{p}(h) - R] + d_{A1}(h) - 2n^{-1}\sum_{j=1}^{n} r_j(h)u(X_j)$$

where

$$d_{A1}(h) = n^{-1}\sum_{j=1}^{n}[\hat{f}_{j,h}(X_j) - f(X_j)]^2 f(X_j)^{-1} w(X_j)$$

and r_j denotes the remainder term of the log Taylor expansion. Now the above expression is decomposed as follows.

$$S + 2R + \delta(h) + d_{A1}(h) - 2n^{-1}\sum_{j=1}^{n} r_j(h)u(X_j)$$

$$= S + 2R + \delta(h) + d_M(h) + D_1(h) + D(h) - 2n^{-1}\sum_{j=1}^{n} r_j(h)u(X_j)$$

where $D_1 = d_{A1} - d_A$, $D = d_A - d_M$,

$$\frac{\delta}{2} = \int \hat{f} f w - n^{-1}\sum_{j}\hat{f}_{j,h}(X_j)w(X_j).$$

Then

$$0 = -2n^{-1} d\log KL(\hat{h}_K)/dh = d'_M(\hat{h}_K) + \delta'(\hat{h}_K) + D'_1(\hat{h}_K)$$
$$+ D'(\hat{h}_K) - 2n^{-1}\sum_{j=1}^{n} r'_j(\hat{h}_K)u(X_j).$$

Using (2.3), (2.4) and Lemma 1 below one finds that

$$0 = d'_M(\hat{h}_K) + \delta'(\hat{h}_K) + D'_1(\hat{h}_K) + D'(\hat{h}_K) - 2n^{-1}\sum_{j=1}^{n} r'_j(\hat{h}_K)u(X_j)$$

$$= (\hat{h}_K - h_0)d''_M(h^*) + \delta'(h_0) + D'(h_0) + o_p(n^{-7/10}), \tag{3.1}$$

where h^* lies inbetween h_0 and $\hat{h}_K$. Now it is easily shown that

$$(\hat{h}_K - h_0)d''_M(h^*) = (\hat{h}_K - h_0)c_3 n^{-2/5} + o_p(n^{-7/10}),$$

so we may refine (3.1) as follows:

$$0 = (\hat{h}_K - h_0)c_3 n^{-2/5} + D'(h_0) + \delta'(h_0) + o_p(n^{-7/10}).$$

Also we may have in a similar fashion

$$0 = (\hat{h}_0 - h_0)c_3 n^{-2/5} + D'(h_0) + o_p(n^{-7/10}).$$

Subtracting:

$$0 = (\hat{h}_K - \hat{h}_0)c_3 n^{-2/5} + \delta'(h_0) + o_p(n^{-7/10}).$$

This result and Lemma 3.5 of Hall and Marron (1987) entail

$$n^{3/10}(\hat{h}_K - \hat{h}_0) \to N(0, \sigma_c^2 c_3^{-2})$$

in distribution.

Lemma 1. *For any $0 < a < b < \infty$ and any $j = 1, \cdots, n$*

$$\sup_{a \le t \le b} \{|D_1'(n^{-1/5}t)| + |r_j'(n^{-1/5}t)u(X_j)|\} = o_p(n^{-7/10}). \tag{3.2}$$

For some $\epsilon > 0$

$$|\hat{h}_K - h_0| = O_p(n^{-1/5-\epsilon}) \tag{3.3}$$

For some $\epsilon > 0$ and any $0 < a < b < \infty$,

$$\sup_{a \le t \le b} \{|D'(n^{-1/5}t)| + |\delta'(n^{-1/5}t)|\} = O_p(n^{-3/5-\epsilon}). \tag{3.4}$$

Furthurmore for any $\eta_1 > 0$ and any non-random h_1 asymptotic to a constant multiple of $n^{-1/5}$,

$$\sup_{|h-h_1| \le n^{-1/5-\eta_1}} n^{7/10}\{|D'(h) - D'(h_1)| + |\delta'(h) - \delta'(h_1)|\} = o_p(1). \tag{3.5}$$

Proof. Proof of (3.2) will follow if one shows that

$$\sup_{a \le t \le b} \{|D_1(n^{-1/5}t)| + |r_j(n^{-1/5}t)u(X_j)|\} = o_p(n^{-9/10}). \tag{3.6}$$

(3.6) can be established by using

$$\hat{f}_{h,j}(x) - \hat{f}_h(x) = (n-1)^{-1}\hat{f}_h(x) - (n-1)^{-1}h^{-1}K((x - X_j)/h).$$

and the fact that

$$\sup_{x,j} |\hat{f}_{h,j}(x) - f(x)| = O_p((nh)^{-1/2}).$$

To treat (3.3), notice that $\hat{h}_K/h_0 \to 1$ in probability. As done in (3.1)

$$2n^{-1}\frac{d}{dh}\log KL(h)|_{h=h_0} = 2n^{-1}\frac{d}{dh}\log KL(h)|_{h=h_0} - 2n^{-1}\frac{d}{dh}\log KL(h)|_{h=\hat{h}_K}$$

$$= d'_A(h_0) + \delta'(h_0) - d'_A(\hat{h}_K) + \delta'(\hat{h}_K) + o_p(n^{-7/10})$$

$$= d'_M(h_0) - d'_M(\hat{h}_K) + O_p(n^{-3/5-\epsilon}),$$

using (3.4) and (3.5). But $2n^{-1}\frac{d}{dh}\log KL(h)|_{h=h_0} = d'_M(h_0) + O_p(n^{-3/5-\epsilon})$ and so

$$O_p(n^{-3/5-\epsilon}) = d'_M(h_0) - d'_M(\hat{h}_K) = (h_0 - \hat{h}_K)d''_M(h^*)$$

where h^* lies between h_0 and $\hat{h}_K$. Using $d''_M(h^*) = c_3n^{-2/5} + o_p(n^{-2/5})$, (3.3) follows. (3.4) and (3.5) will follow from Lemma 3.2 of Hall and Marron (1985), (2.4) of Kim (1997), and (2.1). □

ACKNOWLEDGEMENT

This paper was accomplished with research fund provided by Korea Research Foundation, support for 1997 faculty research abroad. The author is very grateful for the unknown referee's careful reading of the original manuscript.

REFERENCES

Hall, P. (1984) Central limit theorem for integrated error of multivariate nonparametric density estimation. *Journal of Multivariate Analysis.* **14**, 1-16.

Hall, P. and Marron, J. S. (1987) Extent to which least equares cross validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields.* **74**, 567-581.

Kim, T.Y. (1997) Central limit theorem for quadratic errors of nonparametric estimators *Journal of Statistical Planning and Inference.* **64**, 193-204.

Marron, J. S. (1987) A comparison of cross-validation techniques in density estimation. *Annals of Statistics.* **15** 152-162.

Marron, J. S. and Härdle W. (1986) Random approximations to some measures of accuracy in nonparametric curve estimation. *Journal of Multivariate Analysis.* **20**, 91-113.