

5-1-2010

# 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 Ebrahim

*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

Eidous, Omar M.; Marie, Mohammad Abd Alrahem Shafeq; and Ebrahim, Mohammed H. Baker Al-Haj (2010) "A Comparative Study for Bandwidth Selection in Kernel Density Estimation," *Journal of Modern Applied Statistical Methods*: Vol. 9 : Iss. 1 , Article 26.  
DOI: 10.22237/jmasm/1272687900

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 Ebrahim  
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 methods 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  $X_1, X_2, \dots, X_n$  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.)

Marron (1988) presented a list of various methods with discussion, and a survey of smoothing methods for density estimation is provided by Titterton (1985). Sheather (1992) applied several bandwidth selectors to the Old Faithful data. Janssen, et al. (1995) developed and improved scale measures for use in bandwidth selection. Ahmad and Fan (2001) obtained the optimal theoretical bandwidth  $h$  in the general case. Ahmad and Mugdadi (2003) discussed data-based choices of the bandwidth and analyze the kernel density estimation.

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

$$\hat{f}(x; h) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \cdot (1.1)$$

---

Omar M. Eidous is Associate Professor on the Faculty of Science in the Department of Statistics. Email: omarm@yu.edu.jo. Mohammad Abd Alrahem Shafeq Marie is A graduate Student on the Faculty of Science in the Department of Statistics. Email: mohammadmarie@yahoo.com. Mohammed H. Baker Al-Haj Ebrahim is Associate Professor on the Faculty of Science in the Department of Statistics. Email: malhaj@yu.edu.jo.

## BANDWIDTH SELECTION IN KERNEL DENSITY ESTIMATION

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

$$MISE = \int_{-\infty}^{\infty} \left[ \left( Bias \hat{f}(x) \right)^2 + Var \left( \hat{f}(x) \right) \right] dx. \quad (1.2)$$

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

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

Taking the derivative of  $AMISE$  with respect to  $h$  and equating to zero yields,

$$h = k_2^{-2/5} \left\{ \int_{-\infty}^{\infty} K^2(t) dt \right\}^{1/5} \left\{ \int_{-\infty}^{\infty} f''^2(t) dt \right\}^{-1/5} n^{-1/5} = \left\{ \frac{\mu(K)}{k_2^2 R(f'')} \right\}^{1/5}, \quad (1.3)$$

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  $\int_{-\infty}^{\infty} (f''(x))^2 dx$  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:

$$LSCV(h) = \int_{-\infty}^{\infty} \hat{f}^2(x; h) dx - 2n^{-1}(n-1)^{-1} \sum_{i=1}^n \sum_{j \neq i}^n K_h(X_i - X_j) \quad (1.4)$$

According to Rudemo (1982), formula (1.4) is derived based on the exact MISE. If the kernel function is Gaussian density, then

$$\int_{-\infty}^{\infty} \hat{f}^2(x; h) dx$$

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

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

and

$$\sum_{i=1}^n \sum_{j \neq i}^n K_h(X_i - X_j) = \frac{1}{h} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i - X_j)^2}{2h^2}}.$$

Therefore,

$$\text{LSCV}(h) =$$

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

$$- 2n^{-1}(n-1)^{-1} h^{-1} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i - X_j)^2}{2h^2}} \quad (1.5)$$

The optimal bandwidth  $h$  is obtained by minimizing the right side of (1.5) over  $h$ .

Biased Cross-Validation (BCV)

While LSCV method used exact MISE, the biased cross-validation (BCV) is based on the AMISE (Scott & Terrell, 1987). The BCV method suggests the use of the second derivative of the traditional kernel estimator as opposed to the unknown second derivative of  $f(x)$ . The BCV objective function is thus given by:

$$\text{BCV}(h) = \frac{h^4}{4} k_2^2 \mu(\hat{f}''(x; h)) + (nh)^{-1} \mu(K) \quad (1.6)$$

where  $\mu(f) = \int_{-\infty}^{\infty} f^2 dt$  and  $\hat{f}''(x; h)$  is the second derivative of the kernel estimator and  $K$

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

$$\text{BCV}(h) =$$

$$\frac{1}{2nh\sqrt{\pi}} +$$

$$\frac{3}{32n^2 h^5 \sqrt{\pi}} \left( h^4 \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{(X_i - X_j)^2}{4h^2}} \right.$$

$$- h^2 \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2 e^{-\frac{(X_i - X_j)^2}{4h^2}}$$

$$\left. + \frac{1}{12} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^4 e^{-\frac{(X_i - X_j)^2}{4h^2}} \right)$$

The optimal value of  $h$  is obtained by minimizing BCV( $h$ ) over  $h$ .

Direct Plug-In (DPI)

The DPI method is based on the idea of plugging in an estimate of unknown quantity  $\mu(f^{(r)})$  in equation (1.6):

$$\mu(f^{(r)}) = \int_{-\infty}^{\infty} (f^{(r)}(x))^2 dx, \quad r = 2, 4, 6, 8, \dots$$

Sheather and Jones (1991) developed an estimator for  $\mu(f^{(r)})$  based on the kernel estimator with bandwidth  $g$ , which is given by:

$$\mu(f^{(r)}) = n^{-1} \sum_{i=1}^n \hat{f}^{(r)}(X_i; g)$$

$$= n^{-2} \sum_{i=1}^n \sum_{j=1}^n K_g^{(r)}(X_i - X_j). \quad (1.7)$$

According to Wand and Jones (1995), the bias term of the estimator (1.7) can be made to vanish by choosing  $g$  to be equal

$$g = \left[ \frac{-2K^{(r)}(0)}{\mu(f^{(r+2)}) k_2 n} \right]^{1/(r+3)} \quad (1.8)$$

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

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

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:

$$h = \left[ \frac{\mu(K)}{k_2^2 \hat{\psi}_4(\gamma(h)) n} \right]^{1/5}$$

where the pilot bandwidth for the estimation of  $\psi_4$  is a function  $\mathcal{Y}$  of  $h$ . The choice of  $\mathcal{Y}$  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:

$$g_1 = \left[ \frac{-2K^{(4)}(0)}{\hat{\psi}_6 k_2 n} \right]^{1/7}$$

and

$$g_2 = \left[ \frac{-2K^{(6)}(0)}{\hat{\psi}_8 k_2 n} \right]^{1/9}$$

where:

$$\hat{\psi}_8 = \frac{105}{32\sqrt{\pi} \hat{\sigma}^9},$$

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

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

and

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

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,  $K_1, K_2, \dots, K_q$ ,  $q \geq 2$ . After selecting the contrast coefficients  $p_1, p_2, \dots, p_q$ , 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

$$MISE(h)_{CONT} = E \left[ \int_{-\infty}^{\infty} \left( \sum_{j=1}^q p_j \hat{f}_j(x; h) \right)^2 dx \right]$$

and

$$ISE(h)_{CONT} = \int_{-\infty}^{\infty} \left( \sum_{j=1}^q p_j \hat{f}_j(x; h) \right)^2 dx.$$

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

$$\sum_{j=1}^q c_j = 1; c_j = 1/q \text{ for } j=1, \dots, q$$

and

$$\sum_{j=1}^q p_j = 0; p_j = -p_{2j} \text{ for } j=1, \dots, q/2.$$

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,

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

Thus,

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

where:

$$K_1 \left( \frac{x-X_i}{h} \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}}$$

and

$$K_2 \left( \frac{x-X_j}{h} \right) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-X_j)^2}{8h^2}}.$$

Therefore,

$$ISE(h)_{CONT} = \left( \frac{1}{2n^2h^2} \int_{-\infty}^{\infty} \left( \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}} \right)^2 dx - \frac{1}{n^2h^2} \int_{-\infty}^{\infty} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}} \sum_{j=1}^n \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-X_j)^2}{8h^2}} dx + \frac{1}{2n^2h^2} \int_{-\infty}^{\infty} \left( \sum_{j=1}^n \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-X_j)^2}{8h^2}} \right)^2 dx \right) = \frac{3}{8h\sqrt{\pi}} - \frac{1}{n^2h^2} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{(X_i-X_j)^2}{10h^2}}$$

### 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:

$$f_4(x) = \sum_{l=1}^8 \frac{1}{8} \phi_{(2/3)^{l-1}} \{x - 3[(2/3)^{l-1} - 1]\} \quad (1.9)$$

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

$$\phi_A(u) = \frac{1}{\sqrt{2\pi}A} e^{-\frac{1}{2A^2}u^2}.$$

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  $l=1, 2, \dots, k$ . It is assumed that  $f$  has a normal  $k$ -mixture density with parameters  $\{(w_l, \mu_l, \sigma_l^2) : l = 1, 2, \dots, k\}$ .

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

$$\begin{aligned} \text{ISE}(\hat{f}) &= \int_{-\infty}^{\infty} [\hat{f}(x;h) - f(x)]^2 dx \\ &= \frac{1}{n^2} \sum_{i_1=1}^n \sum_{i_2=1}^n \phi_{h\sqrt{2}}(X_{i_1} - X_{i_2}) \\ &\quad - \frac{2}{n} \sum_{i=1}^n \sum_{l=1}^k w_l \phi_{(h^2 + \sigma_l^2)^{1/2}}(X_i - \mu_l) \\ &\quad + U(h, 0) \end{aligned}$$

where

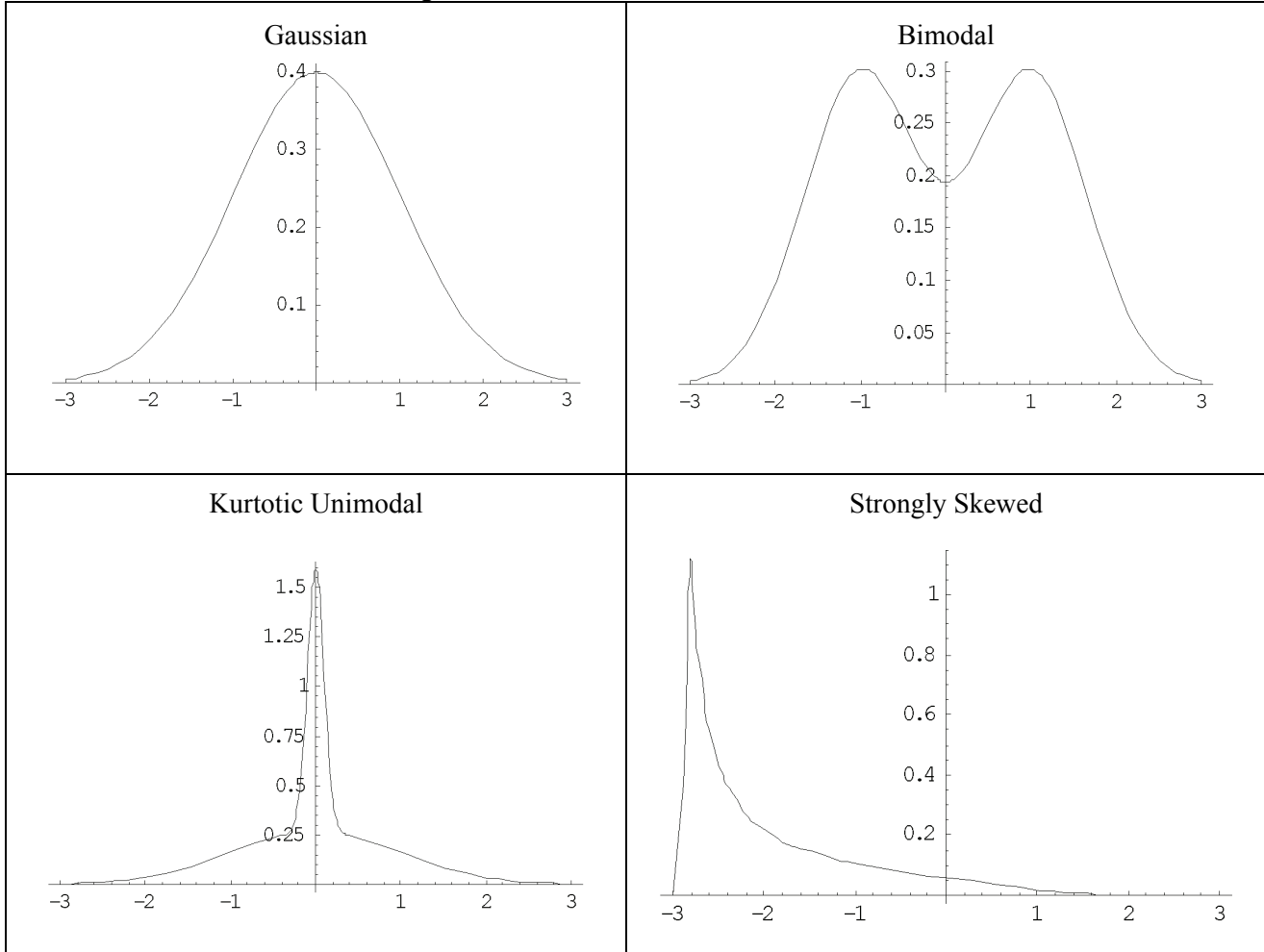
$$U(h, q) = \sum_{l_1=1}^k \sum_{l_2=1}^k w_{l_1} w_{l_2} \phi_{(qh^2 + \sigma_{l_1}^2 + \sigma_{l_2}^2)^{1/2}}(\mu_{l_1} - \mu_{l_2})$$

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

$$RE(\hat{h}) = \frac{\text{MISE}(\hat{h}_*)}{\text{MISE}(\hat{h}_{\text{CONT}})},$$

Figure 1: Some Normal Mixture Densities



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  $h$ .



BANDWIDTH SELECTION IN KERNEL DENSITY ESTIMATION

Table 1: The MISE ( $\hat{f}$ ) for Different Methods to Choose the Value of Bandwidth

Method	MISE ( $\hat{f}$ )				
	Sample Size	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
DPI(2-stage)	50	0.12846	0.23448	0.15910	0.76199
CONT		0.12481	0.22572	0.10643	0.58228
LSCV		0.19144	0.28647	0.25236	0.94230
BCV		0.40578	0.43591	0.39941	0.76748
STE(2-stage)		0.13070	0.23831	0.17873	0.78334
DPI(2-stage)	100	0.12730	0.21665	0.15063	0.75133
CONT		0.12373	0.22057	0.09582	0.56732
LSCV		0.16841	0.26467	0.23532	0.88068
BCV		0.31693	0.38360	0.30008	0.71768
STE(2-stage)		0.12530	0.23352	0.12739	0.76518
DPI(2-stage)	200	0.12215	0.20491	0.14057	0.74043
CONT		0.11160	0.21296	0.09271	0.55792
LSCV		0.15314	0.25512	0.21145	0.83185
BCV		0.25271	0.30184	0.23965	0.60122
STE(2-stage)		0.11947	0.19695	0.12676	0.73857
DPI(2-stage)	500	0.11903	0.19237	0.13929	0.73197
CONT		0.11208	0.20337	0.09019	0.55088
LSCV		0.13948	0.24905	0.19998	0.78332
BCV		0.20559	0.28995	0.16698	0.56810
STE(2-stage)		0.10785	0.18467	0.12599	0.71857

Table 2: The Relative Efficiency (RE) for Different Sample Sizes and Different Normal Mixture Densities

Relative Efficiency	Sample Size	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
$RE(h) = \frac{MISE(\hat{h}_{DPI(2-stage)})}{MISE(\hat{h}_{CONT})}$	50	1.02924	1.03880	1.49487	1.30863
	100	1.02885	0.98222	1.57201	1.32435
	200	1.09453	0.96219	1.51623	1.32712
	500	1.06200	0.94591	1.54440	1.32872
$RE(h) = \frac{MISE(\hat{h}_{LSCV})}{MISE(\hat{h}_{CONT})}$	50	1.53385	1.26913	2.37113	1.61829
	100	1.36110	1.19993	2.45585	1.55235
	200	1.37222	1.19797	2.28076	1.49098
	500	1.24446	1.22461	2.21731	1.42194
$RE(h) = \frac{MISE(\hat{h}_{BCV})}{MISE(\hat{h}_{CONT})}$	50	3.25118	1.93119	3.75279	1.31806
	100	2.56146	1.73913	3.13170	1.26503
	200	2.26442	1.41735	2.58494	1.07761
	500	1.83431	1.42572	1.85142	1.03125
$RE(h) = \frac{MISE(\hat{h}_{STE(2-stage)})}{MISE(\hat{h}_{CONT})}$	50	1.04719	1.05577	1.67932	1.34529
	100	1.01268	1.05871	1.32947	1.34876
	200	1.07052	0.92482	1.36727	1.32379
	500	0.96225	0.90804	1.39694	1.30440

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.

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

- Ahmad, I. A., & Fan, Y. (2001). Optimal bandwidth for kernel density estimators of functions of observations. *Statistics and Probability Letters*, 51(3), 245-251.
- Ahmad, I. A., & Mugdadi, A. R. (2003). Analysis of kernel density estimation of functions of random variables. *Journal of Nonparametric Statistics*, 15, 579-605.
- Ahmad, I. A., & Ran, I. S. (1998). *Kernel contrasts: a data based method of choosing smoothing parameters in nonparametric density estimation*. Unpublished Manuscript.
- Ahmad, I. A., & Ran, I. S. (2004). Kernel contrasts: A data-based method of choosing smoothing parameters in nonparametric density estimation. *Journal of Nonparametric Statistics*, 16, 671-707.
- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71, 353-360.
- Deheuvels, P. (1977). Estimation nonparametrique de la densite par histogrammes generalises. *Rev. Statistical Applications*, 25, 5- 42.
- Fryer, M.J. (1976). Some errors associated with the nonparametric estimation of density functions. *Journal of Instructional Mathematical Applications*, 18, 371-380.
- Hall, P., & Marron, J. S. (1985). Extent to which least-square cross-validation minimizes integrated square error in nonparametric density estimation. *Technical Report, 94*, University of North Carolina, Department of Statistics.
- Janssen, P., Marron, J. S., Veraverbeke, N., & Sarle, W. (1995). Scale measures for bandwidth selection. *Journal of Nonparametric Statistics*, 5, 359-380.
- Marron, J. S. (1988). Automatic smoothing parameter selection: A survey. *Empirical Econometrics*, 13, 187-208.
- Marron, J. S., & Wand, M. P. (1992). Exact Mean Integrated Squared Error. *Annals of Statistics*, 20, 712-736.
- Park, B. U., & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85, 66-72.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9, 65-78.
- Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. New York: Wiley.
- Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82, 1131-1146.
- Scott, D. W., Tapia, R. A., & Thompson, J. R. (1977). Kernel density estimation revisited. *Nonlinear Analysis*, 1, 339-372.
- Sheather, S. J. (1986). An improved data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistical Data Analysis*, 4, 61-65.
- Sheather, S. J. (1992). The performance of six popular bandwidth selection methods on some real data sets (with discussion). *Computational Statistics*, 7, 225-50, 271-281.
- Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, B53*, 683-690.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, 12, 1285-1297.

Titterton, D. M. (1985). Common structure of smoothing techniques in statistics. *International Statistical Review*, 53, 141-170.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman and Hall.