Cross-Validation of Multivariate Densities

Stephan R. Sain, Keith A. Baggerly, and David W. Scott [1]

August 3, 1992

ABSTRACT: In recent years, the focus of study in smoothing parameter selection for kernel density estimation has been on the univariate case, while multivariate kernel density estimation has been largely neglected. In part, this may be due to the perception that calibrating multivariate densities is substantially more difficult. In this paper, we explicitly derive and compare multivariate versions of the bootstrap method of Taylor (1989), the least-squares cross-validation method developed by Bowman (1984) and Rudemo (1982), and a biased cross-validation method similar to that of Scott and Terrell (1987) for multivariate kernel estimation using the product kernel estimator. The theoretical behavior of these cross-validation algorithms is shown to improve (surprisingly) as the dimension increases, approaching the best rate of $O(n^{-1/2})$. Simulation studies suggest that the new biased cross-validation method performs quite well and with reasonable variability as compared to the other two methods. Bivariate examples with heart disease and ozone data are given to illustrate the behavior of these algorithms.

KEY WORDS: Multivariate Kernel Density Estimation, Bootstrap, Biased and Unbiased Cross-Validation.

[1] Stephan R. Sain and Keith A. Baggerly are Graduate Students and David W. Scott is Professor, Department of Statistics, Rice University, Houston, TX 77251-1892. This research was supported in part by the Office of Naval Research under contract N00014-90-J-1176.

1 Introduction

Most work in data-driven methods for kernel density estimation has focused on the univariate case, while multivariate kernel density estimation has been somewhat neglected. Stone (1984) showed the strong theoretical result that, if the underlying multivariate density and its one-dimensional marginals are bounded, smoothing parameters chosen by multivariate least-squares cross-validation (see Bowman (1984) and Rudemo (1982) for details of least-squares cross-validation in the univariate setting) are asymptotically optimal. Terrell (1990) extended the oversmoothing principle of Terrell and Scott (1985) to general multivariate density estimation.

In this paper, we derive and compare multivariate versions of the bootstrap method of Taylor (1989), the least-squares cross-validation method, and a biased cross-validation method similar to that of Scott and Terrell (1987) for multivariate kernel density estimation using the product kernel estimator. These three algorithms represent important and different methods that are widely used in the analysis of univariate data. Higher-order plug-in algorithms for multivariate density estimation have recently been explored by Wand and Jones (1992), but are not included in our study.

The multivariate product kernel estimator is constructed as follows. Let $\mathbf{X}$ be an $n \times d$ data matrix of random vectors $\mathbf{x} = (x_1, \ldots, x_d)$ where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent observations sampled from a multivariate density, $f(\mathbf{x})$, of dimension d. Let x_{ij} denote the ijth entry of $\mathbf{X}$. The multivariate product kernel estimator of $f(\mathbf{x})$ is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{nh_1 \cdots h_d} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{d} K\left(\frac{x_j - x_{ij}}{h_j} \right) \right\} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{d} K_{h_j}\left(x_j - x_{ij} \right) \right\}, \tag{1}$$

where $K_h(u) = h^{-1}K(u/h)$. The same symmetric univariate kernel, K, with variance $\sigma_K^2 = \int u^2 K(u)\, du$, is used in each dimension, but with a different smoothing parameter, h_j, for each dimension. For simplicity, we do not consider data-based rotations of the original data; see Scott (1992) for more on this and other multivariate density estimators.

Appropriate selection of the smoothing parameters is critical to the estimation process. In order to select 'optimal' values of the smoothing parameters, h_j, some criterion function must be used. One such function is the integrated squared error, given by

$$\mathrm{ISE}(\mathbf{h}) = \int_{\Re^d} \left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2 d\mathbf{x}; \tag{2}$$

another is the mean integrated squared error, given by

$$\mathrm{MISE}(\mathbf{h}) = E \int_{\Re^d} \left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2 d\mathbf{x}$$

or, equivalently (by Fubini's Theorem), the integrated mean squared error

$$\mathrm{MISE}(\mathbf{h}) = \mathrm{IMSE}(\mathbf{h}) = \int_{\Re^d} E\left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2 d\mathbf{x}. \tag{3}$$

The goal of data-driven methods such as cross-validation and the bootstrap in multivariate kernel density estimation is to estimate the ISE or MISE using only the data, and to find the vector of smoothing parameters, $\mathbf{h}$, which minimizes the estimated ISE or MISE.

2 The Bootstrap

The use of the bootstrap in univariate kernel density estimation has been studied extensively. The naive approach to using the bootstrap is to resample $x_1^*, \ldots, x_n^*$ from the original data, and then to construct a bootstrap density estimate of the form

$$\hat{f}^*(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i^*}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h\left(x - x_i^*\right).$$

From Equation (3), the global MISE criterion can be computed from the pointwise MSE criterion

$$\mathrm{MSE}(x) = E\left[\left(\hat{f}(x) - f(x)\right)^2\right].$$

The naive bootstrap estimate of $\mathrm{MSE}(x)$ is

$$\mathrm{MSE}^*(x) = E^*\left[\left(\hat{f}^*(x) - \hat{f}(x)\right)^2\right]. \tag{4}$$

However, with the naive bootstrap,

$$E^* \hat{f}^*(x) = E K_h(x - x_1^*) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \hat{f}(x)$$

as x_i^* has a discrete uniform distribution over the original data. Thus the simple bootstrap estimate of the MSE in Equation (4) fails because the bias, $E^*\hat{f}^*(x) - \hat{f}(x)$, vanishes and only the variance is estimated.

Several ideas have been proposed to overcome the problem of the vanishing bias. Hall (1990) restricts the resample size to be less than the original sample size and uses the knowledge of the difference in sample sizes in the construction of the estimator. Taylor (1989) and Faraway and Jhun (1990) propose using the smoothed bootstrap to compute the MSE in Equation (4). Specifically, x_i^* is resampled from the kernel estimate $\hat{f}(x) = n^{-1}\sum_{i=1}^{n} K_h(x - x_i)$ rather than from the raw data themselves (which can be thought of as a kernel estimate with $h = 0$). Taylor (1989) shows that if the smoothed bootstrap technique is used in conjunction with a Normal kernel, no resampling is necessary, as the computations may be given in closed form. It is this method that we will examine in the multivariate setting.

Following Taylor (1989), we consider an estimator of the MISE of the form

$$\widehat{\mathrm{MISE}}(\mathbf{h}) = \int_{\Re^d} E^*\left\{\hat{f}^*(\mathbf{x}) - \hat{f}(\mathbf{x})\right\}^2 d\mathbf{x}, \tag{5}$$

where $\hat{f}(\mathbf{x})$ is the multivariate product kernel estimator given in (1), $\hat{f}^*(\mathbf{x})$ is a multivariate product kernel estimator calculated with data sampled from $\hat{f}(\mathbf{x})$, and the expectation, E^*, is taken with respect to the density $\hat{f}(\mathbf{x})$. By choosing $\mathbf{h}$ to minimize the MISE estimate given in (5), we hope to find a minimum that is close to the minimizer of the unknown MISE.

It is remarkably straightforward to extend Taylor's approach to multiple dimensions. Using the normal kernel for K in Equation (1) and evaluating (5), we obtain as the multivariate bootstrap criterion function (letting $\Delta_{ijk} = (x_{ik} - x_{jk})/h_k$)

$$\mathrm{B}(h_1,\ldots,h_d) = \frac{1}{(2\sqrt{\pi})^d n h_1\cdots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1\cdots h_d}\times$$
$$\sum_{i=1}^{n}\sum_{j\neq i}\left[\frac{n-1}{2^{d/2}n}\exp\left\{-\frac{1}{8}\sum_{k=1}^{d}\Delta_{ijk}^2\right\} + \exp\left\{-\frac{1}{4}\sum_{k=1}^{d}\Delta_{ijk}^2\right\} - \frac{2\times 2^{d/2}}{3^{d/2}}\exp\left\{-\frac{1}{6}\sum_{k=1}^{d}\Delta_{ijk}^2\right\}\right],$$

which is to be minimized (perhaps locally) over $h_1,\ldots,h_d$. We have omitted the $i = j$ terms in the sums to remove the inflated variance which is characteristic of this resampling method (following Taylor (1989)).

3 Least-Squares (Unbiased) Cross-Validation

Whereas the bootstrap method attempts to find the minimum of the MISE, the method of least-squares cross-validation takes a more direct approach by finding a minimum of something between the ISE and the MISE. Expanding Equation (2), we find that

$$\mathrm{ISE}(\mathbf{h}) = R(\hat{f}(\mathbf{x})) - 2\int \hat{f}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + R(f(\mathbf{x})), \tag{6}$$

where $R(\psi(\mathbf{x})) = \int \cdots \int \psi(\mathbf{x})^2 d\mathbf{x}$ is referred to as the roughness of $\psi(\cdot)$. Since $R(f(\mathbf{x}))$ is constant with respect to $\mathbf{h}$, this term may be ignored. Cross-validation, using the leave-one-out estimator

$$\hat{f}_{-i}(\mathbf{x}_i) = \frac{1}{(n-1)h_1 \cdots h_d} \sum_{j \neq i} \left\{ \prod_{k=1}^{d} K\left(\frac{x_{ik} - x_{jk}}{h_k} \right) \right\},$$

allows us to estimate the second term in Equation (6) by noting that

$$E\hat{f}_{-i}(\mathbf{x}_i) = \int \hat{f}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x},$$

which suggests the estimate

$$\mathrm{UCV}(\mathbf{h}) = R(\hat{f}) - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}(\mathbf{x}_i).$$

Thus the expected value of UCV is equal to the expected value of ISE minus the constant $R(f)$. Hence, another name for this method is unbiased cross-validation.

Least-squares cross-validation can easily be generalized to arbitrary dimensions using the multivariate product kernel given in Equation (1). The roughness of $\hat{f}(\mathbf{x})$ when using the standard Normal kernel for K is easily shown to be

$$R(\hat{f}) = \frac{1}{(2\sqrt{\pi})^d n h_1 \cdots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1 \cdots h_d} \sum_{i=1}^{n} \sum_{j \neq i} \exp\left\{ -\frac{1}{4} \sum_{k=1}^{d} \Delta_{ijk}^2 \right\}.$$

The second part of the UCV estimator is given explicitly by

$$\frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}(\mathbf{x}_i) = \frac{2}{(\sqrt{2\pi})^d n^2 h_1 \cdots h_d} \sum_{i=1}^{n} \sum_{j \neq i} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{d} \Delta_{ijk}^2 \right\},$$

where, for simplicity, the factor $n-1$ has been replaced with n. The multivariate least-squares cross-validation function, $\mathrm{UCV}(h_1, \ldots, h_d)$, is then

$$\mathrm{UCV}(h_1, \ldots, h_d) = \frac{1}{(2\sqrt{\pi})^d n h_1 \cdots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1 \cdots h_d} \times$$
$$\sum_{i=1}^{n} \sum_{j \neq i} \left[\exp\left\{ -\frac{1}{4} \sum_{k=1}^{d} \Delta_{ijk}^2 \right\} - (2 \times 2^{d/2}) \exp\left\{ -\frac{1}{2} \sum_{k=1}^{d} \Delta_{ijk}^2 \right\} \right].$$

4 Biased Cross-Validation

The univariate biased cross-validation method investigated by Scott and Terrell (1987) is based on minimizing an estimate of the asymptotic mean integrated squared error (AMISE);

$$\mathrm{AMISE}(h) = \frac{R(K)}{nh} + \frac{1}{4}\sigma_K^4 h^4 R(f''),$$

which is minimized when the smoothing parameter is

$$h_0 = \left[\frac{R(K)}{\sigma_K^4 n R(f'')}\right]^{1/5}. \tag{7}$$

To accomplish this, an estimate of the roughness of the second derivative of the underlying density is needed. The estimator originally proposed by Scott and Terrell (1987) is

$$\hat{R}_1 = \widehat{R(f'')} = R(\hat{f}'') - \frac{R(K'')}{nh^5},$$

where f'' is the second derivative of the underlying density and $\hat{f}''$ is the second derivative of the univariate kernel density estimator with kernel K. The corresponding AMISE estimate is

$$\mathrm{BCV1}(h) = \frac{R(K)}{nh} + \frac{1}{4}\sigma_K^4 h^4 \hat{R}_1.$$

Hall and Marron (1987) have investigated $\hat{R}_1$ and other alternative estimators of $R(f'')$, in particular, one based on the identity

$$R(f'') = \int f''(x)^2 dx = \int f^{iv}(x) f(x) dx = E[f^{iv}(X)],$$

which leads to the estimates

$$\hat{R}_2 = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i}^{iv}(x_i)$$

and

$$\mathrm{BCV2}(h) = \frac{R(K)}{nh} + \frac{1}{4}\sigma_K^4 h^4 \hat{R}_2. \tag{8}$$

The univariate version of BCV2 was first investigated by Jones and Kappenman (1992) and is a special case of smoothed cross-validation (Hall, Marron, and Park (1992)).

Several authors have used higher-order point estimators based on $R(f'')$ to obtain plug-in estimates of h; Sheather and Jones (1991) and Hall et al. (1991), for example. Plug-in estimators go directly from point estimates of $\hat{R}_2$ to estimates of h_0. BCV1, BCV2, UCV, and B, on the other hand, all treat quantities such as $\hat{R}_2$ as functions of h, to be re-evaluated with each potential choice of h. Thus, an estimate of the ISE or AMISE is recomputed for each h, and $\hat{h}$ is taken as the minimizer (often a local minimizer) of the estimated risk function.

It turns out that in the univariate case, the estimator $\hat{R}_2$ has larger variance than the estimator proposed by Scott and Terrell (1987), but improved bias properties. Of particular importance here, $\hat{R}_2$ is easier to extend into a multivariate setting. The ease of implementation and the improved bias quality make it our choice for investigation.

We give an outline of the general derivation of the multivariate BCV2 function for the case $d = 2$. Consider the bivariate form of the AMISE for the product kernel defined in Equation (1) (see Scott (1992) for details):

$$\begin{aligned} \text{AMISE}(h_1, h_2) = \frac{R(K)^2}{nh_1h_2} + \frac{1}{4}\sigma_K^4 \Big\{ h_1^4 \iint f_{x_1x_1}(x_1,x_2)^2 dx_1dx_2 \\ + h_2^4 \iint f_{x_2x_2}(x_1,x_2)^2 dx_1dx_2 + 2h_1^2h_2^2 \iint f_{x_1x_1}(x_1,x_2) f_{x_2x_2}(x_1,x_2) dx_1dx_2 \Big\}, \end{aligned} \tag{9}$$

where K is the univariate kernel with variance σ_K^2 used in (1) and $f_{x_1x_1}$ and $f_{x_2x_2}$ denote the second partials of f with respect to x_1 and x_2, respectively. If we take K to be the standard Normal kernel, then $R(K) = 1/(2\sqrt{\pi})$ and $\sigma_K^2 = 1$.

Using a bivariate form of $\hat{R}_2$ we can estimate the roughness of the second derivatives in the AMISE expression as follows:

$$\iint f_{x_1x_1}(x_1,x_2)^2 dx_1dx_2 = \iint \frac{\partial^4 f(x_1,x_2)}{\partial x_1^4} f(x_1,x_2) dx_1dx_2 = E\left[\frac{\partial^4 f(x_1,x_2)}{\partial x_1^4}\right].$$

The last term in the AMISE expression can be dealt with similarly, noting that

$$\begin{aligned} \iint f_{x_1x_1}(x_1,x_2) f_{x_2x_2}(x_1,x_2) dx_1dx_2 = \\ \iint \frac{\partial^4 f(x_1,x_2)}{\partial x_1^2 \partial x_2^2} f(x_1,x_2) dx_1dx_2 = E\left[\frac{\partial^4 f(x_1,x_2)}{\partial x_1^2 \partial x_2^2}\right]. \end{aligned}$$

These expectations can be estimated respectively by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^4 \hat{f}_{-i}(x_{i1},x_{i2})}{\partial x_1^4} \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^4 \hat{f}_{-i}(x_{i1},x_{i2})}{\partial x_1^2 \partial x_2^2}. \tag{10}$$

Substituting these estimates into (9) and simplifying for the standard Normal kernel, we get as the bivariate form of the new biased cross-validation function

$$\begin{aligned}\mathrm{BCV2}(h_1,h_2) = {} & \frac{1}{4\pi n h_1 h_2} + \frac{1}{4n(n-1)h_1h_2}\times \\ & \sum_{i=1}^{n}\sum_{j\neq i}\left[(\Delta_{ij1}^2+\Delta_{ij2}^2)^2 - 8(\Delta_{ij1}^2+\Delta_{ij2}^2)+8\right]\phi(\Delta_{ij1})\phi(\Delta_{ij2}),\end{aligned}$$

where ϕ is the standard Normal density.

The form of such functions remains simple as dimensionality increases. The general form is

$$\begin{aligned}\mathrm{BCV2}(h_1,\ldots,h_d) = {} & \frac{1}{(2\sqrt{\pi})^d n h_1\cdots h_d} + \frac{1}{4n(n-1)h_1\cdots h_d}\times \\ & \sum_{i=1}^{n}\sum_{j\neq i}\left[\left(\sum_{k=1}^{d}\Delta_{ijk}^2\right)^2 - (2d+4)\left(\sum_{k=1}^{d}\Delta_{ijk}^2\right) + (d^2+2d)\right]\prod_{k=1}^{d}\phi(\Delta_{ijk}).\end{aligned}$$

5 Asymptotic Results

For comparison, we examine the asymptotic standard deviations of these optimal smoothing parameters. We begin with a study in the univariate case summarized in Table 1; a more extensive listing (in a different form) for univariate algorithms may be found in Jones and Kappenman (1992). In the Appendix, we find that for BCV2, the asymptotic standard deviation is given by

$$\sigma(\hat{h}_{\mathrm{BCV2}} - h_0) = \sqrt{2}C(f,K)\sigma_K^{14/5}R[uK^v(u)]^{1/2}n^{-3/10}, \tag{11}$$

where

$$C(f,K) = \frac{1}{20}R(f)^{1/2}R(f'')^{-3/10}R(K)^{-7/10}.$$

If f is the standard Normal density and K is the standard Normal kernel, this reduces to $\sigma(\hat{h}_{\mathrm{BCV2}} - h_0) = 0.4304n^{-3/10}$.

For BCV1, the asymptotic standard deviation is given by

$$\sigma(\hat{h}_{\mathrm{BCV1}} - h_0) = \sqrt{2}C(f,K)\sigma_K^{14/5}R[uK''{*}K'''(u)]^{1/2}n^{-3/10},$$

where $K''{*}K'''(u) = \int K''(w)K'''(u-w)dw$ (i.e. $*$ denotes a convolution). For the standard Normal density and kernel, this reduces to $\sigma(\hat{h}_{\mathrm{BCV1}} - h_0) = 0.0905n^{-3/10}$, considerably

Table 1: Asymptotic standard deviations of $\hat{h}$ for $N(0,1)$ data.

	B	BCV1	UCV	BCV2
$\sigma(\hat{h}-h_0)$	$0.0376n^{-3/10}$	$0.0905n^{-3/10}$	$0.3587n^{-3/10}$	$0.4304n^{-3/10}$
Ratio	1.00	2.41	9.54	11.45

smaller than that of the new biased cross-validation method presented above. Similarly, for the unbiased cross-validation method,

$$\sigma(\hat{h}_{\mathrm{UCV}} - h_0) = C(f,K)\sigma_K^{-6/5} R[uK{*}K'(u) - 2uK'(u)]^{1/2} n^{-3/10}.$$

Simplifying for the standard Normal density and kernel, $\sigma(\hat{h}_{\mathrm{UCV}} - h_0) = 0.3587n^{-3/10}$, which is slightly smaller than that of the new biased cross-validation method. Both of these results were shown by Scott and Terrell (1987).

Jones and Kappenman (1992) show that the expression for $\sigma(\hat{h}_{\mathrm{B}} - h_0)$ is similar to that of $\sigma(\hat{h}_{\mathrm{UCV}} - h_0)$ and is given by

$$\sigma(\hat{h}_{\mathrm{B}} - h_0) = C(f,K)\sigma_K^{-6/5} R\left[uK{*}K{*}K{*}K'(u) - 2uK{*}K{*}K'(u) + uK{*}K'(u)\right]^{1/2} n^{-3/10}.$$

Simplifying for the standard Normal density and kernel, $\sigma(\hat{h}_{\mathrm{B}} - h_0) = 0.0376n^{-3/10}$.

Since $\hat{h} - h_0 = O(n^{-3/10})$ and $h_0 = O(n^{-1/5})$, $(\hat{h} - h_0)/h_0 = O(n^{-1/10})$. This slow rate of convergence has been noted by many authors, and several algorithms with faster rates of convergence have been constructed for univariate data. However, Jones and Kappenman (1992) and Chiu (1992) observed that in simulations the actual ISE's of most of these algorithms are remarkably similar (with h_{ISE} and h_{MISE} the best).

At first, we assumed that the multivariate algorithms would perform more poorly than the univariate ones; however, we were pleasantly surprised to find that such was not the case.

Proposition. Let $\hat{h}$ be the (local) minimizer of one of the cross-validation criteria with the constraint $h_1 = h_2 = \cdots = h_d$. Then, under suitable conditions,

(a) $n^{d/(2d+8)}\left((\hat{h} - h_0)/h_0\right) \xrightarrow{\mathcal{D}} N(0, \alpha_1^2)$, and

(b) $n^{(d+2)/(2d+8)}\left(\hat{h} - h_0\right) \xrightarrow{\mathcal{D}} N(0, \alpha_3^2)$, where α_1 and α_3 are defined in Equations (16) and (17).

Remark 1. Both the relative and absolute performance actually improve as the dimension increases, approaching $O(n^{-1/2})$ in the limit.

Remark 2. The constraint $h_k = h$ is for theoretical convenience, as the minimizer of $\mathrm{AMISE}(h_1, \ldots, h_d)$ cannot be given explicitly, whereas with the constraint it can be. The qualitative behavior of $\hat{h}_k$ should be similar to that of $\hat{h}$.

Proof: (Sketch for BCV2) For the product kernel estimator defined in Equation (1),

$$\mathrm{AMISE}(h, \ldots, h) = \frac{R(K)^d}{nh^d} + \frac{1}{4}\sigma_K^4 h^4 \sum_{k=1}^{d}\sum_{\ell=1}^{d} R(f_{x_k x_\ell}). \tag{12}$$

Leave-one-out estimates of the unknowns $R(f_{x_k x_\ell})$ may be constructed; for example (replacing $n-1$ with n)

$$\hat{R}(f_{x_1 x_1}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^4 \hat{f}_{-i}(\mathbf{x}_i)}{\partial x_1^4} = \frac{1}{n^2 h^d}\sum_{i=1}^{n}\sum_{j\neq i}\left[\frac{1}{h^4} K^{iv}\left(\frac{x_{i1}-x_{j1}}{h}\right)\prod_{k=2}^{d} K\left(\frac{x_{ik}-x_{jk}}{h}\right)\right]$$

so that

$$\begin{aligned}\mathrm{BCV2}(h, \ldots, h) = \frac{R(K)^d}{nh^d} + \frac{\sigma_K^4}{4n^2h^d}\times \\ \sum_{i=1}^{n}\sum_{j\neq i}\left[K^{iv}\left(\frac{x_{i1}-x_{j1}}{h}\right)\prod_{k=2}^{d} K\left(\frac{x_{ik}-x_{jk}}{h}\right) + \cdots\right].\end{aligned} \tag{13}$$

Then $\hat{h}$ is a root of

$$\begin{aligned}\frac{\partial}{\partial h}\mathrm{BCV2}(h, \ldots, h) = \frac{-d\cdot R(K)^d}{nh^{d+1}} - \frac{d\cdot\sigma_K^4}{4n^2h^{d+1}}\times \\ \sum_{i=1}^{n}\sum_{j\neq i}\left[K^{iv}\left(\frac{x_{i1}-x_{j1}}{h}\right)\prod_{k=2}^{d} K\left(\frac{x_{ik}-x_{jk}}{h}\right) + \cdots\right].\end{aligned} \tag{14}$$

(The inside derivative of the bracketed terms in Equation (13) with respect to h gives terms of the same order.) Straightforward approximations similar to those in the Appendix indicate that the expectation of the terms in the square brackets is $O(h^{d+4})$ while the variance of the terms in the square brackets is $O(h^d)$. A theorem of Hall (1984) indicates that the double sum is asymptotically Normal. Thus $\hat{h}$ is the root of a (random) equation of the following form (letting $Z \sim N(0,1)$ and c_1, c_2, and c_3 be constants):

$$\begin{aligned}\frac{\partial}{\partial h}\text{BCV2}(h,\ldots,h) &= \frac{-c_1}{nh^{d+1}} + \frac{c_2}{n^2h^{d+1}}n^2h^{d+4} + Z\sqrt{\frac{c_3^2}{n^4h^{2d+2}}n^2h^d} \\ &= \frac{-c_1}{nh^{d+1}} + c_2h^3 + Z\frac{c_3}{nh^{(d+2)/2}}\end{aligned}$$

or, equivalently,

$$-c_1 + c_2n\hat{h}^{d+4} + c_3Z\hat{h}^{d/2} = 0 \tag{15}$$

when $h = \hat{h}$. Taking $\hat{h}$ to be of the form

$$\hat{h} = h_0(1 + \alpha_1 n^{-\alpha_2}Z),$$

where $h_0 = [c_1/(c_2n)]^{1/(d+4)}$ (which turns out to be the minimizer of (12)), and α_1 and α_2 are constants to be determined, substituting into Equation (15) and equating leading factors of Z in a Taylor's expansion yields

$$\alpha_1 = \frac{1}{d+4}c_1^{-(d+8)/(2d+8)}c_2^{-d/(2d+8)}c_3 \qquad \text{and} \qquad \alpha_2 = d/(2d+8), \tag{16}$$

from which part (a) of the Proposition follows immediately. Since $\hat{h} = h_0 + \alpha_1 h_0 n^{-\alpha_2}Z$, it follows that

$$\alpha_3 = \alpha_1 h_0 = \frac{1}{d+4}c_1^{-(d+6)/(2d+8)}c_2^{-(d+2)/(2d+8)}c_3, \tag{17}$$

which completes part (b) of the Proposition.

We have computed the asymptotic behavior of B, BCV2, and UCV in $\Re^2$ and $\Re^3$. The tedious nature of these calculations can be illustrated by computing the variance of all of the terms that appear in a complete version of Equation (14) when $d = 2$:

$$\begin{aligned}Var\left[\frac{\partial}{\partial h}\text{BCV2}(h,h)\right] &= \frac{\sigma_K^8}{8n^2h^4}R(f)\left[3R(K'')^2 + 4R(K')R(K''')+\right. \\ & R(K)R(K^{iv}) + 2R(K^{iv})R(uK'(u)) + 8R(K''')R(uK''(u)) + \\ & \left. 12R(K'')R(uK'''(u)) + 8R(K')R(uK^{iv}(u)) + 2R(K)R(uK^{v}(u))\right].\end{aligned}$$

In Table 2 we simply present the numerical results for the case of a standard Normal kernel and $N(\mathbf{0}, \mathbf{I}_d)$ data, for which

$$h_0 = \left[\frac{4}{(d+2)n}\right]^{1/(d+4)}.$$

Table 2: Asymptotic standard deviations for $d = 1, 2, 3$ with $N(\mathbf{0}, \mathbf{I}_d)$ data.

	B	UCV	BCV2
$\sigma(\hat{h} - h_0)$ in $\Re^1$	$0.038n^{-3/10}$	$0.359n^{-3/10}$	$0.430n^{-3/10}$
$\sigma(\hat{h} - h_0)$ in $\Re^2$	$0.024n^{-1/3}$	$0.243n^{-1/3}$	$0.354n^{-1/3}$
$\sigma(\hat{h} - h_0)$ in $\Re^3$	$0.017n^{-5/14}$	$0.187n^{-5/14}$	$0.260n^{-5/14}$

For $\mathbf{x} \in \Re^2$, $\hat{h}_{\mathrm{B}} - h_0 = 0.024n^{-1/3}Z$, $\hat{h}_{\mathrm{UCV}} - h_0 = 0.243n^{-1/3}Z$, and $\hat{h}_{\mathrm{BCV2}} - h_0 = 0.354n^{-1/3}Z$, whereas for $\mathbf{x} \in \Re^3$, $\hat{h}_{\mathrm{B}} - h_0 = 0.017n^{-5/14}Z$, $\hat{h}_{\mathrm{UCV}} - h_0 = 0.187n^{-5/14}Z$, and $\hat{h}_{\mathrm{BCV2}} - h_0 = 0.260n^{-5/14}Z$. These were calculated symbolically.

6 Simulation Results

In a small study, one hundred samples of various sizes were taken from a $N(\mathbf{0}, \mathbf{I}_d)$ distribution. For each sample, local minima of the bootstrap and both cross-validation functions were found using the S-Plus function *nlmin* (Statistical Sciences, Inc. (1991)). This optimizer is based on a general quasi-Newton method. For simplicity, the same smoothing parameter was used in each dimension when $d > 1$. Specifically, the formula $h = n^{-1/(4+d)}$ was used to provide a starting value for the iterations and the bandwidth was taken to be the local minimizer found. It is well known that some small percentage of such cross-validation curves have more than one minimizer, but this algorithm will choose only one. In some cases, we checked the algorithm by plotting the criteria over a grid of h values.

Worton (1989) showed that for multivariate Normal $N(\mathbf{0}, \sigma^2\mathbf{I}_d)$ data and a Normal product kernel method (i.e., the same kernel and smoothing parameter used in each dimension), the exact mean integrated squared error (MISE) can be calculated for a given bandwidth as follows:

$$(4\pi)^{\frac{d}{2}}\mathrm{MISE}(h) = n^{-1}h^{-d} + \sigma^{-d} + (1 - n^{-1})(\sigma^2 + h^2)^{-d/2} - 2^{-(d+2)/2}(2\sigma^2 + h^2)^{-d/2}. \quad (18)$$

We computed the values h_{MISE} that minimize (18) for each given sample size. These values together with the simulation results are summarized in Figures 1, 2, and 3 for dimensions $1 \le d \le 3$ and 100 repetitions at each sample size.

The bootstrap estimates of the smoothing parameters are, on average, considerably

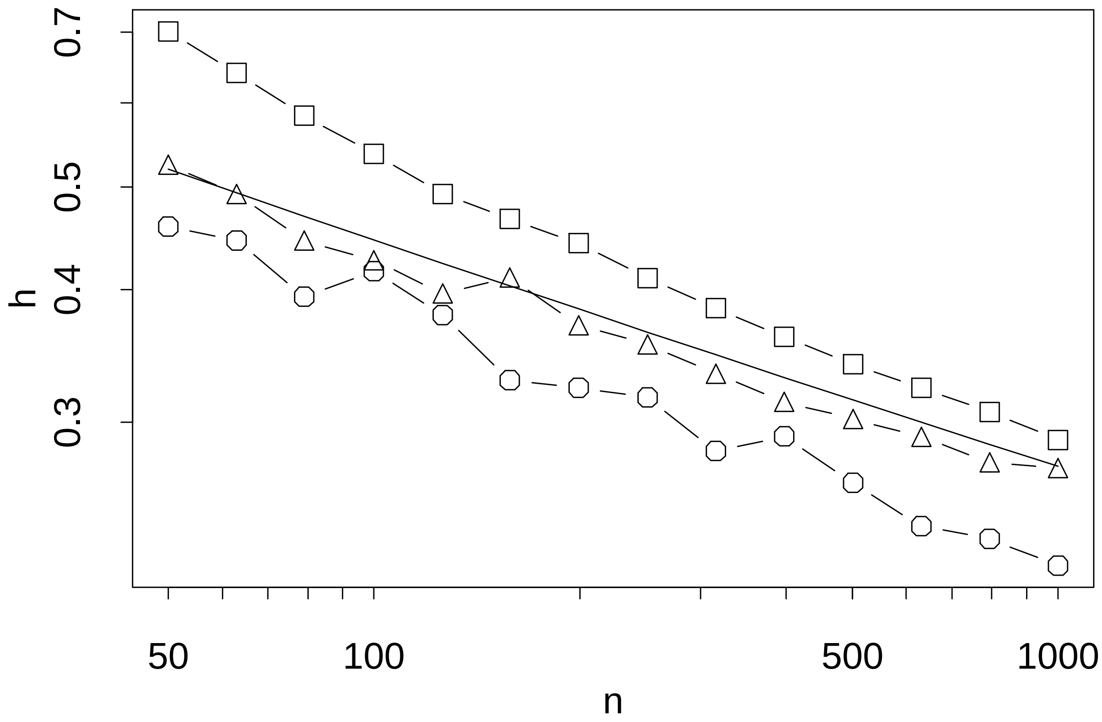

Figure 1: Means of estimated values of univariate smoothing parameters for 100 simulations with $N(0,1)$ data. Squares indicate B, octagons indicate UCV, and triangles indicate BCV2. The solid line indicates the optimal MISE bandwidth values for Normal data.

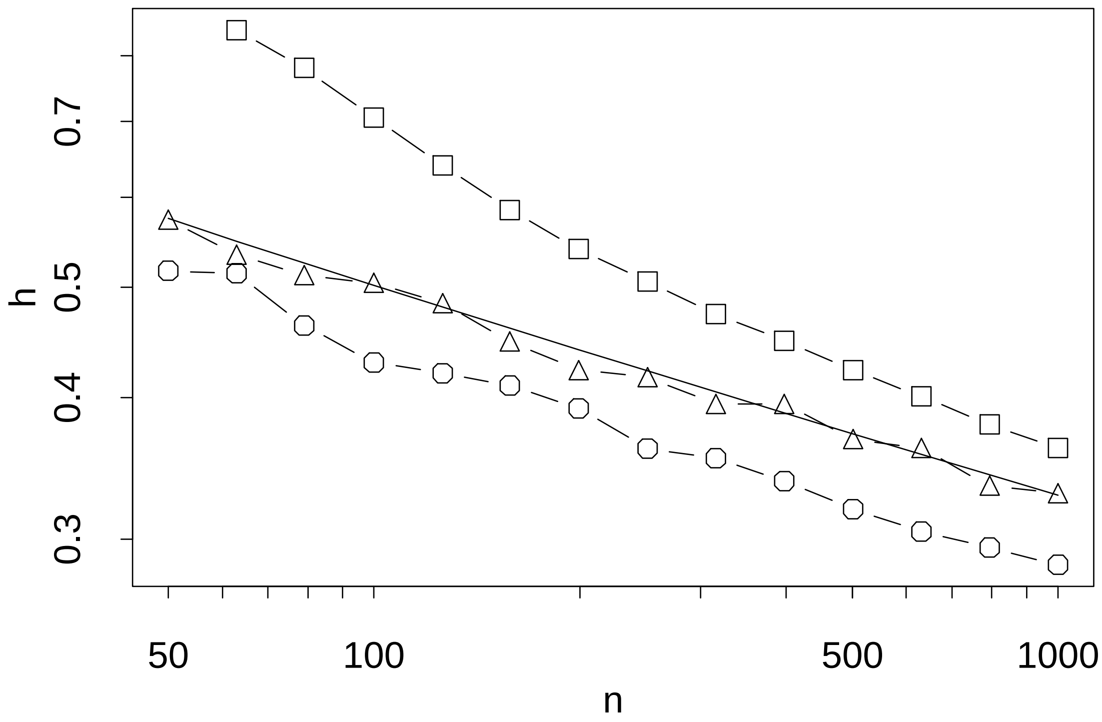

Figure 2: Bivariate version of Figure 1 with $N(\mathbf{0}, \mathbf{I}_2)$ data.

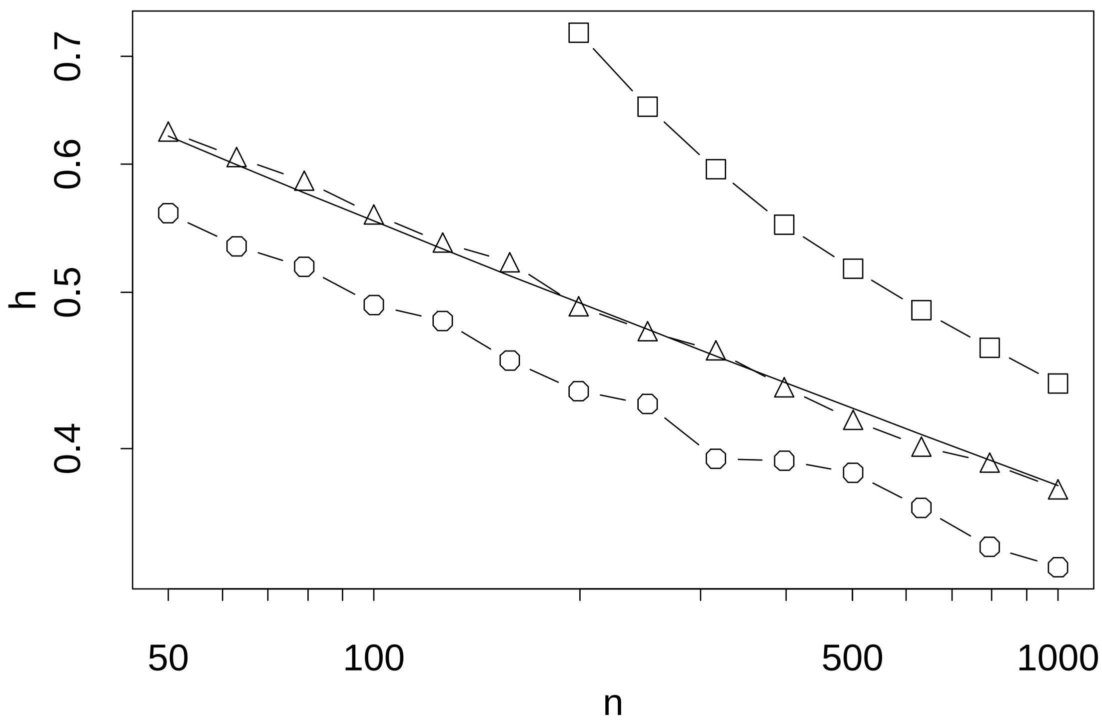

Figure 3: Trivariate version of Figure 1 with $N(\mathbf{0}, \mathbf{I}_3)$ data.

larger than the optimal MISE values. Furthermore, for smaller sample sizes in each dimension ($n < 100$ for $d = 2$ and $n < 250$ for $d = 3$), the optimizer was not able to find a minimum more than 50 percent of the time; the criterion was monotone decreasing in these cases. Figure 4 shows histograms of data-based BCV2, UCV, and B bandwidths taken from the simulation study with sample size 500. The dashed lines represent the optimal MISE values ($h_{\mathrm{MISE}} = 0.315, 0.372$, and 0.424, for $d = 1, 2$, and 3 respectively) computed from Worton's formula given in Equation (18). For the B values, a small positive bias is noted as d increases.

For small samples, the means of the unbiased cross-validation bandwidths are closer to the optimal MISE values than those obtained by the bootstrap. However, these means are smaller than the optimal MISE value. This is due to the UCV function's attempt to capture apparent structure or chance multimodality in the sample, although none truly exists in the underlying density; see Figure 4. Note that there is a large grouping around the exact value h_{MISE}, yet a significant fraction of the values are considerably smaller than the optimal h_{MISE}. This tendency is diminished in higher dimensions.

The new biased cross-validation method, BCV2, performs remarkably well, with the mean bandwidth for each sample size and dimension being very close to the optimal MISE values. The better approximation suggested by the asymptotic results is apparent in Figure 4.

The actual sample standard deviations of the estimated smoothing parameters are given in Table 3 for sample sizes $n = 50, 100, 500$, and 1000. The notation 'NA' is used in the table to refer to the cases when the optimizer was unable to find a minimizer at that sample size. Note that this only occurs with the bootstrap. When the optimizer was able to find a minimum more than 50 percent of the time, the bootstrap estimates had standard deviations that were considerably smaller than those generated by either cross-validation algorithm, regardless of the dimension of the data.

Contrary to the theoretical results in the univariate and bivariate settings, the biased cross-validation algorithm, BCV2, yielded standard deviations slightly lower than the unbiased version, due in part to the unbiased cross-validation algorithm's hunt for nonexistent structure. As the dimensionality of the data increases, the difference between the

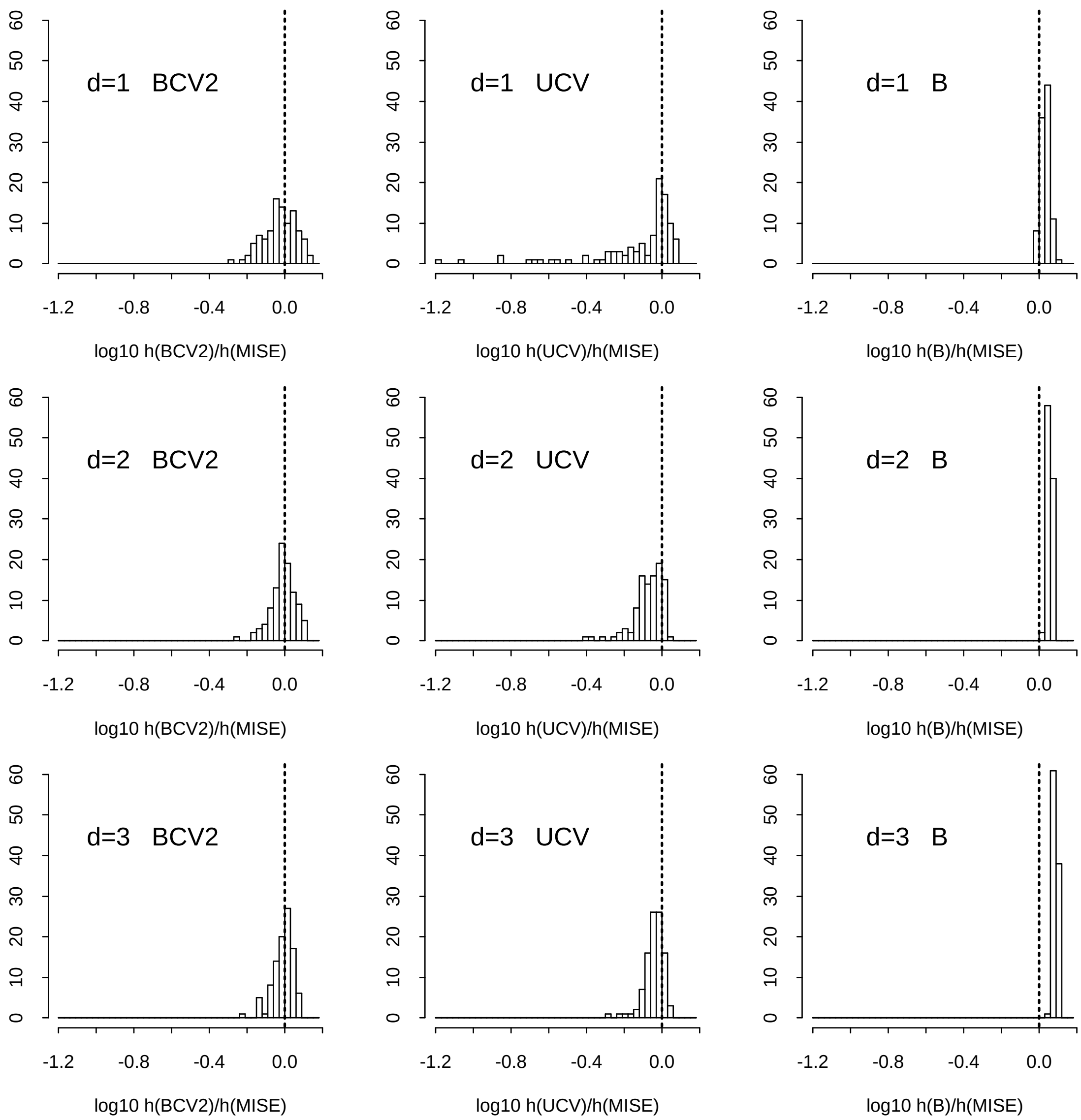

Figure 4: Histograms of the 100 optimal smoothing parameters from samples of size 500 used in Figures 1,2, and 3 (divided by h_{MISE} on a log scale). The dashed line shows the location of the optimal MISE smoothing parameter.

Table 3: Sample standard deviations of estimated cross-validation smoothing parameters. The theoretical values of $\sigma(\hat{h}-h_0)$ given in the last column are for $n = 1000$.

		n = 50	n = 100	n = 500	n=1000	$\sigma(\hat{h}-h_0)$
	B	0.090	0.0517	0.0174	0.0116	0.0047
d = 1	UCV	0.197	0.1243	0.0922	0.0669	0.0452
	BCV2	0.137	0.0836	0.0576	0.0507	0.0542
	B	NA	0.0506	0.0132	0.0082	0.0024
d = 2	UCV	0.150	0.1191	0.0529	0.0449	0.0243
	BCV2	0.116	0.1069	0.0525	0.0391	0.0354
	B	NA	NA	0.0124	0.0088	0.0015
d = 3	UCV	0.110	0.0846	0.0425	0.0318	0.0159
	BCV2	0.137	0.1048	0.0499	0.0352	0.0220

standard deviations of the two cross-validation methods diminishes, with the unbiased cross-validation method having a slightly lower standard deviation. Note that UCV and B perform poorly relative to the asymptotic theory; see the last two columns of Table 3.

A second study was done to examine the performance of the three algorithms with bimodal data. The results are summarized in Table 4. A bimodal density has often proved difficult for BCV1 and plug-in algorithms which tend to oversmooth. (For the plug-in algorithms, the large variance of the data leads to too large a choice of the auxiliary smoothing parameter, leading to oversmoothing.) Samples of various sizes were taken from a bimodal Normal mixture density, $f(\mathbf{x}) = \frac{1}{2}N(\boldsymbol{\mu}_1, \mathbf{I}_d)+\frac{1}{2}N(\boldsymbol{\mu}_2, \mathbf{I}_d)$, where $\boldsymbol{\mu}_1 = (2,0,\ldots,0)^t$ and $\boldsymbol{\mu}_2 = -\boldsymbol{\mu}_1$, so that the distance between the means was 4 regardless of dimension. Optimal values of the smoothing parameters were chosen as the global minima on the range of h from $n^{-1/(d+4)}/8$ to $n^{-1/(d+4)} \times 8$, where n was the sample size.

Figure 5 shows histograms of the global minima of the criterion functions for the three methods using the same 100 samples of size 500 in each dimension. Marron and Wand (1992) give an expression for calculating the univariate MISE for kernel estimation of Normal mixture densities using a Normal kernel. For the Normal mixture density used in this simulation, the value of h that minimizes the MISE for $n = 500$ is given in Table 4. The dashed line in the univariate histograms represents this optimal value. In 2 and 3 dimensions, the dashed lines indicate $h_{\text{AMISE}} = 0.397$ and 0.440, respectively.

Table 4: Summary of results for bimodal simulation studies.

		Mean of $\hat{h}$		Std. Dev. of $\hat{h}$	
		n = 200	n = 500	n = 200	n=500
d = 1	B	1.587	1.053	0.0922	0.5617
	UCV	0.353	0.285	0.1170	0.1037
	BCV2	0.433	0.339	0.0901	0.0563
	MISE	0.440	0.359		
	AMISE	0.412	0.343		
d = 2	B	1.603	1.028	0.1017	0.5616
	UCV	0.367	0.283	0.1287	0.1034
	BCV2	0.505	0.419	0.0752	0.0543
	AMISE	0.463	0.397		
d = 3	B	1.604	1.063	0.0913	0.5551
	UCV	0.346	0.287	0.1387	0.0882
	BCV2	0.559	0.476	0.0816	0.0472
	AMISE	0.502	0.440		

The biased cross-validation method, BCV2, performed very well, yielding estimated smoothing parameters close to the optimal values. Although most of the least-squares cross-validation smoothing parameters are centered about the optimal values, a substantial fraction are considerably smaller. In contrast to the unimodal case, this problem persists as the dimension increases. The bootstrap method produced global minima split almost equally between the values h_{AMISE} and $3h_{\text{AMISE}}$ in each dimension. The bimodal nature of the data apparently 'confuses' the bootstrap algorithm, resulting in significantly larger bandwidths (on average) and much larger variability. However, while UCV tends to undersmooth, B tends to oversmooth.

Figure 6 shows plots of the criterion function curves for 5 samples of size 500 taken from the same bimodal Normal mixture density given above. Note that the function values have been rescaled for ease of comparison. This rescaling has no effect on the position of the local minima. The new biased cross-validation algorithm gives clear minima near the optimal MISE value of $h_{\text{MISE}} = 0.359$. The curves generated with the unbiased cross-validation algorithm show minima, yet the neighborhood about the minima is fairly flat. Note that two of the minima are significantly smaller than the optimal value even at this sample size. The bootstrap method consistently yields risk curves with two local minima

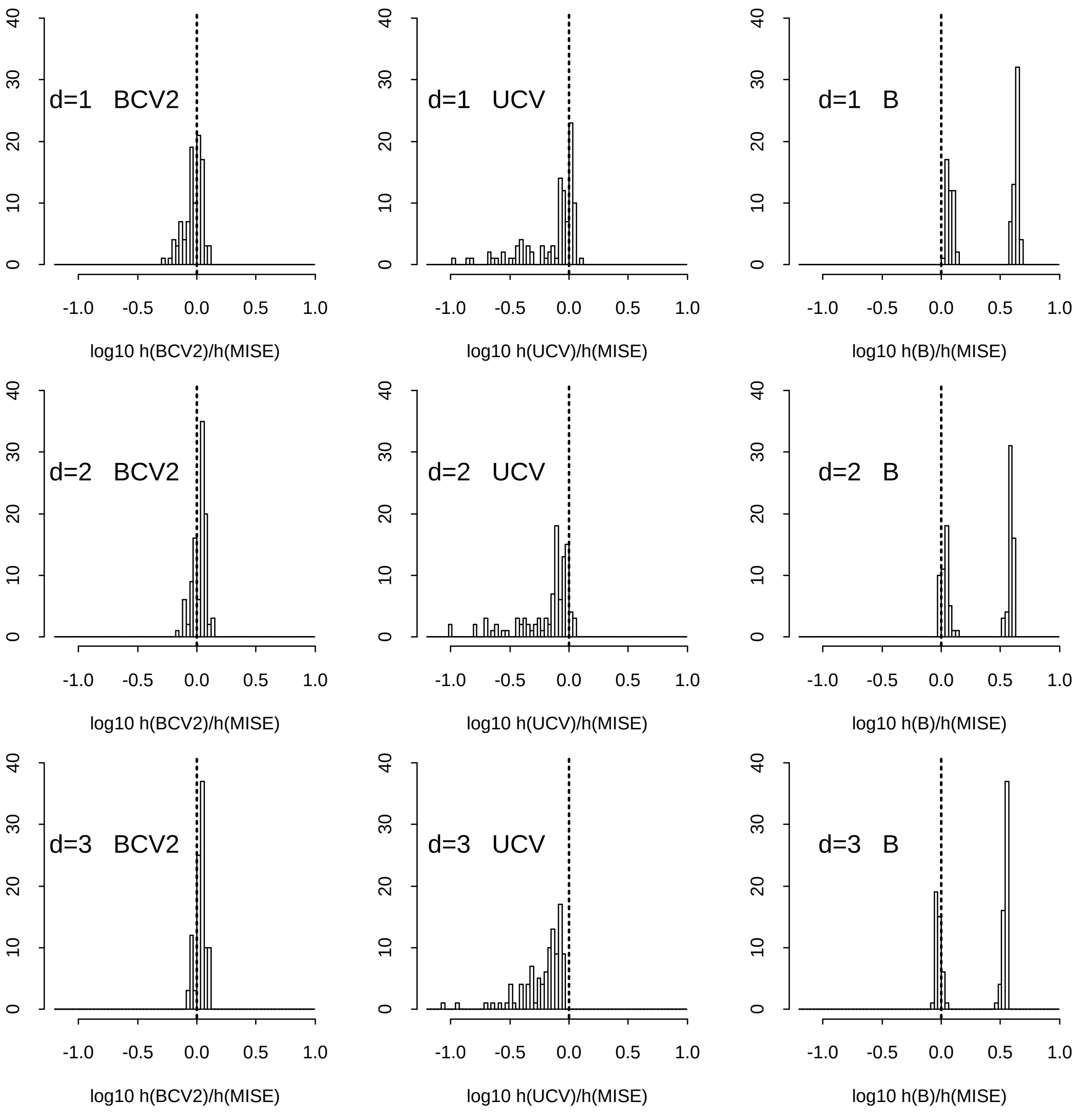

Figure 5: Histograms of 100 local minima for BCV2, UCV, and B for data sets of size 500 sampled from a bimodal density in dimensions 1, 2, and 3. The bandwidths are divided by h_{MISE} and plotted on a $\log_{10}$ scale. The dashed lines represent h_{MISE}.

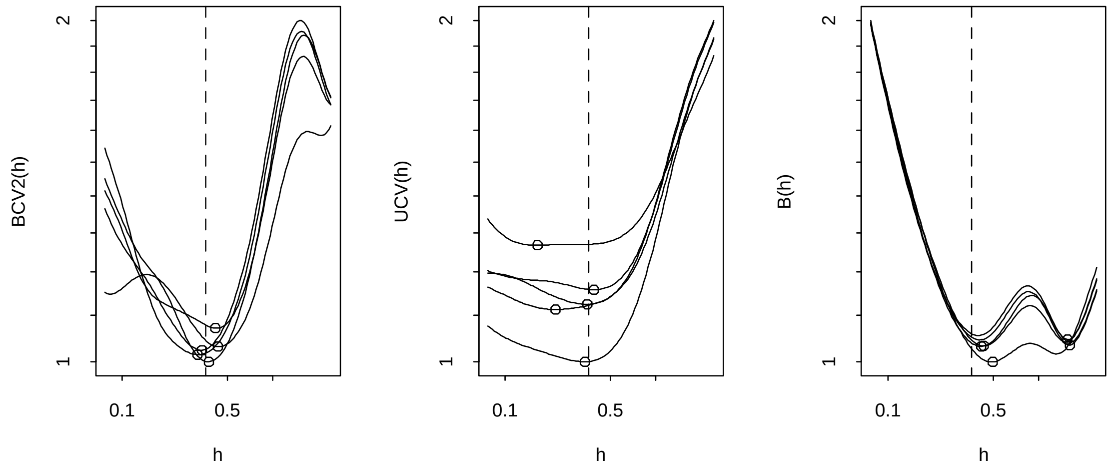

Figure 6: Plots of rescaled BCV2, UCV, and B curves for five data sets of size 500 sampled from a univariate bimodal Normal mixture. Circles indicate global minima. The dashed lines indicate the optimal MISE value.

for these data. The smaller local minimum corresponds to this method's attempts to find the bimodality in the data, while the larger local minimum corresponds to a value of the smoothing parameter that would make the estimate almost unimodal. In fact, in the 100 simulations with $n = 500$ and $d = 2$ summarized in Table 4, 46% of the global minima were centered around 0.426 with a standard deviation of 0.0312. The other 54% were centered around 1.541 with a standard deviation of 0.0780. Note that the standard deviations about these modes are much smaller than the standard deviation of the global minima given in Table 4.

7 Case Studies

The first data set examined is 320 lipid (cholesterol and triglyceride) levels of men with heart disease from Scott et al. (1978). The authors found 2 primary clusters in these data. The data were log transformed to reduce skewness and then standardized to obtain equal variances in each dimension.

Figure 7 shows a surface representing the values of the bootstrap criterion function, $\mathrm{B}(h_1, h_2)$. In this case, a local minima of the surface was found at $h_1 = 0.471$ and

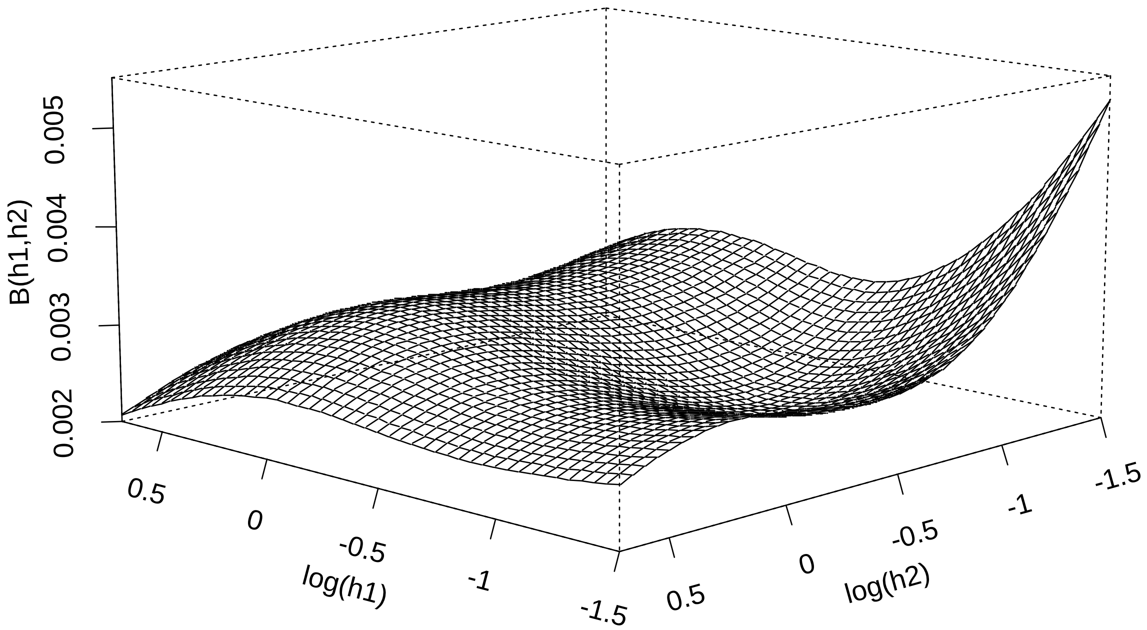

Figure 7: $\mathrm{B}(h_1, h_2)$ surface for lipid data.

$h_2 = 0.485$. The Normal product kernel estimate of this data set using these values of the smoothing parameters is shown in Figure 8. Note that this estimate is considerably oversmoothed, making it hard to identify the bimodality discussed in Scott et al. (1978). Using the same data, the unbiased cross-validation method gives smoothing parameter values of $h_1 = 0.341$ and $h_2 = 0.180$. The biased cross-validation method yields smoothing parameters of $h_1 = 0.373$ and $h_2 = 0.174$. In this case, the two cross-validation algorithms produce very similar density estimates. The product kernel estimate using the biased cross-validation values is shown in Figure 9. The bimodality is clearly present. Note that for a standard Normal data set of this size, the optimal MISE values of the smoothing parameters are $h_i = 0.403$, $i = 1, 2$.

The second data set is much larger, with 6,541 data points representing hourly measurements of ozone and temperature gathered during 1989 in Houston (see Ensor, Sain, and Scott (1992)). Random uniform noise was added to the raw data to account for rounding. A power transformation of the form $y \leftarrow x^\gamma$, where $\gamma = 1/2$ for ozone and $\gamma = 2$ for temperature, was then applied to significantly reduce skewness, and, finally, the data were

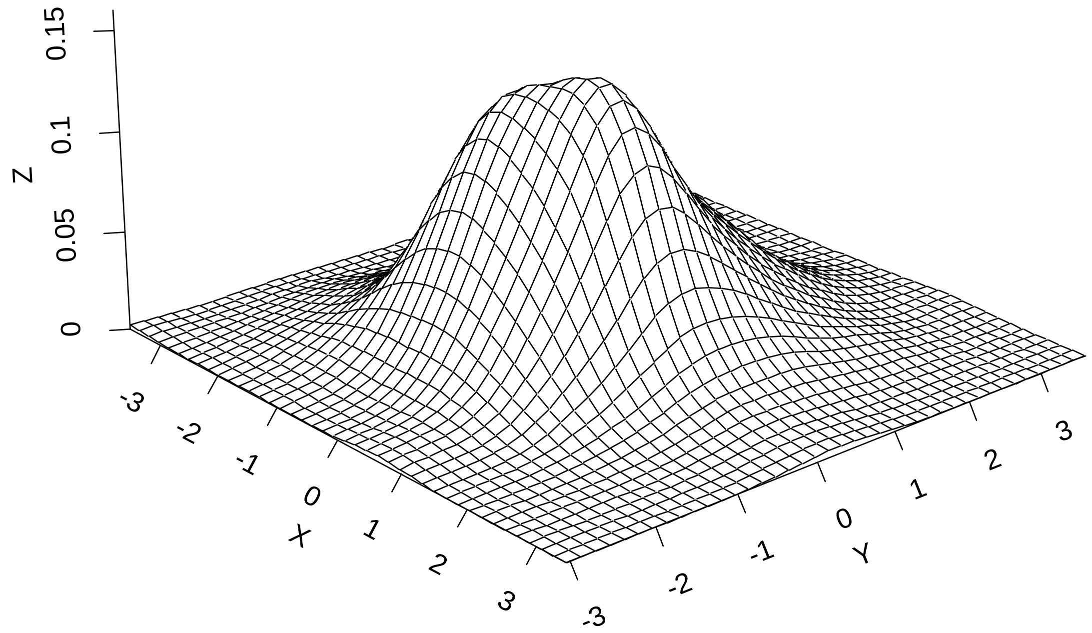

Figure 8: Kernel estimate of lipid data using bootstrap estimates of smoothing parameters.

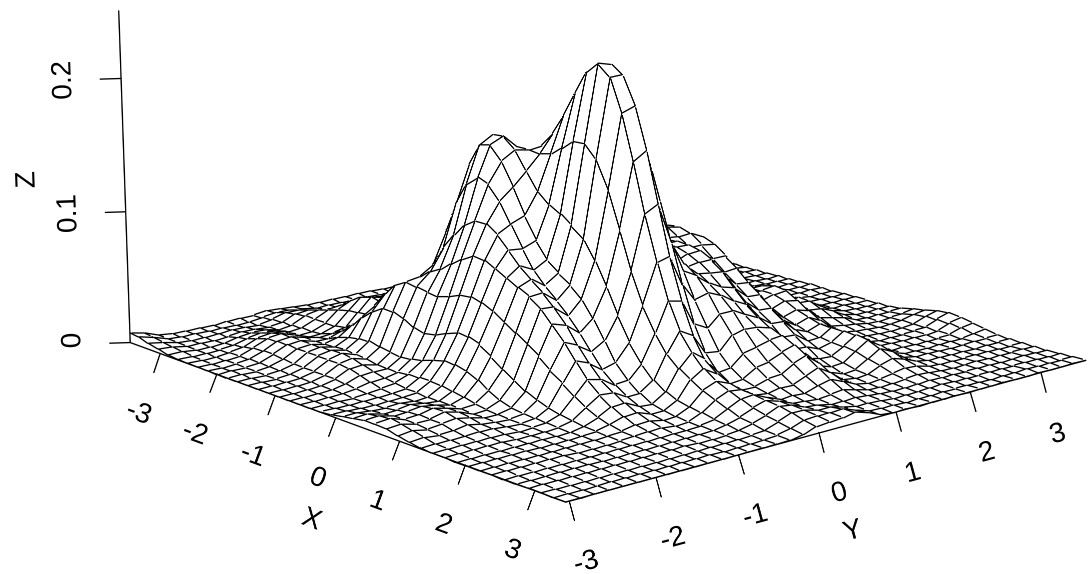

Figure 9: Kernel estimate of lipid data using BCV2 estimates of smoothing parameters.

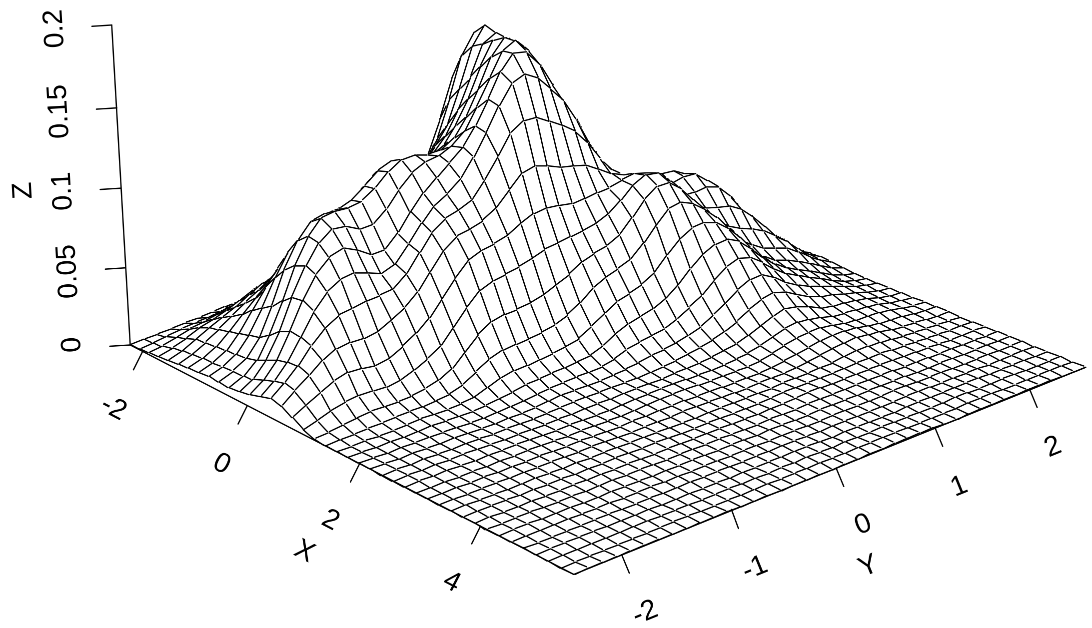

Figure 10: Kernel estimate of ozone data using bootstrap estimates of smoothing parameters.

standardized to equalize the variances.

In this case, the three algorithms all produced quite different smoothing parameters. Figure 10 shows the kernel estimate of the ozone data with the optimal bootstrap smoothing parameters of $h_1 = 0.250$ and $h_2 = 0.183$. Figure 11 shows the kernel estimate with the optimal UCV smoothing parameters of $h_1 = 0.101$ and $h_2 = 0.0530$. Figure 12 shows the kernel estimate with the optimal BCV2 smoothing parameters of $h_1 = 0.248$ and $h_2 = 0.0952$. Of the three algorithms, the biased cross-validation method yields the preferred estimates of the smoothing parameters. The bootstrap values give an estimate that is somewhat oversmoothed, while the unbiased cross-validation algorithm yields values that undersmooth the estimate. At this point, it is useful to recall that 'oversmoothed' is a relative term. An independent observer, viewing only the bootstrap estimate, might think that it was too rough. The UCV and BCV estimates are needed to provide an appropriate scale reference. Note that for a standard Normal data set of this size, the optimal MISE values of the smoothing parameters are $h_i = 0.235$, $i = 1, 2$.

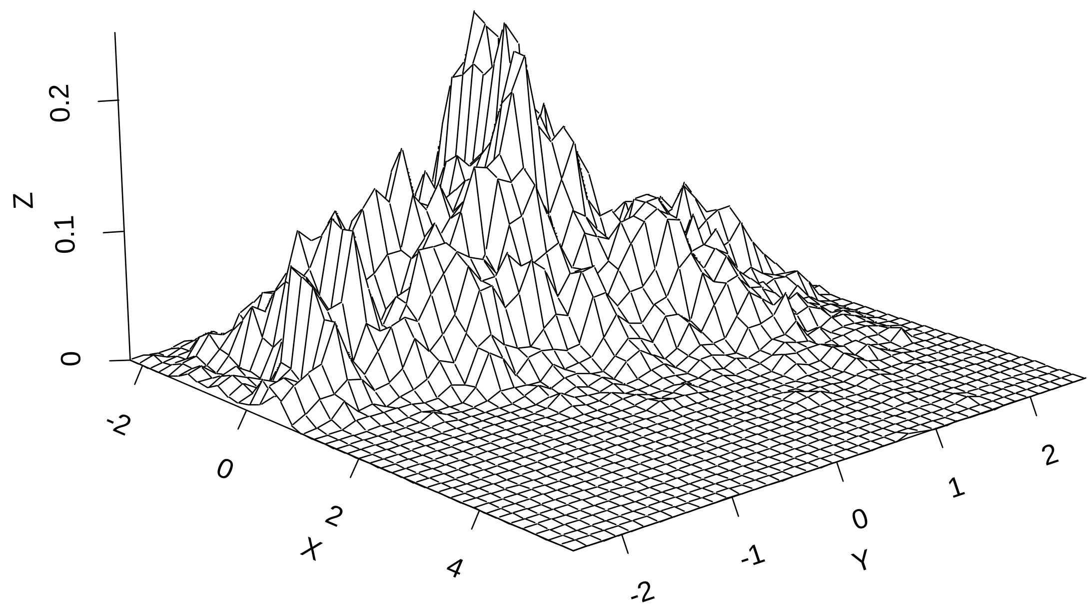

Figure 11: Kernel estimate of ozone data using UCV estimates of smoothing parameters.

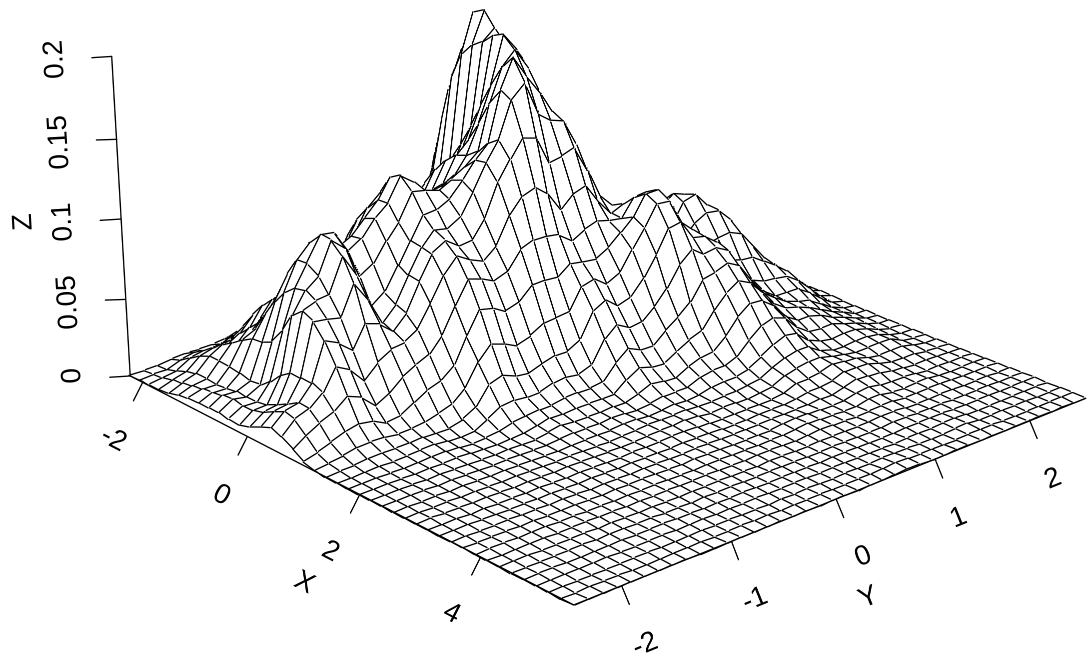

Figure 12: Kernel estimate of ozone data using BCV2 estimates of smoothing parameters.

8 Discussion

Of all the methods studied here, BCV2 seems to perform the best. Simulations show that this method produces values very close to the ideal with reasonable variability. The least-squares cross-validation method also produces values close to the ideal, yet often times it gives values of smoothing parameters that attempt to show structure in the data when that structure does not actually exist. This results in estimates that are seriously undersmoothed. The bootstrap, although by far the least variable of the three (with unimodal data), gives values of the smoothing parameters that clearly oversmooth. Also, it has problems with finding minima in small data sets. We expected UCV and BCV2 to behave similarly. The fact that BCV2 does not exhibit a tendency to choose very small bandwidths may be due to the 'rigid' integrated variance approximation $R(K)/nh$ used in Equation (8). In our experience, UCV exhibits a propensity to choose very small bandwidths even with thousands of data points. It is a limitation of the asymptotic theory that the behavior of UCV(h) away from h_0 is poorly understood.

Another minor point in favor of the new biased cross-validation algorithm is computing time. One iteration of the bootstrap function takes nearly one and a half times as long as an iteration of the unbiased cross-validation function, and nearly three times as long as an iteration of the new biased cross-validation function; a direct consequence of the number of exponentials that must be calculated in the double sums.

In spite of the promising theoretical results, there is no doubt that the problem of finding appropriate smoothing parameters in kernel density estimation grows in complexity as the dimensionality of the data increases, especially when the data are nearly rank deficient. Much less variable cross-validation algorithms can be constructed using plug-in methods (see Wand and Jones (1992)), which use estimates such as that in Equation (10), but with auxiliary smoothing parameters. However, the technology for choosing these auxiliary smoothing parameters is as yet immature. In practice, several methods can and should be employed to give a flavor of the estimates, from which the best (or consensus) estimate can be chosen. We do suggest that BCV2 be one of these methods.

Appendix

Assume that K is symmetric and that K and f are sufficiently differentiable. The univariate BCV2 bandwidth $\hat{h}$ is a root of

$$\frac{\partial}{\partial h}\mathrm{BCV2}(h) = \frac{-R(K)}{nh^2} - \frac{\sigma_K^4 \hat{\theta}_h}{4h} \tag{19}$$

where

$$\hat{\theta}_h = \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i}\left\{K^{iv}\left(\frac{x_i - x_j}{h}\right) + \frac{x_i - x_j}{h}K^{v}\left(\frac{x_i - x_j}{h}\right)\right\} \tag{20}$$

Making the change of variables $u = (x_i - x_j)/h$,

$$E\hat{\theta}_h = \iint [K^{iv}(u) + uK^{v}(u)] f(x) f(x - hu)\, du\, dx.$$

Plugging in the Taylor's series

$$f(x - hu) = f(x) - huf'(x) + \cdots + \frac{1}{24}h^4 u^4 f^{iv}(x) + \cdots$$

and noting that $\int u^4 K^{iv}(u) = 24$ and $\int u^5 K^{v}(u) = -120$ are the first nonzero terms when integrating over u gives

$$E\hat{\theta}_h = -4h^4 R(f'') + O(h^6).$$

Next, rewriting $\hat{\theta}_h$ as $[n(n-1)h]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\gamma_{ij}$ in Equation (20),

$$\begin{aligned} Var(\hat{\theta}_h) &= [n(n-1)h]^{-2} Cov\left\{\sum_{i=1}^{n}\sum_{j\neq i}\gamma_{ij}, \sum_{i'=1}^{n}\sum_{j'\neq i'}\gamma_{i'j'}\right\} \\ &= [n(n-1)h]^{-2}\left\{2n(n-1)Var(\gamma_{ij}) + 4n(n-1)(n-2)Cov(\gamma_{ij}, \gamma_{ij'})\right\}, \end{aligned}$$

where $i \neq j \neq j'$. Now $E\gamma_{ij} \approx -4h^5 R(f'')$ from above, and $Cov(\gamma_{ij}, \gamma_{ij'}) = O(h^{10})$. Thus $Var(\hat{\theta}_h)$ is dominated by $E\gamma_{ij}^2$, which may be approximated by

$$\begin{aligned} E\gamma_{ij}^2 &= h\iint [K^{iv}(u)^2 + 2uK^{iv}(u)K^{v}(u) + u^2 K^{v}(u)^2] f(x) f(x - hu)\, du\, dx \\ &\approx h\int f(x)^2 dx \int [K^{iv}(u)^2 + 2uK^{iv}(u)K^{v}(u) + u^2 K^{v}(u)^2]\, du \\ &= hR(f)R[uK^{v}(u)] \end{aligned}$$

as $\int 2uK^{iv}(u)K^{v}(u)du = -\int K^{iv}(u)^2 du$. Therefore,

$$Var(\hat{\theta}_h) \approx 2R(f)R[uK^v(u)]/(n^2h),$$

and

$$\hat{\theta}_h \approx -4h^4R(f'') + Z\sqrt{2R(f)R[uK^v(u)]/(n^2h)},$$

where $Z \sim N(0,1)$. The Normality follows from Hall (1984); see Scott and Terrell (1987) and Jones and Kappenman (1992). The determining Equation (19) becomes

$$\frac{\partial}{\partial h}\mathrm{BCV2}(h) = \frac{-R(K)}{nh^2} + \sigma_K^4 h^3 R(f'') - Z\frac{\sigma_K^4\sqrt{2R(f)R[uK^v(u)]}}{4nh^{3/2}}. \tag{21}$$

Recalling from (7) that $h_0^5 = R(K)/[n\sigma_K^4 R(f'')]$, and choosing the root $\hat{h}$ in (21) to be of the form

$$\hat{h}_{\mathrm{BCV2}} \approx h_0 + \alpha_1 n^{-\alpha_2} Z, \tag{22}$$

straightforward perturbation arguments yield $\alpha_2 = 3/10$ and

$$\alpha_1 = \frac{\sqrt{2}}{20}R(f)^{1/2}R(f'')^{-3/10}R(K)^{-7/10}\sigma_K^{14/5}R[uK^v(u)]^{1/2}.$$

Hence from (22),

$$n^{3/10}(\hat{h}_{\mathrm{BCV2}} - h_0) \xrightarrow{\mathcal{D}} N(0, \alpha_1^2),$$

which completes the demonstration of Equation (11).

References

Bowman, A.W. (1984), "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates," *Biometrika*, **71**, 353-360.

Chiu, S.T. (1992), "A Comparative Study of Bandwidth Selectors," Technical Report, Department of Statistics, Colorado State University.

Ensor, K.B., Sain, S.R., and Scott, D.W. (1992), "Ozone and Air Quality in Houston, 1980-1990," Technical Report, Department of Statistics, Rice University.

Faraway, J.J. and Jhun, M. (1990), "Bootstrap Choice of Bandwidth for Density Estimation," *Journal of the American Statistical Association*, **85**, 1119-1122.

Hall, P. (1984), "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, **14**, 1-16.

Hall, P. (1990), "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems," *Journal of Multivariate Analysis*, **32**, 177-203.

Hall, P. and Marron, J.S. (1987), "Estimation of Integrated Squared Density Derivatives," *Statistics and Probability Letters*, **6**, 109-115.

Hall, P., Marron, J.S., and Park, B.U. (1992), "Smoothed Cross-Validation," *Probability Theory and Related Fields*, in press.

Hall, P., Sheather, S., Jones, M.C., and Marron, J.S. (1991), "On Optimal Data-Based Bandwidth Selection in Kernel Density Estimation," *Biometrika*, **78**, 263-269.

Jones, M.C. and Kappenman, R.F. (1992), "On a Class of Kernel Density Estimate Bandwidth Selectors," *Scandinavian Journal of Statistics*, in press.

Marron, J.S. and Wand, M.P. (1992), "Exact Mean Integrated Squared Error," *The Annals of Statistics*, in press.

Rudemo, M. (1982), "Empirical Choice of Histogram and Kernel Density Estimators," *Scandinavian Journal of Statistics*, **9**, 65-78.

Scott, D.W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York: John Wiley.

Scott, D.W., Gotto, A.M., Cole, J.S., and Gorry, G.A. (1978), "Plasma Lipids as Collateral Risk Factors in Coronary Heart Disease – A Study of 371 Males with Chest Pain," *Journal of Chronic Diseases*, **31**, 337-345.

Scott, D.W. and Terrell, G.R. (1987), "Biased and Unbiased Cross-Validation in Density Estimation," *Journal of the American Statistical Association*, **82**, 1131-1146.

Sheather, S. and Jones, M.C. (1991), "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society B*, **53**, 683-690.

Statistical Sciences, Inc. (1991), *S-Plus Reference Manual*, Seattle: Statistical Sciences, Inc.

Stone, C.J. (1984), "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," *The Annals of Statistics*, **12**, 1285-1297.

Taylor, C.C. (1989), "Bootstrap Choice of the Smoothing Parameter in Kernel Density Estimation," *Biometrika*, **76**, 705-712.

Terrell, G.R. (1990), "The Maximal Smoothing Principle in Density Estimation," *Journal of the American Statistical Association*, **85**, 470-477.

Terrell, G.R. and Scott, D.W. (1985), "Oversmoothed Nonparametric Density Estimates," *Journal of the American Statistical Association*, **80**, 209-214.

Wand, M.P. and Jones, M.C. (1992), "Multivariate Plug-In Bandwidth Selection," Technical report, Department of Statistics, Rice University.

Worton, B.J. (1989), "Optimal Smoothing Parameters for Multivariate Fixed and Adaptive Kernel Methods," *Journal of Statistical Computation and Simulation*, **32**, 45-57.