A Comparative Study for Bandwidth Selection in Kernel Density Estimation

Omar M. Eidous
Yarmouk University, Irbid, Jordan, omarm@yu.edu.jo

Mohammad Abd Alrahem Shafeq Marie
Yarmouk University, Irbid, Jordan, mohammadmarie@yahoo.com

Mohammed H. Baker Al-Haj Ebrahem
Yarmouk University, Irbid, Jordan, malhaj@yu.edu.jo

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Available at: http://digitalcommons.wayne.edu/jmasm/vol9/iss1/26

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
A Comparative Study for Bandwidth Selection in Kernel Density Estimation

Omar M. Eidous    Mohammad Abd Alrahem Shafeq Marie    Mohammed H. Baker Al-Haj Ebrahem
Yarmouk University, Irbid, Jordan

Nonparametric kernel density estimation method does not make any assumptions regarding the functional form of curves of interest; hence it allows flexible modeling of data. A crucial problem in kernel density estimation method is how to determine the bandwidth (smoothing) parameter. This article examines the most important bandwidth selection methods, in particular, least squares cross-validation, biased cross-validation, direct plug-in, solve-the-equation rules and contrast methods. Methods are described and expressions are presented. The main practical contribution is a comparative simulation study that aims to isolate the most promising methods. The performance of each method is evaluated on the basis of the mean integrated squared error for small-to-moderate sample size. Simulation results show that the contrast method is the most promising method based on the simulated families considered.

Key words: Probability Density Function, Bandwidth, Least Squares Cross-Validation, Biased Cross-Validation, Contrast Method, Direct Plug-In, Solve-The-Equation Rules.

Introduction
The Kernel method is widely used in nonparametric density estimation. It produces a kernel estimator for the unknown probability density function \( p.d.f \) \( f(x) \). Many researchers have observed that the choice of the bandwidth (smoothing) parameter, \( h \), is crucial for the effective performance of the kernel estimator (for example, see Scott, 1992). A method that uses the data \( X_1, X_2, \ldots, X_n \) to produce a value for the bandwidth \( h \) is termed a bandwidth selector or data-driven selector.

Various data-driven methods for selecting the bandwidth have been proposed and studied. Most of these methods are based on minimizing the MISE or AMISE. Unfortunately, none of these are available in practice because all of them depend on the unknown probability density function. (See Bowman, 1984; Stone, 1984; Hall & Marron, 1985; Scott & Terrell, 1987; Sheather & Jones, 1991.)


Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a continuous univariate distribution with an unknown pdf \( f(x) \), then the kernel density estimator of \( f(x) \), \( x \in \mathbb{R} \) is defined by Silverman (1986) as

\[
\hat{f}(x; h) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) \cdot (1.1)
\]
where $K_h(u) = h^{-1}K(uh^{-1})$. $K$ is the kernel function and is assumed to be symmetric (Silverman, 1986), and $h$ is the bandwidth (or the smoothing parameter) that controls the degree of smoothing applied to the data. Both $K$ and $h$ are under the control of the user, therefore, their determination is necessary in order to analyze results about the kernel estimator.

The bandwidth can be chosen to minimize the asymptotic mean integrated square error, or AMISE (Silverman, 1986). In this case, $h$ can be obtained by minimizing

$$ MISE = \int \left( \text{Bias} \hat{f}(x) \right)^2 + \text{Var} \left( \hat{f}(x) \right) dx. $$  \hspace{1cm} (1.2)

If $\text{Bias} \hat{f}(x)$ and $\text{Var} (\hat{f}(x))$ are substituted into (1.2), then $h$ is obtained by solving the following equation:

$$ \min_h \text{AMISE} = \min_h \int_{-\infty}^{\infty} \left( \frac{1}{2} h^2 f''(x) k_2 \right)^2 + \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(t) dt dx. $$

Taking the derivative of $\text{AMISE}$ with respect to $h$ and equating to zero yields,

$$ h = k_2^{-2/5} \left[ \int_{-\infty}^{\infty} K^2(t) dt \right]^{1/5} \left[ \int_{-\infty}^{\infty} f'^2(t) dt \right]^{-1/5} n^{-1/5} $$

$$ = \left\{ \frac{\mu(K)}{k_2^2 R(f'')n} \right\}^{1/5}, $$ \hspace{1cm} (1.3)

where

$$ k_2 = \int_{-\infty}^{\infty} t^2 K(t) dt, $$

$$ \mu(K) = \int_{-\infty}^{\infty} K^2(t) dt, $$

and

$$ R(f'') = \int_{-\infty}^{\infty} f'^2(t) dt. $$

Formula (1.3) is disappointing because the optimal bandwidth is a function of the second derivative of the density function being estimated. Therefore, unless the true density is known, it is impossible to know the optimal bandwidth. Moreover, when the true density is known, no estimation problem exists. Nonetheless, the quantity $\int (f''(x))^2 dx$ in (1.3) can be estimated by using a kernel estimator.

### Methodology

#### Selecting the Bandwidth

The practical implementation of the kernel density estimator requires specification of the bandwidth $h$. A widely used criterion is to choose an $h$ that minimizes the AMISE: the bandwidth controls the smoothness of the fitted density curve. Note that a larger $h$ provides a smoother estimate with smaller variance and larger bias, while a smaller $h$ produces a rougher estimate with larger variance and smaller bias.

Most methods for choosing the bandwidth presented in the literature are proposed when the underlying probability density function, $f(x)$ has support $(-\infty, \infty)$. In addition, by surveying the literature, it was found that the methods represented herein are commonly used to estimate the smoothing parameter $h$ in practice.

#### Least squares cross-validation (LSCV)

Least squares cross-validation (LSCV), proposed by Rudemo (1982) and Bowman (1984), is a completely automatic method for choosing the bandwidth $h$. Following Rudemo’s (1982) derivations, the optimal bandwidth estimator can be obtained by minimizing:

$$ \text{LSCV}(h) = \int \hat{f}^2(x; h) dx - 2n^{-1}(n-1)^{-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_h(X_i - X_j) $$ \hspace{1cm} (1.4)

According to Rudemo (1982), formula (1.4) is derived based on the exact MISE. If the kernel function is Gaussian density, then
\[
\int_{-\infty}^{\infty} \hat{f}^2(x; h) dx = \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{x - X_i}{h}\right) K\left(\frac{x - X_j}{h}\right) dx
\]

is the Gaussian kernel. Because \( k_2 = 1 \) and \( \mu(K) = \frac{1}{2\sqrt{\pi}} \), BCV(\( h \)) is given by

\[
BCV(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{3}{32n^2 h^5 \sqrt{\pi}} \left( h^4 \sum_{i=1}^{n} \sum_{j=1}^{n} e^{\frac{(X_i - X_j)^2}{4h^2}} - h^2 \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - X_j)^2 e^{\frac{(X_i - X_j)^2}{4h^2}} + \frac{1}{12} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - X_j)^4 e^{\frac{(X_i - X_j)^2}{4h^2}} \right)
\]
The problem is persistent because it is apparent from (1.8) that the optimal bandwidth $g$ for estimating $\mu(f^{(r)})$ depends on $\mu(f^{(r+2)})$. To overcome this problem Sheather and Jones (1991) suggested estimating $\mu(f^{(r)})$ at some stage and using a simple estimate of bandwidth $g$ chosen with reference to a parametric family, usually a normal density.

Thus, a family of DPI bandwidth selectors exist which depends on the number of stages of functional estimation before a normal scale (NS) is used. Such a rule will be called an $l$-stage DPI bandwidth selector and is denoted by $\hat{h}_{DPI,l}$. The NS may be considered to be a zero-stage DPI bandwidth selector. Wand and Jones (1995) pointed out that no method exists for objective choice of the number of iterations that should be used. If $f$ is a normal density with mean 0 and variance $\sigma^2$, then according to Wand and Jones (1995), $r$ will be

$$\mu(f^{(r)}) = \frac{(-1)^r r!}{(2\sigma)^{r+1}(r/2)!\sqrt{\pi}}.$$  

Note that simulation results presented for the DPI method in the simulation are based on the use of a two-stage DPI bandwidth selector to find the bandwidth. An algorithm for the two-stage DPI method is given by Sheather and Jones (1991).

Solve-the-Equation (STE)

The solve-the-equation (STE) rule is based on the formula for the AMISE-optimal bandwidth. Many authors (Scott, et al., 1977; Sheather, 1986; Park & Marron, 1990; Sheather & Jones, 1991) have required that $h$ be chosen to satisfy the relationship:

$$h = \left[ \frac{\mu(K)}{k_2 \hat{\psi}_4(\gamma(h)) n} \right]^{1/5}.$$  

where the pilot bandwidth for the estimation of $\psi_4$ is a function $\gamma$ of $h$. The choice of $\gamma$ may be denoted by:

$$\gamma(h) = \left[ \frac{2K(4)(0)k_2}{R(K)} \right]^{1/7} (-\hat{\psi}_4(g_1)/\hat{\psi}_6(g_2))^{1/7} h^{5/7}$$

where $\hat{\psi}_4(g_1)$ and $\hat{\psi}_6(g_2)$ are kernel estimates of $\psi_4$ and $\psi_6$, respectively (Sheather & Jones, 1991). The choice of $g_1$ and $g_2$ may be determined by using:

$$g_1 = \left[ \frac{-2K(4)(0)}{\hat{\psi}_4(k_2 n)} \right]^{1/7}$$

and

$$g_2 = \left[ \frac{-2K(6)(0)}{\hat{\psi}_6(k_2 n)} \right]^{1/9}$$

where:

$$\hat{\psi}_4 = \frac{105}{32\sqrt{\pi} \sigma^9},$$

$$\hat{\psi}_6 = \frac{-15}{16\sqrt{\pi} \sigma^7},$$

and

$$K(4)(0) = 3/\sqrt{2\pi},$$

$$K(6)(0) = -15/\sqrt{2\pi}.$$  

Note that this two-stage STE bandwidth selector was used to find the bandwidth in the simulation and the algorithm used to find the $\hat{h}_{STE,2}$ was based on Sheather and Jones (1991).

Contrast Method (CONT)

Ahmad and Ran (1998) introduced the concept of kernel contrast to select the bandwidth $h$ by studying its finite sample and asymptotic properties. The first step in the CONT method is to define the kernel density estimations $\hat{f}_j(x;h)$ based on $q$ kernels, $K_1, K_2, ..., K_q$, $q \geq 2$. After selecting the contrast coefficients $p_1, p_2, ..., p_q$, where

$$\sum_{j=1}^{q} p_j = 0,$$  

the bandwidth that minimizes the MISE($h$)$_{CONT}$ is selected. However, a reasonable choice for estimating $h$ is to minimize
ISE(\(h\))_{CONT}, which does not depend on the unknown density function \(f(t)\). This method was proposed by Ahmad and Ran (2004), where

\[
MISE(h)_{CONT} = E \left[ \int \left( \sum_{j=1}^{q} p_j f_j(x; h) \right)^2 dx \right]
\]

and

\[
ISE(h)_{CONT} = \int \left( \sum_{j=1}^{q} p_j \hat{f}_j(x; h) \right)^2 dx.
\]

Ahmad and Mugdadi (2003) showed that the estimator based on the ISE(\(h\))_{CONT} for \(f(x)\) is consistent. The density estimation using a kernel contrast is denoted by

\[
\hat{f}(x; h) = \sum_{j=1}^{q} c_j \hat{f}_j(x; h).
\]

The kernels may have an equal weight if \(q\) is chosen as an even integer, where

\[
\sum_{j=1}^{q} c_j = 1; c_j = 1/q \text{ for } j=1, \ldots, q
\]

and

\[
\sum_{j=1}^{q} p_j = 0; p_j = -p_j \text{ for } j=1, \ldots, q/2.
\]

The simulation results in this article were found by taking, \(p_1 = -p_2, p_2 = -1, c_1 = c_2 = 1/2\), where \(K_1, K_2\) are the two kernels N(0,1) and N(0,4), respectively. Therefore,

\[
ISE(h)_{CONT} = \frac{1}{2} \int_{-\infty}^{\infty} \left( \sum_{i=1}^{n} \frac{1}{nh} K_1 \left( \frac{x-X_i}{h} \right) - \sum_{j=1}^{n} \frac{1}{nh} K_2 \left( \frac{x-X_j}{h} \right) \right)^2 dx.
\]

Thus,

\[
ISE(h)_{CONT} = \left( \frac{1}{2n^2h^2} \int_{-\infty}^{\infty} \left( \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}} \right)^2 dx \right)
\]

- \left( \frac{1}{n^2h^2} \int_{-\infty}^{\infty} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}} \sum_{j=1}^{n} \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-X_j)^2}{8h^2}} dx \right)

+ \left( \frac{1}{2n^2h^2} \int_{-\infty}^{\infty} \left( \sum_{j=1}^{n} \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-X_j)^2}{8h^2}} \right)^2 dx \right)

= \frac{3}{8h\sqrt{\pi}} - \frac{1}{n^2h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\frac{(X_i-X_j)^2}{16h^2}}

\]

Simulation Study

A simulation study was conducted to compare the several methods discussed for selecting the bandwidth of a kernel density estimator. The methods compared to estimate the bandwidth \(h\) - and consequently \(f(x)\) - are: least squares cross-validation (LSCV), biased cross-validation (BCV), direct plug-in (DPI), solve-the-equation (STE) rules and contrast (CONT). It is important to understand the effects
BANDWIDTH SELECTION IN KERNEL DENSITY ESTIMATION

of the different methods for the estimator of $f(x)$ for different values of the sample size, $n$. In this study, four different normal mixture densities were simulated; these densities are (Marron & Wand, 1992):

a. Gaussian:
$$f_1(x) = \phi(x).$$

b. Kurtotic Unimodal:
$$f_2(x) = \frac{2}{3} \phi(x) + \frac{1}{3} \phi_{1/10}(x).$$

c. Bimodal:
$$f_3(x) = \frac{1}{2} \phi_{2/3}(x + 1) + \frac{1}{2} \phi_{2/3}(x - 1).$$

d. Strongly Skewed:
$$f_4(x) = \sum_{i=1}^{8} \frac{1}{8} \phi_{(2/3)^{i/2}} \{x - 3[(2/3)^{i/2} - 1]\}$$

(1.9)

where $A = A^{-1} \phi(u / A)$ and $\phi$ denotes the probability density function (pdf) of a standard normal variable, that is,
$$\phi_A(u) = \frac{1}{\sqrt{2\pi A}} e^{-\frac{1}{2}u^2 A^{-2}}.$$

These densities represent Symmetric, Kurtotic Unimodal, Bimodal and Strongly Skewed distributions respectively. Figure 1 displays the shapes of these densities, which are a small subset of fifteen normal mixtures used by Marron and Wand (1992).

The general normal mixtures density is given by (Marron & Wand, 1992):
$$f(x) = \sum_{l=1}^{k} w_l \phi_{\sigma_l} (x - \mu_l)$$

where $-\infty < \mu_l < \infty, \sigma_l > 0$ and $w_l$ is a vector with positive entries summing to unity (weight), for $l=1, 2, ..., k$. It is assumed that $f$ has a normal $k$-mixture density with parameters $\{(w_l, \mu_l, \sigma_l^2) : l = 1, 2, ..., k\}$.

Fryer (1976) and Deheuvels (1977) first showed that the MISE could be calculated exactly when both the underlying density and the kernel function are Gaussian. The integrated squared error (ISE) of the estimator - if the true underlying density is known to be $f(x)$ as in equation (1.37) - is given by Marron and Wand (1992) as

$$\text{ISE}(\hat{f}) = \int_{-\infty}^{\infty} \left[ \hat{f}(x; h) - f(x) \right]^2 dx$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_{h_0^2}(X_i - X_j)$$

$$- \frac{2}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{k} w_l \phi_{(h_0^2 + \sigma_l^2)^{1/2}} (X_i - \mu_l)$$

$$+ U(h, 0)$$

where
$$U(h, q) = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{l=1}^{k} w_l w_{l_j} \phi_{(q h^2 + \sigma_l^2 + \sigma_{l_j}^2)^{1/2}} (\mu_l - \mu_{l_j})$$

and the kernel function $K$ is the standard normal. Thus, it is more appropriate to analyze the expected value of the ISE, called the MISE.

For each normal mixture density in (1.9) and each sample size $n = 50, 100, 200, 500$ that were simulated from $f(x)$, 1,000 samples were artificially repeated from each $f(x)$. For each sample, the bandwidth $h$ based on LSCV, BCV, DPI, CONT and STE methods were obtained. Subsequently, for each sample the ISE values were obtained by using (1.9) according to the simulated density $f(x)$. Subsequently, the MISE values were empirically determined as the mean of the ISE values obtained in each sample. Table 1 displays the simulation results and the MISEs against the sample sizes for the different underlying normal mixture densities. Moreover, the relative efficiencies of the contrast (CONT) method against LSCV, BCV, DPI and STE methods are given in Table 2. The rule of relative efficiency is given by

$$RE(h) = \frac{\text{MISE}(\hat{h}_s)}{\text{MISE}(\hat{h}_{\text{CONT}})}.$$
where \( \hat{h}_* \) is the bandwidth which computed from the other methods (see Table 2).

Conclusion
Tables 1 and 2 show the main results of the simulation study. To provide insight into the effect of the sample size and different normal mixture densities on the performance of the various bandwidth selection methods, the following conclusions can be drawn:

1. The MISE for the kernel estimator \( \hat{f}(x; h) \) decreases as the sample size increases for all simulated functions and for all different methods, which coincides with the theoretical properties of the kernel estimator.

2. In terms of the MISE of \( \hat{f}(x; h) \), the performance of the BCV method is acceptable when the data are simulated from a very skewed density \( f_4(x) \), while its performance is inefficient for the other densities.

3. The MISE values of \( \hat{f}(x; h) \) when \( h \) is estimated based on the LSCV or BCV method are large compared with the MISE values produced by the other methods for all simulated densities and for all sample sizes.

Note that conclusions 2 and 3 suggest that these two methods should be disregarded as global method to select the bandwidth \( h \).
Table 1: The MISE (\(\hat{f}\)) for Different Methods to Choose the Value of Bandwidth

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Size</th>
<th>(f_1(x))</th>
<th>(f_2(x))</th>
<th>(f_3(x))</th>
<th>(f_4(x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPI(2-stage)</td>
<td>50</td>
<td>0.12846</td>
<td>0.23448</td>
<td>0.15910</td>
<td>0.76199</td>
</tr>
<tr>
<td>CONT</td>
<td>50</td>
<td>0.12481</td>
<td>0.22572</td>
<td>0.10643</td>
<td>0.58228</td>
</tr>
<tr>
<td>LSCV</td>
<td>50</td>
<td>0.19144</td>
<td>0.28647</td>
<td>0.25236</td>
<td>0.94230</td>
</tr>
<tr>
<td>BCV</td>
<td>50</td>
<td>0.40578</td>
<td>0.43591</td>
<td>0.39941</td>
<td>0.76748</td>
</tr>
<tr>
<td>STE(2-stage)</td>
<td>50</td>
<td>0.13070</td>
<td>0.23831</td>
<td>0.17873</td>
<td>0.78334</td>
</tr>
<tr>
<td>DPI(2-stage)</td>
<td>100</td>
<td>0.12730</td>
<td>0.21665</td>
<td>0.15063</td>
<td>0.75133</td>
</tr>
<tr>
<td>CONT</td>
<td>100</td>
<td>0.12373</td>
<td>0.22057</td>
<td>0.09582</td>
<td>0.56732</td>
</tr>
<tr>
<td>LSCV</td>
<td>100</td>
<td>0.16841</td>
<td>0.26467</td>
<td>0.23532</td>
<td>0.88068</td>
</tr>
<tr>
<td>BCV</td>
<td>100</td>
<td>0.31693</td>
<td>0.38360</td>
<td>0.30008</td>
<td>0.71768</td>
</tr>
<tr>
<td>STE(2-stage)</td>
<td>100</td>
<td>0.12530</td>
<td>0.23352</td>
<td>0.12739</td>
<td>0.76518</td>
</tr>
<tr>
<td>DPI(2-stage)</td>
<td>200</td>
<td>0.12215</td>
<td>0.20491</td>
<td>0.14057</td>
<td>0.74043</td>
</tr>
<tr>
<td>CONT</td>
<td>200</td>
<td>0.11160</td>
<td>0.21296</td>
<td>0.09271</td>
<td>0.55792</td>
</tr>
<tr>
<td>LSCV</td>
<td>200</td>
<td>0.15314</td>
<td>0.25512</td>
<td>0.21145</td>
<td>0.83185</td>
</tr>
<tr>
<td>BCV</td>
<td>200</td>
<td>0.25271</td>
<td>0.30184</td>
<td>0.23965</td>
<td>0.60122</td>
</tr>
<tr>
<td>STE(2-stage)</td>
<td>200</td>
<td>0.11947</td>
<td>0.19695</td>
<td>0.12676</td>
<td>0.73857</td>
</tr>
<tr>
<td>DPI(2-stage)</td>
<td>500</td>
<td>0.11903</td>
<td>0.19237</td>
<td>0.13929</td>
<td>0.73197</td>
</tr>
<tr>
<td>CONT</td>
<td>500</td>
<td>0.11208</td>
<td>0.20337</td>
<td>0.09019</td>
<td>0.55088</td>
</tr>
<tr>
<td>LSCV</td>
<td>500</td>
<td>0.13948</td>
<td>0.24905</td>
<td>0.19998</td>
<td>0.78332</td>
</tr>
<tr>
<td>BCV</td>
<td>500</td>
<td>0.20559</td>
<td>0.28995</td>
<td>0.16698</td>
<td>0.56810</td>
</tr>
<tr>
<td>STE(2-stage)</td>
<td>500</td>
<td>0.10785</td>
<td>0.18467</td>
<td>0.12599</td>
<td>0.71857</td>
</tr>
</tbody>
</table>
4. The DPI and STE methods produce similar results in term of their MISE values for all densities and for all sample sizes. The DPI method performs better than the STE method for small sample sizes and as the sample size increases the STE is better than the DPI method. This indicates that the convergence rate of the STE method is faster than that of the DPI method.

5. The performance of the CONT method generally is better than the performance of the other methods. A significant improvement for the CONT method over the other methods is clearly demonstrated in the bimodal ($f_3(x)$) and the strongly skewed ($f_4(x)$) models.

6. The relative efficiency values in Table 2 show that, for most of the densities and sample sizes, a considerable gain in the relative efficiency for the CONT method is achieved. The relative efficiency values are less than one in some cases, which indicates that the performance of the corresponding method is better than CONT method, but the relative efficiency remains acceptable in these cases.

7. Comparing the MISE values for different methods when the data are simulated from $f_4(x)$ to the MISE values when the data are simulated from the other densities, it may be concluded that $f_4(x)$ is difficult to estimate by any of the methods considered. That is, the strongly skewed density contains features that cannot be recovered from the sample sizes considered.

8. On the basis of the simulation results, the CONT method may be recommended as a global method to select the bandwidth $h$ in kernel density estimation.

<table>
<thead>
<tr>
<th>Table 2: The Relative Efficiency ($RE$) for Different Sample Sizes and Different Normal Mixture Densities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Efficiency</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>RE(h) = $\frac{MISE(\hat{h}<em>{DPI (2-stage)})}{MISE(\hat{h}</em>{CONT})}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>RE(h) = $\frac{MISE(\hat{h}<em>{LSCV})}{MISE(\hat{h}</em>{CONT})}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>RE(h) = $\frac{MISE(\hat{h}<em>{BCV})}{MISE(\hat{h}</em>{CONT})}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>RE(h) = $\frac{MISE(\hat{h}<em>{STE (2-stage)})}{MISE(\hat{h}</em>{CONT})}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
This study has shown that the CONT method is a useful technique for choosing the bandwidth of the kernel estimator. The CONT method produces reasonable estimates for \( f(x) \) in almost all cases considered (see Table 2). Although the conclusions are based on four different densities, many other candidate shapes exist for the densities from which it is assumed that the data was obtained (Marron & Wand, 1992). Therefore, it is not possible to claim that the CONT method performs better than the other methods for any set of data. However, based on the simulation study, the different methods can be ranked in ascending order (best to worst) according to their performances as follows:

1. CONT
2. DPI (2-stage) and STE (2-stage)
3. LSCV
   and lastly,
4. BCV

References


