Journal of Statistical
Theory & Methods
1995, Vol. 6, pp 85~96

Jackknife Kernel Density Estimation Using Uniform Kernel Function in the Presence of k's Unidentified Outliers

Jungsoo Woo[1], Jangchoon Lee[2]

Abstract The purpose of this paper is to propose the kernel density estimator and the jackknife kernel density estimator in the presence of k's unidentified outliers, and to compare the small sample performances of the proposed estimators in a sense of mean integrated square error(MISE).

Keywords : Jackknife kernel density estimation, unidentified outlier, MISE

1. Introduction

Suppose that $X_1,\cdots,X_n$ are independent and identically distributed random variables with distribution function $F(x)$ and probability density function $f(x)$, and let

$$\hat{f}_n(x) = \int_{-\infty}^{\infty} \frac{1}{h(n)} K\left(\frac{x-y}{h(n)}\right) d\hat{F}_n(y) = \frac{1}{nh(n)} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h(n)}\right), \tag{1}$$

where $K(x) \geq 0$, $K(x)dx = 1$, $\sup_x |K(y)| < \infty$, $\int_{-\infty}^{\infty} |K(y)| dy < \infty$, $\lim_{n\to\infty} |K(y)| = 0$. Here $K(\cdot)$ is called the kernel function and $h(n)$ is called the bandwidth.

The jackknife method for density estimation was first introduced and illustrated by Sommers(1972), and later developed by Schucany and Sommers(1977). Rustagi and Dynin(1983) studied the effect of jackknifing by using leave-out rules and defined pseudovalues in case of density estimates. Kang(1989) proposed the jackknife and generalized second order jackknife estimators of the probability density function. And Lee(1994) studied the kernel density estimator and the

[1] Department of Statistics, Yeungnam University, Kyungbuk, Republic of Korea
[2] Daegu Junior College, Daegu, Republic of Korea

jackknife kernel density estimator in the presence of an unidentified outlier.

The purpose of this thesis is to propose the kernel density estimator and the jackknife kernel density estimator in the presence of k's unidentified outliers, and to compare the small-sample performancesof the proposed estimators in a sense of mean integrated square error(MISE).

2. Jackknife Kernel Density Estimation in the Presence of k's Unidentified Outliers

We shall consider the maximum likelihood estimator(MLE) and the jackknife estimator of the probability density function through the kernel functions when k's ($<2/n$) unidentified outliers are present. Let $X_1,\cdots,X_n$ be independent random variables, and $X_1,\cdots,X_{n-k}$ be identically distributed random variables with absolutely continuous distribution function $F(x)$ and probability density function $f(x)$, but $X_{n-k+1},\cdots,X_n$ are a random variables with absolutely continuous distribution function $F_o(x)$ and probability density function $f_o(x)$. Then, we define a kernel density estimator

$$\tilde{f}_n(x)=\int_{-\infty}^{\infty}\frac{1}{h(n)}K\left(\frac{x-y}{h(n)}\right)d\tilde{F}_n(y)$$

$$=\frac{1}{nh(n)}\left[\sum_{j=1}^{n-k}K\left(\frac{x-X_j}{h(n)}\right)+\sum_{l=n-k+1}^{n}K\left(\frac{x-X_l}{h(n)}\right)\right], \qquad (2)$$

where a kernel function $K(y)$ satisfies conditions (1).

The expectation and variance of the kernel density estimator defined by (2) have the following forms

$$E\left[\tilde{f}_n(x)\right]=\frac{n-k}{nh(n)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy+\frac{k}{nh(n)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy, \qquad (3)$$

$$VAR\left[\tilde{f}_n(x)\right]=\frac{1}{nh(n)^2}\left[\frac{n-k}{n}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f(y)dy+\frac{k}{n}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right.$$
$$\left.-\frac{n-k}{n}\left\{\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\right\}^2-\frac{k}{n}\left\{\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right\}^2\right]. \qquad (4)$$

Theorem 2.1 The kernel density estimator defined by (2) is asymptotically unbiased at all points x at which the probability density functions $f(x)$ and

$f_o(x)$ are continuous if the constant $h(n)$ satisfies $\lim_{n\to\infty} h(n)=0$ and the kernel satisfies (1).

Theorem 2.2 If $\lim_{n\to\infty} h(n)=0$, $\lim_{n\to\infty} nh(n)=\infty$ and the kernel satisfies conditions (1), then $\tilde{f}_n(x)$ is MSE consistent estimator of the probability dentisty function at all point x of continuity of $f(x)$ and $f_o(x)$.

Let $\tilde{F}_{n-1}^{-i}(x)$ be the empirical distribution function of the random samples $X_1,\cdots,X_n$ with the observation X_i removed, where random variables $X_{n-k+1},\cdots,X_n$ are k's unidentified outliers. Then the kernel estimator of probability density function $f(x)$ with X_i removed is defined by

$$\tilde{f}_{n-1}^{-i}(x)=\frac{1}{h(n-1)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)d\tilde{F}_{n-1}^{-i}(y), \tag{5}$$

where {h(n-1)} is a sequence of constants based on n-1 observations.

The pseudovalues of the jackknife and the jackknife kernel density estimator of the probablity density function, when k's unidentified outliers are present, are defined by the following; for any real number $R\neq 1$.

$$J_n(\tilde{f}_n(x))=\frac{\tilde{f}_n(x)-R\tilde{f}_{n-1}^{-i}(x)}{1-R},\quad i=1,\cdots,n, \tag{6}$$

$$J(\tilde{f}_n(x))=\frac{1}{n(1-R)}\left[\frac{1}{h(n)}\left\{\sum_{j=1}^{n-k}K\left(\frac{x-X_j}{h(n)}\right)+\sum_{l=n-k+1}^{n}K\left(\frac{x-X_l}{h(n)}\right)\right\}\right.$$
$$\left.-\frac{R}{h(n-1)}\left\{\sum_{j=1}^{n-k}K\left(\frac{x-X_j}{h(n-1)}\right)+\sum_{l=n-k+1}^{n}K\left(\frac{x-X_l}{h(n-1)}\right)\right\}\right]. \tag{7}$$

The expectation of the jackknife kernel density estimator defined by (7) has the following form;

$$E\left[J(\tilde{f}_n(x))\right]=\frac{1}{n(1-R)}\left[\left\{\frac{n-k}{h(n)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy+\frac{k}{h(n)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_0(y)dy\right\}\right.$$
$$\left.-\left\{\frac{(n-k)R}{h(n-1)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(y)dy+\frac{kR}{h(n-1)}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right\}\right]. \tag{8}$$

Theorem 2.3 If the constant $h(n)$ satisfies $\lim_{n\to\infty} h(n)=0$ and the kernel satisfies (1), and in addition $\lim_{n\to\infty} R\neq 1$ exists, then the jackknife kernel density estimator defined by (7) is asymptotically unbiased at all points x at which the $f(x)$ and $f_o(x)$ are continuous.

$$VAR\left[J(\tilde{f}_n(x))\right]=\frac{1}{n(1-R)^2}\left[\frac{n-k}{nh(n)^2}\left\{\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f(y)dy-\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\right)^2\right\}\right.$$
$$+\frac{k}{nh(n)^2}\left\{\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f_o(y)dy-\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)^2\right\}$$
$$+\frac{(n-k)R^2}{nh(n-1)^2}\left\{\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n-1)}\right)f(y)dy-\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(y)dy\right)^2\right\}$$
$$+\frac{kR^2}{nh(n-1)^2}\left\{\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy-\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right)^2\right\}$$
$$+\frac{2(n-k)R}{nh(n)h(n-1)}\left\{\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)K\left(\frac{x-y}{h(n-1)}\right)f(y)dy\right.$$
$$\left.-\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\right)\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(y)dy\right)\right\}$$
$$+\frac{2kR}{nh(n)h(n-1)}\left\{\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right.$$
$$\left.\left.-\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right)\right\}\right]. \tag{9}$$

Theorem 2.4 $J(\tilde{f}_n(x))$ is MSE consistent estimator of the probability density functionat all continuous points x of $f(x)$ and $f_o(x)$, if $\lim_{n\to\infty}h(n)=0$, $\lim_{n\to\infty}nh(n)=\infty$, $\lim_{n\to\infty}\frac{h(n)}{h(n-1)}=1$,and there exists $\lim_{n\to\infty}R\neq 1$.

When k's unidentified outliers are present, the general expressions for the mean integrated square error of the kernel density and the jackknife kernel density estimators of the density function are given by

$$MISE\left[\tilde{f}_n(x)\right]=\frac{n-k}{n^2h(n)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f(y)dydx$$
$$+\frac{k}{n^2h(n)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f_o(y)dydx+\frac{(n-k)(n-k-1)}{n^2h(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\right)^2dx$$

$$+\frac{2k(n-k)}{n^2h(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)dx$$

$$-\frac{2(n-k)}{nh(n)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(x)f(y)dydx-\frac{2k}{nh(n)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(x)f_o(y)dydx$$

$$+\frac{k(k-1)}{n^2h(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)^2dx+\int_{-\infty}^{\infty}f(x)^2dx, \tag{10}$$

$$MISE\left[J(\tilde{f}_n(x))\right]=\frac{1}{n(1-R)^2}\left[\frac{n-k}{nh(n)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f(y)dydx\right.$$

$$+\frac{k}{nh(n)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f_o(y)dydx$$

$$+\frac{(n-k)(n-k-1)}{nh(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n)}\right)f(y)dy\right)^2dx$$

$$+\frac{k(k-1)}{nh(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)^2dx$$

$$+\frac{2k(n-k)}{nh(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)dx$$

$$+\frac{(n-k)R^2}{nh(n-1)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n-1)}\right)f(y)dydx$$

$$+\frac{kR^2}{nh(n-1)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n-1)}\right)f_o(y)dydx$$

$$+\frac{(n-k)(n-k-1)R^2}{nh(n-1)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K^2\left(\frac{x-y}{h(n-1)}\right)f(y)dy\right)^2dx$$

$$+\frac{k(k-1)R^2}{nh(n-1)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right)^2dx$$

$$+\frac{2k(n-k)R^2}{nh(n-1)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right)dx$$

$$-\frac{2k(n-k)R}{nh(n)h(n-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)K\left(\frac{x-y}{h(n-1)}\right)f(y)dydx$$

$$-\frac{2kR}{nh(n)h(n-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dydx$$

$$
\begin{aligned}
&-\frac{2(n-k)(n-k-1)R}{nh(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(y)dy\right)dx\\
&-\frac{2k(n-k)R}{nh(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right)dx\\
&-\frac{2k(n-k)R}{nh(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(y)dy\right)dx\\
&-\frac{2k(k-1)R}{nh(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f_o(y)dy\right)\Bigg]\\
&-\frac{2}{1-R}\left[\frac{n-k}{nh(n)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(x)f(y)dydx\right.\\
&+\frac{k}{nh(n)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(x)f_o(y)dydx\\
&-\frac{(n-k)R}{nh(n-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(x)f(y)dydx\\
&\left.-\frac{kR}{nh(n-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n-1)}\right)f(x)f(y)dydx\right]+\int_{-\infty}^{\infty}f(x)^2dx\cdot \qquad (11)
\end{aligned}
$$

3. MISE of the Kernel Density and the Jackknife Kernel Dendity Estimators When k's Unidentified Scale and Location Transited Outliers are Present

Now we assume $X_1,\cdots,X_{n-k}$ are the random samples from the exponential distribution which has the probability density function as follows ;

$$f(x)=\exp(-x),\ \ x\geq 0,$$

whereas $X_{n-k+1},\cdots,X_n$ are k's unidentified outlier observations from an exponential distribution which has the probability density function as follow ;

$$f(x)=\exp\{-b(x-c)\},\ \ 0<b,\ 0\leq c\leq x\ ,$$

and K(x) be a rectangular kernel, $k(x)=1/2$, if $|x|\leq 1$.

And we shall choose the constants R in the jackknife kernel density estimator of the form (7) in order to compare with the jackknife kernel density estimator defined by Rustagi and Dynin, and Kang. The several constants are as follow;

$$R = R_o,\quad R_1 = -\frac{h(n)^2}{h(n-1)^2},\ \text{and}\ R_2 = \frac{h(n)^2}{h(n-1)^2}$$

where R_o is a constant which is minimizing the mean integrated square error of the jackknife kernel density estimators with the forms (11). R_o is given by

$$R_0 = \frac{ICOV(\tilde{V}_n,\ \tilde{V}_{n-1}) - IVAR(\tilde{V}_n) + TA}{IVAR(\tilde{V}_{n-1}) - ICOV(\tilde{V}_n,\ \tilde{V}_{n-1}) + TB}$$

where,

$$MA_1(n) = \frac{1}{h(n)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h(n)}\right) f(y)\,dy\,dx,$$

$$MA_2(n) = \frac{1}{h(n)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h(n)}\right) f_o(y)\,dy\,dx,$$

$$MB_1(n) = \frac{1}{h(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f(y)\,dy\right)^2 dx,$$

$$MB_2(n) = \frac{1}{h(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f_o(y)\,dy\right)^2 dx,$$

$$MC_1(n) = \frac{1}{h(n)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f(x) f(y)\,dy\,dx,$$

$$MC_2(n) = \frac{1}{h(n)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f(x) f_o(y)\,dy\,dx,$$

$$MD_1(n) = \frac{1}{h(n)h(n-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) K\left(\frac{x-y}{h(n-1)}\right) f(y)\,dy\,dx,$$

$$MD_2(n) = \frac{1}{h(n)h(n-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) K\left(\frac{x-y}{h(n-1)}\right) f_o(y)\,dy\,dx,$$

$$ME_1(n) = \frac{1}{h(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f(y)\,dy \int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n-1)}\right) f(y)\,dy\right) dx,$$

$$ME_2(n) = \frac{1}{h(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f_o(y)\,dy \int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n-1)}\right) f_o(y)\,dy\right) dx,$$

$$ME_3(n) = \frac{1}{h(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f_o(y)\,dy \int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n-1)}\right) f(y)\,dy\right) dx,$$

$$ME_4(n) = \frac{1}{h(n)h(n-1)}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) f(y)\,dy \int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n-1)}\right) f_o(y)\,dy\right) dx,$$

$$ME_5(n)=\frac{1}{h(n)^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f(y)dy\int_{-\infty}^{\infty}K\left(\frac{x-y}{h(n)}\right)f_o(y)dy\right)dx,$$

$$IVAR(\tilde{V}_n)=\frac{n-k}{n}\left(MA_1(n)-MB_1(n)\right)+\frac{k}{n}\left(MA_2(n)-MB_2(n)\right),$$

$$ICOV(\tilde{V}_n,\tilde{V}_{n-1})=\frac{n-k}{n}\left(MD_1(n)-ME_1(n)\right)+\frac{k}{n}\left(MD_2(n)-ME_2(n)\right),$$

$$TA=\frac{(n-k)^2}{n}\left(ME_1(n)-MB_1(n)\right)+\frac{k}{n}\left(ME_2(n)-MB_2(n)\right)+\frac{k(n-k)}{n}\left(ME_3(n)+ME_4(n)\right)$$
$$-2ME_5(n)\big)-(n-k)\left(MC_1(n-1)-MC_1(n)\right)-k\left(MC_2(n-1)-MC_2(n)\right),$$

$$TB=\frac{(n-k)^2}{n}\left(MB_1(n-1)-ME_1(n)\right)+\frac{k}{n}\left(MB_2(n-1)-ME_2(n)\right)-k\left(MC_2(n-1)+MC_2(n)\right)$$
$$+\frac{k(n-k)}{n}\left(2ME_5(n-1)-ME_3(n)-ME_4(n)\right)-(n-k)\left(MC_1(n-1)-MC_1(n)\right).$$

We shall choose several bandwidth $h(n)$ from Devrole and Gyorfi(1985) for an underlying exponential distribution as follows ;

$$h(n)=\left\{\frac{8}{n\pi}\left(\int\sqrt{f}\Big/\int|f|\right)^2\right\}^{\frac{1}{3}}=\left(\frac{8}{n\pi}\right)^{\frac{1}{3}},\quad h(n)=\left\{6\Big/\left(n\int(f')^2\right)\right\}^{\frac{1}{3}}=\left(\frac{12}{n}\right)^{\frac{1}{3}},$$

$$h(n)=\left\{15\Big/\left(n\int(f')^2\right)\right\}^{\frac{1}{5}}=\left(\frac{30}{n}\right)^{\frac{1}{5}}.$$

Now we can obtain the mean integrated square errors of the kernel density estimator $\tilde{f}_n(x)$ and the jackknife kernel density estimator $J(\tilde{f}_n(x))$ for the rectangular kernel are given by

$$MISE\left[\tilde{f}_n(x)\right]=\frac{1}{4n^2h(n)^2}\Big[(n-k)(n-k-1)\left(\exp\{-2h(n)\}-1\right)$$
$$+\frac{2k(n-k)}{b(1+b)}\left(b^2\exp\{-2h(n)\}+\exp\{-2bh(n)\}-b^2-1\right)+\frac{k(k-1)}{b}\left(\exp\{-2bh(n)\}-1\right)\Big]$$
$$+\frac{1}{nh(n)}\left[(n-k)\exp\{-h(n)\}+\frac{1}{1+b}\left(b\exp\{-h(n)\}+\exp\{-bh(n)\}\right)\right]-\frac{1}{2h(n)}+\frac{1}{2},\qquad(12)$$

$$MISE\left[J(\tilde{f}_n(x))\right]=\frac{1}{n(1-R)^2}\left[\frac{n}{2h(n)}+\frac{nR^2}{2h(n-1)}-\frac{nR}{h(n-1)}+\frac{1}{4nh(n)^2}E_1(n)\right.$$
$$+\frac{R^2}{4nh(n-1)^2}E_1(n-1)-\frac{R}{2nh(n)h(n-1)}\left\{(n-k)(n-k-1)\left(\exp\{-(h(n)+h(n-1))\}\right.\right.$$

$$-\exp\{-(h(n-1)-h(n))\}\big)+\frac{k(k-1)}{b}\big(\exp\{-b(h(n)+h(n-1))\}-\exp\{-b(h(n-1)-h(n))\}\big)$$
$$+\frac{2k(n-k)}{b(1+b)}\big(b^2\big(\exp\{-(h(n)+h(n-1))\}-\exp\{-(h(n-1)-h(n))\}\big)$$
$$+\exp\{-b(h(n)+h(n-1))\}-\exp\{-b(h(n-1)-h(n))\}\big)\big\}\Big]$$
$$-\frac{1}{1-R}\left[\frac{1}{h(n)}E_2(n)-\frac{R}{h(n-1)}E_2(n-1)\right]+\frac{1}{2}, \tag{13}$$

where

$$E_1(n)=(n-k)(n-k-1)\big(\exp\{-2h(n)\}-1\big)+\frac{k(k-1)}{b}\big(\exp\{-2bh(n)\}-1\big)$$
$$+\frac{2k(n-k)}{b(1+b)}\big(b^2\exp\{-2h(n)\}+\exp\{-2bh(n)\}-b^2-1\big),$$
$$E_2(n)=1-\frac{n-k}{n}\exp\{-h(n)\}-\frac{k}{n(1+b)}\big(\exp\{-bh(n)\}+b\exp\{-h(n)\}\big).$$

If $K(x)$ is the rectangular kernel function, then the mean integrated square errors of the kernel density and the jackknife kernel density estimators in the scale and location transited outlier case are the same as the scale transited outlier, on account of the rectangular kernel and properties of location transition. If we set $b=1$, then the results (12) and (13) are the same as the mean integrated square errors of the kernel density estimator $\tilde{f}_n(x)$ and jackknife kernel density estimator $J(\tilde{f}_n(x))$ for a rectanguler kernel in Kang.

From the results (12) and (13), when the kernel is the rectangular and the underlying distribution is assumed the exponential, Table 1 and 2 show the exact numerical values of the mean integrated square errors of the kernel density and the jackknife kernel density estimators in the presence of the outlier numbers k=2's scale and location transited unidentified outliers and when the k are changed for the small sample sizes $n=5(5)30$, several bandwidths $h(n)$, several constants R, and the nuisance parameter b and c.

From Table 1 and 2, we can obtain the small sample performances as followings ;

(1) As b goes to 1 and the smaller outlier numbers k, the MISE tend to decrease.

(2) For given all bandwidths $h(n)$ and the outlier numbers k are changed, the jackknife kernel density estimator $J(\tilde{f}_n(x))$ with $R=R_o$ is better in the sense of MISE than the kernel density estimator $\tilde{f}_n(x)$ and the jackknife kernel density estimators with $R=R_1$ and $R=R_2$ in the small sample size.

Table 1. The Mean Integrated Square Errors of the Kernel and the Jackknife Kernel Density Estimators in the Presence of k=2's Scale and Location Transited Unidentified Outliers When the Underlying Distribution is Assumed the Exponential.

bandwidth	n	$\tilde{f}_n(x)$		$J(\tilde{f}_n(x))$					
				$R=R_o$		$R=R_1$		$R=R_2$	
		b=0.1	b=0.5	b=0.1	b=0.5	b=0.1	b=0.5	b=0.1	b=0.5
	5	.26538	.21535	.26300	.21152	.26302	.21152	.67952	.62887
	10	.17150	.15542	.17120	.15511	.17126	.15518	.72415	.71094
	15	.13918	.13093	.13911	.13086	.13920	.13095	.78525	.77926
$(8/n\pi)^{1/3}$	20	.12154	.11636	.12152	.11634	.12162	.11644	.84025	.83684
	25	.10997	.10634	.10997	.10634	.11006	.10643	.88901	.88681
	30	.10152	.09887	.10159	.09887	.10168	.09896	.93286	.93132
	5	.28269	.23783	.28269	.23779	.28384	.23952	.50135	.45778
	10	.20273	.18745	.20235	.18704	.20392	.18868	.50634	.49471
	15	.17200	.16372	.17149	.16321	.17286	.16457	.53413	.52880
$(12/n)^{1/3}$	20	.15421	.14879	.15368	.14827	.15486	.14944	.56156	.55851
	25	.14207	.13816	.14156	.13766	.14259	.13867	.58701	.58503
	30	.13304	.13003	.13255	.12955	.13346	.13045	.61030	.60891
	5	.28759	.24386	.28757	.24376	.28849	.24512	.64431	.60184
	10	.21896	.20416	.21834	.20305	.21988	.20512	.65861	.64768
	15	.19444	.18631	.19359	.18544	.19513	.18700	.68556	.68064
$(30/n)^{1/5}$	20	.18055	.17515	.17962	.17421	.18109	.17568	.70929	.70649
	25	.17114	.16716	.17018	.16621	.17157	.16760	.72946	.72769
	30	.16412	.16101	.16316	.16006	.16448	.16137	.74740	.74615

$R_0 = Min.MISE\,(R)$ $R_1 = -h(n)^2/h(n-1)^2$ $R_2 = h(n)^2/h(n-1)^2$

Table 2. The Mean Integrated Square Errors of the Kernel and the Jackknife Kernel Density Estimators When the Outlier Numbers k are Changed and the Kernel Function is $(8/n\pi)^{1/3}$ in the Underlying Distribution is Assumed the Exponential.

estimator	n	k=2	k=4	k=6	k=8	k=10	k=12	k=14
	5	.21353						
	10	.15541	.16869					
	15	.13093	.13809	.14699				
$\tilde{f}_n(x)$	20	.11636	.12101	.12669	.13339			
	25	.10634	.10968	.11370	.11840	.12378	12984	
	30	.09887	.10142	.10446	.10798	.11199	.11649	.12147
	5	.21151						
	10	.15511	.16835					
	15	.13085	.13801	.14690				
$R=R_o$	20	.11634	.12099	.12666	.13337			
	25	.10633	.10967	.11369	.11839	.12377	.12984	
	30	.09887	.10142	.10446	.10798	.11199	.11649	.12147
	5	.21151						
	10	.15518	.16840					
	15	.13094	.13809	.14698				
$R=R_1$	20	.11643	.12108	.12675	.13345			
	25	.10642	.10976	.11378	.11848	.12386	.12991	
	30	.09895	.10151	.10455	.10807	.11208	.11657	.12150
	5	.62886						
	10	.71094	.72229					
	15	.77928	.78456	.79216				
$R=R_2$	20	.83684	.83984	.84410	.84988			
	25	.88699	.88873	.89173	.89536	.89985	.90539	
	30	.93094	.93223	.93428	.93684	.94017	.94400	.94860

$R_0 = Min.MISE\,(R)$, $R_1 = -h(n)^2/h(n-1)^2$, $R_2 = h(n)^2/h(n-1)^2$, k<2/n

(3) For given all bandwidths $h(n)$ and the outlier numbers k are changed, the jackknife kernel density estimator with $R = -h(n)^2/h(n-1)^2$ is better than the jackknife kernel density estimator with $R = h(n)^2/h(n-1)^2$ in the sense of MISE.

(4) When the Rectangular kernel case, the kernel density estimator and jackknife kernel density estimator which have $R = h(n)^2/h(n-1)^2$ are not MSE consistent.

(5) For $n = 5(5)30$, the bandwidth $(8/n\pi)^{1/3}$ is preferable.

References

1. Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation : The L1 View.* Wiley, New York.
2. Kang, S. B. (1989). *The Generalized Jackknife Density Estimation,* Ph.D thesis. Hyosung Women's University.
3. Lee, J. C. (1994). *Kernel Density Estimation in the Presence of an Unidentified Outlier,* Ph.D thesis. Yeongnam University.
4. Rustagi, J. S. and Dynin, S. (1985). *The Jackknife Estimate of the Hazard Function.* Technical Report, Ohio State University. Columbus, Ohio.
5. Schucany, W. R., Gray, H. L. and Owen, D. B. (1971). *On Bias Reduction in Estimation.* J.Amer. Statist. Assoc., 66, 524-533.
6. Schucany, W. R. and Sommers, J. P. (1977). *Improvement of Kernel Type Density Estimators.* J. Amer. Statist. Assoc., 72, 420-423.