Cross Validation Bandwidth Selection for Derivatives of Multidimensional Densities

Matthew Baird

RAND Labor & Population

WR-1060

October 2014

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

Cross Validation Bandwidth Selection for Derivatives of Multidimensional Densities

Matthew Baird
The RAND Corporation*

October 2, 2014

Abstract

Little attention has been given to the effect of higher order kernels for bandwidth selection for multidimensional derivatives of densities. This paper investigates the extension of cross validation methods to higher dimensions for the derivative of an unconditional joint density. I present and derive different cross validation criteria for arbitrary kernel order and density dimension, and show consistency of the estimator. Doing a Monte Carlo simulation study for various orders of kernels in the Gaussian family and additionally comparing a weighted integrated square error criterion, I find that higher order kernels become increasingly important as the dimension of the distribution increases. I find that standard cross validation selectors generally outperform the weighted integrated square error cross validation criteria. Using the infinite order Dirichlet kernel tends to have the best results.

*I would like to thank Rosa Matzkin for her financial support and research help on this project, as well as UCLA Econometrics Pro-Seminar participants for helpful advice, including Jinyong Hahn and Conan Snider

1 Introduction

There are many cases when a researcher is interested in information about the shape of a distribution of random variables. Methods are well developed for nonparametric estimation of the density created by the underlying data generating process, including methods for selecting the bandwidth parameters for kernel density estimators.¹ There is also substantial progress in developing techniques for estimating derivatives of distributions. More sophisticated methods have been developed for choosing bandwidths for multivariate densities (Zhang et al., 2005) and for derivatives of univariate densities (Wu 1997, Bearnse and Rilstone 2008), but little work has been done on densities of multivariate densities. In this paper, I demonstrate consistency of the multidimensional derivative of density estimators, derive cross validation criteria, and investigate how well product kernels perform in estimating high dimension derivatives of unconditional densities.

I test various orders of Gaussian kernels with an extension of the univariate cross validation criteria as well as with an additional cross validation criterion: a weighted integrated square error criterion. The standard cross validation is more robust against poor bandwidth selection than the weighted integrated square error criterion with regards both to a mean square error (MSE) criterion and a maximum deviation criterion.

Throughout the paper, I give my attention to product kernels for their ease of use. Cacoullos (1966) first presented product kernels as an option in estimating multivariate densities. This disallows potentially more accurate kernels; however, the product kernel is the easiest to use and to derive cross validation criteria for, and is frequently used in practice. The orders of the kernel (defined as the first non-zero moment of the kernel) used are 2,4,6,8,10, and the infinite order (the Dirichlet kernel). This paper investigates only cross validation methods instead of plug-in methods, as plug-in methods become increasingly difficult to formulate with higher dimension kernels, necessitating solving potentially intractable systems of equations. Further, Loader (1999) challenges the previous work that suggested the superiority of plug-in methods over cross validation, and demonstrates that this is not true in many cases (e.g., when there is misspecification of the pilot bandwidths).

Hansen (2005) and Turlach (1993) find that kernel order and bandwidth choice (respectively) are more important than choice of kernel family. For that reason, similar to the work of Hansen and Wand and Schucany (1990), I restrict attention to different orders of the Gaussian kernel. It is not difficult to take the generalized criteria presented here and adapt them for use on a different kernel family. Marron (1994) shows that higher order kernels perform well when the curvature of what is being estimated is roughly constant, and poorly when there are abrupt changes in curvature on neighborhoods about the size of the bandwidth. Wand and Schucany (1990) examine Gaussian kernels from orders 2 to 10; they compare the efficiencies of these kernels theoretically to the optimal kernels, and show that for low order kernels, they are very close, and for ν derivative low (i.e., the zeroth derivative). The worst case they present for the first derivative has a relative efficiency of 0.76. Marron and Wand (1992) also show that the bandwidth that minimizes MISE is close to that which minimizes

¹Li and Racine (2007) provide an excellent review.

AMISE (the plug-in estimator) only for sample sizes close to 1 million, discouraging use of plug-in methods.

The cross validation methods are generalizations of the cross validation criteria set forth by Hardle, Marron and Wand (1990) for the univariate density derivative. They demonstrate that, for the univariate case and the first derivative, there is not much loss in efficiency (in the sense that Silverman uses the term) from using the Gaussian kernel instead of the optimal kernel. Various research connect to derivations of consistency and optimal convergence. Marron and Hardle (1986) generalizes the procedures for a variety of nonparametric estimators, including density estimators. Li and Racine (2007) present numerous derivations.

Chacon, Duong, and Wand (2011) investigate derivatives of multidimensional densities, just as in this paper, but focus their attention just on second order kernels. They derive the MISE and show convergence rates. They also allow for a bandwidth matrix, and show that this general bandwidth matrix generally achieves better simulation results than that of using a diagonal matrix (as is done in product kernels, and so in this paper).

Section 2 presents the direct cross validation criterion as well as a weighted cross validation criterion, and demonstrates consistency of the multivariate kernel estimator. Section 3 presents simulation results. Section 4 concludes. Proof and criteria derivations, along with further results tables, are in the Appendix in Section 5.

2 Bandwidth Selection for the Derivative of a Joint Density

Assume we are interested in estimating the derivative of a q^{th} -dimensional density, for arbitrary dimensionality $q \geq 1$. Let G_x be the set of bandwidths associated with variables x . Given the estimator

$$\hat{f}(x) = \frac{1}{n \prod_{x \in G_x} h_s} \sum_{i=1}^n \prod_{s \in G_x} K\left(\frac{x_s - x_{is}}{h_s}\right)$$

Then, the estimator for the derivative of the density is

$$\frac{\partial \hat{f}(x)}{\partial x_k} = \frac{1}{n h_k \prod_{x \in G_x} h_s} \sum_{i=1}^n K'\left(\frac{x_k - x_{ik}}{h_k}\right) \prod_{s \in G_x \setminus k} K\left(\frac{x_s - x_{is}}{h_s}\right)$$

I first demonstrate conditions for consistent estimation of the derivative of the density, and then derive two cross validation criteria connected to the integrated square error, and show that these criteria provide bandwidth estimators that satisfy the consistency criteria. I then perform a Monte Carlo study to compare the performance of different criteria and kernel orders across different data sizes and dimensions.

2.1 Consistency

Theorem 1. *Consider the estimation of the derivative of a q^{th} dimensional density using an r^{th} order kernel $K(\cdot)$. If the following conditions hold*

1. *The kernel integrates to 1*
2. *The kernel is symmetric*
3. *The $r^{th}-1$ moment of the kernel is 0*
4. *The r^{th} moment of the kernel is positive and finite*
5. *$f(x)$ is $(r+1)$ times differentiable*

Then,

$$MSE\left(\frac{\partial \hat{f}(x)}{\partial x_k}\right) = O\left(\left(\sum_{t=1}^q h_t^{r+1}\right)^2 + \frac{1}{nh_k \prod_{s \in G_x} h_s}\right)$$

Further, if the bandwidths converge to zero as the sample size goes to infinity, but more slowly than the sample size increases, then the last statement directly implies that $\frac{\partial \hat{f}(x)}{\partial x_k} \rightarrow \frac{\partial f(x)}{\partial x_k}$ in mean square error (MSE), implying also convergence in probability, and consistency.

The proof of this theorem is contained in Section 5.1.1.

2.2 Cross Validation Criteria: Integrated Square Error

For the standard cross validation method, the criterion minimizes the integrated difference between the true and the estimated densities. This criterion is the same for any dimension of the density. Let $f(x)$ be the true density, and $\hat{f}(x)$ be the product kernel estimator of $f(x)$. Then the cross validation criterion (Integrated Square Error, or ISE) is given by

$$ISE(h) = \int \cdots \int \left(\frac{\partial \hat{f}(x)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right)^2 dx$$

Section 5.1.2 provides the derivation for the cross validation: for given dimension, I express a criterion which yields equivalent minimizing arguments and is not a function of any unknowns. This is given by

$$ISE^*(h) = \frac{1}{n^2 h_k^2 \prod_{s \in G_x} h_s} \sum_{i=1}^n \sum_{j=1}^n \left[\int K'(\tilde{x}_{ijs} + \bar{x}_{is}) K'(\bar{x}_{is}) d\bar{x}_{is} \right] \times \cdots \\ \prod_{s \in G_x \setminus k} \left(\int K(\tilde{x}_{ijs} + \bar{x}_{is}) K(\bar{x}_{is}) d\bar{x}_{is} \right) + \frac{2 \sum_{i=1}^n \sum_{j \neq i} K''(\tilde{x}_{ijk}) \prod_{s \in G_x \setminus k} K(\tilde{x}_{ijs})}{n(n-1) \left(\prod_{s \in G_x} h_s \right)^2 h_k^2}$$

where $x = (x_1, \dots, x_K)' \in G_x$ is the point of evaluation, and $x_i = (x_{i1}, \dots, x_{iK})'$ is one observation in the data. h_s is the bandwidth for the s^{th} random variable, and $K(\cdot)$ is the kernel chosen for the estimation procedure.

2.3 Cross Validation Criteria: Weighted Integrate Square Error

I propose an alternative criterion which weights the difference between the true and the estimated derivatives of the densities. The weight is provided by an estimate of the density. The criterion is

$$WISE(h) = \int \dots \int \left(\frac{\partial \hat{f}(x; h)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right)^2 \hat{f}(x; b) dx$$

where b (and equivalently $\hat{f}(x; b)$) is estimated prior to searching for the optimal bandwidth h . A joint search would require including $\int \dots \int \hat{f}(x; b)^2 f(x)^2 dx$. As $f(x)$ is unknown, and the sample analogue cannot simply be used as is done in the other cases, this evaluation becomes more difficult to evaluate. For this reason, b is estimated before using cross validation methods.

Similar to the direct integrated square error cross validation derivation, I derive an expression that has the same minimizing argument and is a function only of observed data. This is done in Section 5.1.3, and comes out to be

$$\begin{aligned} WISE^*(h) = & \frac{1}{h_k^2 n^3 \prod_{s \in G_x} h_s^2 b_s} \sum_{i=1}^n \sum_{j=1}^n \sum_{m=1}^n \int K' \left(\frac{x_k - x_{ik}}{h_k} \right) K' \left(\frac{x_k - x_{jk}}{h_k} \right) K \left(\frac{x_k - x_{mk}}{b_k} \right) dx_k \dots \\ & \times \prod_{s \in G_x \setminus k} \int K \left(\frac{x_s - x_{is}}{h_s} \right) K \left(\frac{x_s - x_{js}}{h_s} \right) K \left(\frac{x_s - x_{ms}}{b_s} \right) dx_s \\ & + 2 \frac{1}{n(n-1)^2 h_k^2 \prod_{s \in G_x} h_s^2} \sum_{m=1}^n \left[\sum_{i \neq m} K'' \left(\frac{x_{mk} - x_{ik}}{h_k} \right) \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{is}}{h_s} \right) \sum_{j \neq m} \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{js}}{b_s} \right) \right. \\ & \left. + \sum_{i \neq m} K' \left(\frac{x_{mk} - x_{ik}}{h_k} \right) \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{is}}{h_s} \right) \sum_{j \neq m} K' \left(\frac{x_{mk} - x_{jk}}{b_k} \right) \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{js}}{b_s} \right) \right] \end{aligned}$$

As employed by Hardle, Marron and Wand (1990) for the derivative of a univariate density, the theorems of Marron and Hardle (1986) can be applied here for the derivative of a multivariate density to show that the criteria of ISE and WISE both converge to the optimal criteria of MISE, and that the bandwidths resulting are not only valid for consistent estimation, but converge to the optimal MISE bandwidths.

3 Simulation Results

Wand and Schucany (1990) show that the $2r^{th}$ degree Gaussian kernel can be represented by

$$G_{2r}(x) = \frac{(-1)^r \phi^{(2r-1)}(x)}{2^{r-1}(r-1)!x}$$

I use this to derive the cross validation criteria for the different order kernels. The infinite order (Dirichlet) kernel, as Hansen (2005) presents it, is $K(x) = \frac{\sin x}{\pi x}$.

I run a Monte Carlo with 100 simulations on the data, with various sample sizes (100, 500, 1000, 5000) and various density dimensions (1,2,4,8). Data is simulated from a multivariate normal distribution with zero mean and variance-covariance matrix with variances given by $\sigma_i^2 = 1$ and covariance elements given by $\sigma_{ij} = 0.4$. The simulations use Gaussian kernels of order 2, 4, 6, 8, and 10, as well as the infinite dimension Dirichlet kernel. The weighted integrated square error is also tested, which uses a second order kernel. Two statistics of the simulations are examined: the mean square error ($\sum_i (\partial \hat{f}(x_i)/\partial x_k - \partial f(x_i)/\partial x_k)^2$) and the maximum absolute deviation ($\max_i |\partial \hat{f}(x_i)/\partial x_k - \partial f(x_i)/\partial x_k|$).

Figure 1, an estimated univariate derivative of a density for 500 observations for a single simulation, demonstrates different MSEs and Maximum Deviations, to provide intuition for the closeness of fit for the later reported statistics. For this simulation, the Dirichlet kernel performed the best, while the weighted integrated square error cross validation and the second order standard cross validation perform comparably.

Tables 1 - 4 in Section 5.2 show the Monte Carlo results for the simulation study. Higher order kernels outperform lower order kernels virtually universally. The Dirichlet kernel is overall the best performing, although often the higher order kernels are almost as good. Most of the gains from using higher order kernels are exhausted after a few increases in order—often, just increasing the kernel order to 4 improves the mean square error on average and the maximum deviations, but increasing the order beyond that does little more. The second order term in the Taylor expansion, the bias of the 2nd order kernels, is more important to control for than higher order terms. Higher order kernels’ improvement is most dramatic for low sample sizes and small dimensions, but the effect persists both with higher sample sizes and dimensions. Larger sample sizes yield more accurate results, and the weighted square error does poorly for small sample sizes, and only slightly worse than its 2nd order counterpart for large sample sizes. With the univariate case, it seems to do slightly better than its 2nd order counterpart, but higher dimension densities are poorly estimated using the WISE minimization criterion.

To offer comparison across dimension size, I look at how large the absolute maximum heights of the derivative of the densities are. While the higher dimensional densities have lower MSEs and maximum deviations, they are also coming from densities with much lower maximum and average absolute heights. The maximum and average heights of the densities are shown in Table 5. The ratios of the best maximum deviation for the various sample sizes and densities (i.e., the maximum that is the smallest for all kernel orders) to both the average and the maximum absolute derivative of the density height are calculated (Tables 6 and 7). For example, for a univariate density derivative and $N = 100$, the best average

maximum deviation comes from using the Dirichlet kernel, which has an average maximum deviation of 0.117 (Table 1). Table 5 shows that, for a univariate density of the type used in the Monte Carlo simulations, the maximum absolute height is 0.242, and the average absolute height is 0.1628.

This helps to frame how good of an estimate the .117 is—it is smaller than the average absolute height, but not by much. In fact, the ratio of the average maximum deviation in the simulations to the maximum absolute height is $0.117/0.2420=0.4835$, meaning that the density estimator is off on average by almost half of the highest height of the true derivative of the density, or 0.718 of the average height. Clearly, this is not a satisfactory outcome.

Higher values of N quickly yield much better results. This accentuates how much worse some of the higher densities estimates are, even with the best estimators used here. Looking at the ratio of the average maximum deviation to the average absolute height in Table 7, for $N = 1000$, the univariate density yields the acceptable ratio of 0.3274 (compare this to the plot in Figure 1, where the Dirichlet kernel has a ratio of 0.2872). For an 8-dimensional density derivative, this fraction jumps all the way up to 7.09, much too high for meaningful estimation of the underlying true derivative of the density. Theoretically, for some N high enough, the ratio would fall back into an acceptable range; however, the sample sizes that would be required for accurate estimation would be too large to be tractable.²

Overall, these results suggest that when trying to estimate a derivative of an unconditional multivariate density, higher order kernels and the straightforward cross validation criterion yield the best outcomes.

4 Conclusion

Researchers have developed and employed various methods to evaluate derivatives of univariate densities non-parametrically. However, little attention has been given to multivariate cases. In this paper, I examine cross validation methods for higher dimension densities, comparing different kernel orders and various criteria.

The primary limitation of the results of this paper is that the Monte Carlo studies only estimate Gaussian densities, although with covariance. Nonetheless, I argue that these results are more generalizable. A Monte Carlo simulation to test out the criteria suggests the complications inherent with estimating high dimensional density derivatives, as the accuracy of the estimates sharply decreases with increased dimension. The computational requirements are prohibitively large for increasing the sample size to compensate for the larger dimension when the dimension is overly large. However, when high dimensional derivative densities need to be estimated, the simulations suggest that higher order kernels are more effective, and in particular, use of the Dirichlet infinite order kernel performs the best in the class of Gaussian kernels. Weighted integrated square error methods are both slower and less effective generally, so I recommend for derivatives of multidimensional densities using the direct cross validation methods.

²The estimation of the bandwidth parameter using the methods in this paper would take far too long, even with parallel processing, as was used in this project.

5 Appendix

5.1 Proofs and Derivations

5.1.1 Proofs

Theorem 1. Consider the estimation of the derivative of a q^{th} dimensional density using an r^{th} order kernel $K(\cdot)$. If the following conditions hold

1. The kernel integrates to 1: $\int \cdots \int \prod_{s \in G_x} K(x_s) dx_1 \cdots dx_q = 1$
2. The kernel is symmetric: $K(x_s) = K(-x_s)$
3. The $r^{th}-1$ moment of the kernel is 0: $\int \cdots \int x_k^{r-1} \prod_{s \in G_x} K(x_s) dx_1 \cdots dx_q = 0$
4. The r^{th} moment of the kernel is positive and finite: $\int \cdots \int x_k^r \prod_{s \in G_x} K(x_s) dx_1 \cdots dx_q > 0$ and $\int \cdots \int x_k^r \prod_{s \in G_x} K(x_s) dx_1 \cdots dx_q < \infty$
5. $f(x)$ is $(r+1)$ times differentiable

Then,

$$MSE \left(\frac{\partial \hat{f}(x)}{\partial x_k} \right) = O \left(\left(\sum_{t=1}^q h_t^{r+1} \right)^2 + \frac{1}{nh_k \prod_{s \in G_x} h_s} \right)$$

and, if as $n \rightarrow \infty$, $\max_j \{h_j\} \rightarrow 0$ and $nh_k \prod_{s \in G_x} h_s \rightarrow \infty$, then $\frac{\partial \hat{f}(x)}{\partial x_k} \rightarrow \frac{\partial f(x)}{\partial x_k}$ in MSE, implying also convergence in probability, and consistency.

Proof. We can divide the MSE into bias squared and variance. First, examine the bias:

$$\begin{aligned} bias \left(\frac{\partial \hat{f}(x)}{\partial x_k} \right) &= E \left(\frac{\partial \hat{f}(x)}{\partial x_k} \right) - \frac{\partial f(x)}{\partial x_k} \\ &= E \left[\frac{1}{nh_k \prod_{s \in G_x} h_s} \sum_{i=1}^n K' \left(\frac{x_k - x_{ik}}{h_k} \right) \prod_{s \in G_x \setminus k} K \left(\frac{x_s - x_{is}}{h_s} \right) \right] - \frac{\partial f(x)}{\partial x_k} \\ &= \frac{1}{h_k \prod_{s \in G_x} h_s} \int \cdots \int K' \left(\frac{x_k - x_{ik}}{h_k} \right) \prod_{s \in G_x \setminus k} K \left(\frac{x_s - x_{is}}{h_s} \right) f(x_i) dx_i - \frac{\partial f(x)}{\partial x_k} \end{aligned}$$

Substitute $z_{is} = \frac{x_{is} - x_s}{h_s}$, Note that this implies $x_{is} = x_s + h_s z_{is}$ and $dx_{is} = h_s dz_{is}$. This yields

$$\frac{1}{h_k} \int \cdots \int K'(-z_{ik}) \prod_{s \in G_x \setminus k} K(-z_{is}) f(x + h z_i) dz_i - \frac{\partial f(x)}{\partial x_k}$$

Next, assume that $K(-z_{is}) = K(z_{is})$, implying $K'(-z_{is}) = -K'(z_{is})$. Then this is

$$\frac{-1}{h_k} \int \cdots \int K'(z_{ik}) \prod_{s \in G_x \setminus k} K(z_{is}) f(x + h z_i) dz_i - \frac{\partial f(x)}{\partial x_k}$$

Using integration by parts, and noting that $f(x_i + h z_i) K(z_k)|_{-\infty}^{\infty} = 0$, this is equal to

$$\int \cdots \int \prod_{s \in G_x} K(z_{is}) f_k(x + h z_i) dz_i - \frac{\partial f(x)}{\partial x_k}$$

Take a $(r + 1)$ degree Taylor expansion of $f_k(x + h z_i)$ around x

$$\begin{aligned} f_k(x + h z_i) = & \sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i \in [0, r-1]} \frac{\partial^{\sum_{i=1}^q j_i+1} f(x)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} \\ & + \sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i = r} \frac{\partial^{\sum_{i=1}^q j_i+1} f(x)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} \\ & + \sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i = r+1} \cdots \sum_{j_q = (r+1) \times 1 (t=q)}^{r+1} \frac{\partial^{\sum_{i=1}^q j_i+1} f(\xi)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} \end{aligned}$$

where $\xi \in [x, x + h z_i]$. Substituting this in,

$$\begin{aligned} \text{bias} \left(\frac{\partial \hat{f}(x)}{\partial x_k} \right) = & \int \cdots \int \prod_{s \in G_x} K(z_{is}) \sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i \in [0, r-1]} \frac{\partial^{\sum_{i=1}^q j_i+1} f(x)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} dz_i \\ & + \int \cdots \int \prod_{s \in G_x} K(z_{is}) \sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i = r} \frac{\partial^{\sum_{i=1}^q j_i+1} f(x)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} dz_i \\ & + \int \cdots \int \prod_{s \in G_x} K(z_{is}) \sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i = r+1} \frac{\partial^{\sum_{i=1}^q j_i+1} f(\xi)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} dz_i \end{aligned}$$

Consider the first term; this becomes

$$\sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i \in [0, r-1]} \frac{\partial^{\sum_{i=1}^q j_i+1} f(x)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \int \cdots \int \prod_{s \in G_x} K(z_{is}) \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} dz_i$$

Given Assumption 3, all of the integrals evaluate to zero here except for the one where all j are equal to zero. When all s are equal to zero, Assumption 1 says that $\int \cdots \int \prod_{s \in G_x} K(z_{is}) dz_i = 1$. Therefore,

$$\sum_{j_1, \dots, j_q | j_i \geq 0, \sum_i j_i \in [0, r-1]} \frac{\partial^{\sum_{i=1}^q j_i+1} f(x)}{\partial x_k \partial^{j_1} x_1 \partial^{j_2} x_2 \cdots \partial^{j_q} x_q} \int \cdots \int \prod_{s \in G_x} K(z_{is}) \prod_{\ell=1}^q \frac{(h_\ell z_{i\ell})^{j_\ell}}{j_\ell!} dz_i = \frac{\partial f(x)}{\partial x_k}$$

Next, consider the second term. Now, by Assumption 3, all of the summands that have any $j_i \in (0, r)$ is equal to zero. Therefore, the only remaining elements are those where one is equal to r and the rest are equal to zero (these, by Assumption 2, integrate to one).

Therefore, the second term is

$$\begin{aligned} & \frac{1}{r!} \sum_{t=1}^q h_t^r \frac{\partial^{r+1} f(x)}{\partial x_k \partial^r x_t} \int \cdots \int \prod_{s \in G_x} K(z_{is}) z_{it}^r dz_i = \frac{1}{r!} \sum_{t=1}^q h_t^r \frac{\partial^{r+1} f(x)}{\partial x_k \partial^r x_t} \int K(z_{it}) z_{it}^r dz_{it} \\ & = \frac{\int K(x) x^r dx}{r!} \sum_{t=1}^q h_t^r \frac{\partial^{r+1} f(x)}{\partial x_k \partial^r x_t} \end{aligned}$$

Next, consider the final term. All summands that include $z_{i\ell}$ for $\ell \in [0, r]$ equal zero.

$$\begin{aligned} & \frac{1}{(1+r)!} \sum_{t=1}^q h_t^{r+1} \frac{\partial^{r+2} f(\xi)}{\partial x_k \partial^{r+1} x_t} \int \cdots \int \prod_{s \in G_x} K(z_{is}) z_{it}^{r+1} dz_i \\ & = \frac{1}{(1+r)!} \sum_{t=1}^q h_t^{r+1} \frac{\partial^{r+2} f(\xi)}{\partial x_k \partial^{r+1} x_t} \int K(z_{it}) z_{it}^{r+1} dz_{it} \end{aligned}$$

Using Assumption 5, for some C

$$\frac{1}{(1+r)!} \sum_{t=1}^q h_t^{r+1} \frac{\partial^{r+2} f(\xi)}{\partial x_k \partial^{r+1} x_t} \int K(z_{it}) z_{it}^{r+1} dz_{it} < C \sum_{t=1}^q h_t^{r+1} = O\left(\sum_{t=1}^q h_t^{r+1}\right)$$

Combining these results,

$$bias\left(\frac{\partial \hat{f}(x)}{\partial x_k}\right) = \frac{\int K(x) x^r dx}{r!} \sum_{t=1}^q h_t^r \frac{\partial^{r+1} f(x)}{\partial x_k \partial^r x_t} + O\left(\sum_{t=1}^q h_t^{r+1}\right)$$

Next, examine the variance.

$$\begin{aligned} var\left(\frac{\partial \hat{f}(x)}{\partial x_k}\right) &= var\left(\frac{1}{nh_k \prod_{s \in G_x} h_s} \sum_{i=1}^n K'\left(\frac{x_k - x_{ik}}{h_k}\right) \prod_{s \in G_x \setminus k} K\left(\frac{x_s - x_{is}}{h_s}\right)\right) \\ &= \frac{1}{nh_k^2 \prod_{s \in G_x} h_s^2} \left[E\left(K'\left(\frac{x_k - x_{ik}}{h_k}\right)^2 \prod_{s \in G_x \setminus k} K\left(\frac{x_s - x_{is}}{h_s}\right)^2\right) \right. \\ & \quad \left. - E\left(K'\left(\frac{x_k - x_{ik}}{h_k}\right) \prod_{s \in G_x \setminus k} K\left(\frac{x_s - x_{is}}{h_s}\right)\right)^2 \right] \end{aligned}$$

Consider the terms separately. The first term is

$$\begin{aligned}
& \frac{1}{nh_k^2 \prod_{s \in G_x} h_s^2} \left[E \left(K' \left(\frac{x_k - x_{ik}}{h_k} \right)^2 \prod_{s \in G_x \setminus k} K \left(\frac{x_s - x_{is}}{h_s} \right)^2 \right) \right] \\
&= \frac{1}{nh_k \prod_{s \in G_x} h_s} \int \cdots \int K' (z_{ik})^2 \prod_{s \in G_x \setminus k} K (z_{is})^2 \left[f(x) + \sum_{t=1}^q \frac{\partial f(\xi)}{\partial x_k} h_t z_{it} \right] dz_i \\
&= \frac{f(x)}{nh_k \prod_{s \in G_x} h_s} \int K' (x)^2 dx \left(\int K (x)^2 dx \right)^{q-1} + O \left(\frac{1}{nh_k \prod_{s \in G_x} h_s} \right)
\end{aligned}$$

Next consider the second term

$$\begin{aligned}
\frac{1}{nh_k^2 \prod_{s \in G_x} h_s^2} E \left(K' \left(\frac{x_k - x_{ik}}{h_k} \right) \prod_{s \in G_x \setminus k} K \left(\frac{x_s - x_{is}}{h_s} \right) \right)^2 &= \frac{1}{nh_k^2} \left(\int \cdots \int K' (z_{ik}) \prod_{s \in G_x \setminus k} K (z_{is}) dz_i \right)^2 \\
&= O \left(\frac{1}{nh_k^2} \right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{var} \left(\frac{\partial \hat{f}(x)}{\partial x_k} \right) &= \frac{f(x)}{nh_k \prod_{s \in G_x} h_s} \int K' (x)^2 dx \left(\int K (x)^2 dx \right)^{q-1} + O \left(\frac{1}{nh_k \prod_{s \in G_x} h_s} \right) - O \left(\frac{1}{nh_k^2} \right) \\
&= \frac{f(x)}{nh_k \prod_{s \in G_x} h_s} \int K' (x)^2 dx \left(\int K (x)^2 dx \right)^{q-1} + O \left(\frac{1}{nh_k \prod_{s \in G_x} h_s} \right)
\end{aligned}$$

Combining these results

$$\begin{aligned}
MSE \left(\frac{\partial \hat{f}(x)}{\partial x_k} \right) &= \left(\frac{\int K (x) x^r dx}{r!} \sum_{t=1}^q h_t^r \frac{\partial^{r+1} f(x)}{\partial x_k \partial^r x_t} + O \left(\sum_{t=1}^q h_t^{r+1} \right) \right)^2 \\
&\quad + \frac{f(x)}{nh_k^2 \prod_{s \in G_x} h_s} \int K' (x)^2 dx \left(\int K (x)^2 dx \right)^{q-1} + O \left(\frac{1}{nh_k \prod_{s \in G_x} h_s} \right) \\
&= O \left(\left(\sum_{t=1}^q h_t^{r+1} \right)^2 + \frac{1}{nh_k \prod_{s \in G_x} h_s} \right)
\end{aligned}$$

Therefore, if as $n \rightarrow \infty$, $\max_j \{h_j\} \rightarrow 0$ and $nh_k \prod_{s \in G_x} h_s \rightarrow \infty$, then the last statement directly implies that $\frac{\partial \hat{f}(x)}{\partial x_k} \rightarrow \frac{\partial f(x)}{\partial x_k}$ in MSE, implying also convergence in probability, and consistency. $\square$

5.1.2 Derivation of CV bandwidth for Derivative of Joint

I consider the derivative of a joint density with respect to variable $x_k \in x$.

$$\begin{aligned} ISE(h) &= \int \cdots \int \left(\frac{\partial \hat{f}(x)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right)^2 dx \\ &= \int \cdots \int \frac{\partial \hat{f}(x)}{\partial x_k}^2 dx - 2 \int \cdots \int \frac{\partial \hat{f}(x)}{\partial x_k} \frac{\partial f(x)}{\partial x_k} dx + \int \cdots \int \frac{\partial f(x)}{\partial x_k}^2 dx \\ &= ISE_1(h) - 2 * ISE_2(h) + ISE_3 \end{aligned}$$

Again, ISE_3 is not a function of the bandwidth selection, so minimizing ISE is identical to minimizing

$$ISE^*(h) = ISE_1(h) - 2 * ISE_2(h)$$

Examine each term separately

$$\begin{aligned} ISE_1(h) &= \frac{1}{n^2 (\prod_{s \in G_x} h_s)^2 h_k^2} \sum_{i=1}^n \sum_{j=1}^n \int \cdots \int K' \left(\frac{x_k - x_{ik}}{h_k} \right) K' \left(\frac{x_k - x_{jk}}{h_k} \right) \cdots \\ &\quad \times \prod_{s \in G_x \setminus k} K \left(\frac{x_s - x_{is}}{h_s} \right) K \left(\frac{x_s - x_{js}}{h_s} \right) dx \end{aligned}$$

Let $\bar{x}_{is} = \frac{x_s - x_{is}}{h_s}$ and $\tilde{x}_{ijs} = \frac{x_{is} - x_{js}}{h_s}$. Then

$$\begin{aligned} ISE_1(h) &= \frac{1}{n^2 (\prod_{s \in G_x} h_s)^2 h_k^2} \sum_{i=1}^n \sum_{j=1}^n \int \cdots \int \left[\int h_k K'(\tilde{x}_{ijk} + \bar{x}_{ik}) K'(\bar{x}_{ik}) d\bar{x}_{ik} \right] \cdots \\ &\quad \times \prod_{s \in G_x \setminus k} (h_s K(\tilde{x}_{ijs} + \bar{x}_{is}) K(\bar{x}_{is}) d\bar{x}_{is}) \\ &= \frac{1}{n^2 h_k^2 \prod_{s \in G_x} h_s} \sum_{i=1}^n \sum_{j=1}^n \left[\int K'(\tilde{x}_{ijs} + \bar{x}_{is}) K'(\bar{x}_{is}) d\bar{x}_{is} \right] \prod_{s \in G_x \setminus k} \left(\int K(\tilde{x}_{ijs} + \bar{x}_{is}) K(\bar{x}_{is}) d\bar{x}_{is} \right) \end{aligned}$$

As in the joint density case, the evaluation of these integrals depend on the kernel chosen. Section 1 provides an example of this, the analogy of which carries directly into this example, except now derivatives of the kernel must also be taken prior to integration of that section.

Next, examine $ISE_2(h)$:

$$ISE_2(h) = \int \cdots \int \frac{\partial \hat{f}(x)}{\partial x_k} \frac{\partial f(x)}{\partial x_k} dx$$

Integrating by parts,

$$\begin{aligned} \int \cdots \int \frac{\partial \hat{f}(x)}{\partial x_k} \frac{\partial f(x)}{\partial x_k} dx &= - \int \cdots \int \frac{\partial^2 \hat{f}(x)}{\partial x_k^2} f(x) dx = -E \left(\frac{\partial^2 \hat{f}(x)}{\partial x_k^2} \right) \\ &= \frac{-1}{n(n-1) \left(\prod_{s \in G_x} h_s \right)^2 h_k^2} \sum_{i=1}^n \sum_{j \neq i} K''(\tilde{x}_{ijk}) \prod_{s \in G_x \setminus k} K(\tilde{x}_{ijs}) \end{aligned}$$

Putting these together

$$\begin{aligned} ISE^*(h) &= \\ \frac{1}{n^2 h_k^2 \prod_{s \in G_x} h_s} \sum_{i=1}^n \sum_{j=1}^n \left[\int K'(\tilde{x}_{ijs} + \bar{x}_{is}) K'(\bar{x}_{is}) d\bar{x}_{is} \right] \prod_{s \in G_x \setminus k} \left(\int K(\tilde{x}_{ijs} + \bar{x}_{is}) K(\bar{x}_{is}) d\bar{x}_{is} \right) \\ + \frac{2}{n(n-1) \left(\prod_{s \in G_x} h_s \right)^2 h_k^2} \sum_{i=1}^n \sum_{j \neq i} K''(\tilde{x}_{ijk}) \prod_{s \in G_x \setminus k} K(\tilde{x}_{ijs}) \end{aligned}$$

5.1.3 Derivation of Weighted CV bandwidth for Derivative of Joint

I consider the derivative of a joint density with respect to variable $x_k \in x$. Let h be the bandwidth that the researcher is trying to minimize with respect to, and $\hat{f}(x; b)$ be the weighting function. b is estimated prior to this, so that the weighting is independent of the bandwidth selection h . Then, the weighted integrated square error is given by

$$\begin{aligned} WISE(h) &= \int \cdots \int \left(\frac{\partial \hat{f}(x; h)}{\partial x_k} - \frac{\partial f(x)}{\partial x_k} \right)^2 \hat{f}(x; b) dx \\ &= \int \cdots \int \frac{\partial \hat{f}(x; h)}{\partial x_k}^2 \hat{f}(x; b) dx - 2 \int \cdots \int \frac{\partial \hat{f}(x; h)}{\partial x_k} \frac{\partial f(x)}{\partial x_k} \hat{f}(x; b) dx \\ &\quad + \int \cdots \int \frac{\partial f(x)}{\partial x_k}^2 \hat{f}(x; b) dx \\ &= WISE_1(h) - 2 * WISE_2(h) + WISE_3 \end{aligned}$$

$WISE_3$ is not a function of the bandwidth (h) selection, so minimizing $WISE$ is identical to minimizing

$$WISE^*(h) = WISE_1(h) - 2 * WISE_2(h)$$

$$\begin{aligned}
WISE_1(h) &= \int \cdots \int \frac{\partial \hat{f}(x; h)}{\partial x_k} \hat{f}(x; b) dx \\
&= \int \cdots \int \frac{1}{h_k^2 n^3 \prod_{s \in G_x} h_s^2 b_s} \sum_{i=1}^n \sum_{j=1}^n \sum_{m=1}^n K' \left(\frac{x_k - x_{ik}}{h_k} \right) K' \left(\frac{x_k - x_{jk}}{h_k} \right) K \left(\frac{x_k - x_{mk}}{b_k} \right) \cdots \\
&\quad \times \prod_{s \in G_x \setminus k} K \left(\frac{x_s - x_{is}}{h_s} \right) K \left(\frac{x_s - x_{js}}{h_s} \right) K \left(\frac{x_s - x_{ms}}{b_s} \right) dx \\
&= \frac{1}{h_k^2 n^3 \prod_{s \in G_x} h_s^2 b_s} \sum_{i=1}^n \sum_{j=1}^n \sum_{m=1}^n \int K' \left(\frac{x_k - x_{ik}}{h_k} \right) K' \left(\frac{x_k - x_{jk}}{h_k} \right) K \left(\frac{x_k - x_{mk}}{b_k} \right) dx_k \cdots \\
&\quad \times \prod_{s \in G_x \setminus k} \int K \left(\frac{x_s - x_{is}}{h_s} \right) K \left(\frac{x_s - x_{js}}{h_s} \right) K \left(\frac{x_s - x_{ms}}{b_s} \right) dx_s
\end{aligned}$$

This depends on the choice of the kernel, but is estimable. Next,

$$WISE_2(h) = \int \cdots \int \frac{\partial \hat{f}(x; h)}{\partial x_k} \frac{\partial f(x)}{\partial x_k} \hat{f}(x; b) dx$$

Integrating by parts, and assuming that the value of the derivative is bounded, this becomes

$$\begin{aligned}
WISE_2(h) &= -E \left[\frac{\partial^2 \hat{f}(x; h)}{\partial x_k^2} \hat{f}(x; b) + \frac{\partial \hat{f}(x; h)}{\partial x_k} \frac{\partial \hat{f}(x; b)}{\partial x_k} \right] \\
&= C_1 \sum_{m=1}^n \left[\sum_{i \neq m} K'' \left(\frac{x_{mk} - x_{ik}}{h_k} \right) \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{is}}{h_s} \right) \sum_{j \neq m} \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{js}}{b_s} \right) \right] \\
&\quad + C_2 \sum_{m=1}^n \left[\sum_{i \neq m} K' \left(\frac{x_{mk} - x_{ik}}{h_k} \right) \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{is}}{h_s} \right) \sum_{j \neq m} K' \left(\frac{x_{mk} - x_{jk}}{b_k} \right) \prod_{s \in G_x} K \left(\frac{x_{ms} - x_{js}}{b_s} \right) \right]
\end{aligned}$$

where

$$C_1 = \frac{1}{n(n-1)^2 h_k^2 \prod_{s \in G_x} h_s b_s}, \quad C_2 = \frac{1}{n(n-1)^2 h_k b_k \prod_{s \in G_x} h_s b_s}$$

All of the elements are now evaluated in order to estimate the *WISE* and minimize with respect to h . However, note that it become very complicated to analytically evaluate the integral in $WISE_1$ for a given kernel. The researcher will most likely need to limit attention to simpler (generally lower order) kernels. Here, I give the example of the 2nd order Gaussian Kernel for exposition. In that case,

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\{-.5x^2\} \equiv \phi(x), \quad K'(x) = -x\phi(x), \quad K''(x) = (x^2 - 1)\phi(x)$$

For $\int K\left(\frac{x-r}{h}\right) K\left(\frac{x-s}{h}\right) K\left(\frac{x-t}{b}\right) dx$

$$\begin{aligned}
&= \int \frac{1}{(2\pi)^{3/2}} \exp \left\{ -.5 \left(\left(\frac{x-r}{h} \right)^2 + \left(\frac{x-s}{h} \right)^2 + \left(\frac{x-t}{b} \right)^2 \right) \right\} dx \\
&= \int \frac{1}{(2\pi)^{3/2}} \exp \left\{ -\frac{1}{2\Omega} ((x-\mu)^2 + D) \right\} dx \\
&= \frac{\sqrt{\Omega} \exp\{\frac{1}{2}\Omega^{-1}D\}}{2\pi} \int \frac{1}{\sqrt{2\pi\Omega}} \exp \left\{ -\frac{1}{2\Omega} (x-\mu)^2 \right\} dx \\
&= \frac{\sqrt{\Omega} \exp\{\frac{1}{2}\Omega^{-1}D\}}{2\pi}
\end{aligned}$$

where

$$\Omega = \frac{b^2 h^2}{2b^2 + h^2}, \quad \mu = \frac{b^2 r + b^2 s + h^2 t}{2b^2 + h^2}, \quad D = \frac{r^2 b^2 + s^2 b^2 + t^2 h^2}{2b^2 + h^2} - \mu^2$$

Similarly, for $\int K'\left(\frac{x-r}{h}\right) K'\left(\frac{x-s}{h}\right) K\left(\frac{x-t}{b}\right) dx$

$$\begin{aligned}
&= \int \frac{(x-r)(x-s)}{h^2(2\pi)^{3/2}} \exp \left\{ -.5 \left(\left(\frac{x-r}{h} \right)^2 + \left(\frac{x-s}{h} \right)^2 + \left(\frac{x-t}{b} \right)^2 \right) \right\} dx \\
&= \int \frac{x^2 - (r+s)x + rs}{h^2(2\pi)^{3/2}} \exp \left\{ -\frac{1}{2\Omega} ((x-\mu)^2 + D) \right\} dx \\
&= \frac{\sqrt{\Omega} \exp\{\frac{1}{2}\Omega^{-1}D\}}{2h^2\pi} \int \frac{1}{\sqrt{2\pi\Omega}} (x^2 - (r+s)x + rs) \exp \left\{ -\frac{1}{2\Omega} (x-\mu)^2 \right\} dx \\
&= \frac{\sqrt{\Omega} \exp\{\frac{1}{2}\Omega^{-1}D\}}{2h^2\pi} (\text{Var}(x) + E[x]^2 - (r+s)E[x] + rs) \\
&= \frac{\sqrt{\Omega} \exp\{\frac{1}{2}\Omega^{-1}D\}}{2h^2\pi} (\Omega + \mu^2 - (r+s)\mu + rs)
\end{aligned}$$

5.2 Tables

Table 1: Monte Carlo Simulation Results: (100 Sims): Derivative of 1-Dimensional Density

	Mean MSE				Mean Max. Dev.			
	N=100	N=500	N=1000	N=5000	N=100	N=500	N=1000	N=5000
2nd Order	0.914 (9.05)	0.0253 (0.149)	0.0186 (0.116)	0.0155 (0.132)	0.439 (2.92)	0.184 (0.385)	0.146 (0.366)	0.113 (0.357)
4th Order	0.0102 (0.0201)	0.0108 (0.0685)	0.00253 (0.00696)	0.000962 (0.00197)	0.139 (0.131)	0.114 (0.248)	0.0734 (0.0895)	0.0479 (0.0498)
6th Order	0.0114 (0.024)	0.00428 (0.0097)	0.00186 (0.00407)	0.000919 (0.00234)	0.141 (0.142)	0.0899 (0.0907)	0.0631 (0.0684)	0.0431 (0.053)
8th Order	0.0125 (0.0268)	0.00454 (0.0104)	0.0018 (0.00439)	0.000923 (0.00267)	0.145 (0.15)	0.0915 (0.0954)	0.0603 (0.0686)	0.0408 (0.0556)
10th Order	0.015 (0.0301)	0.00482 (0.0101)	0.0019 (0.0048)	0.000942 (0.00286)	0.158 (0.164)	0.0951 (0.0963)	0.061 (0.0718)	0.0405 (0.0564)
Dirichlet	0.011 (0.0347)	0.00448 (0.0106)	0.00221 (0.0115)	0.00114 (0.00427)	0.117 (0.145)	0.0824 (0.0873)	0.0533 (0.0874)	0.0367 (0.0611)
WISE	0.0525 (0.233)	0.092 (0.473)	0.0189 (0.0765)	0.00444 (0.0204)	0.241 (0.49)	0.335 (0.89)	0.167 (0.358)	0.0985 (0.171)

Standard Deviations of MSE Simulation Estimates in Parentheses

Table 2: Monte Carlo Simulation Results: (100 Sims): Derivative of 2-Dimensional Density

	Mean MSE				Mean Max. Dev.			
	N=100	N=500	N=1000	N=5000	N=100	N=500	N=1000	N=5000
2nd Order	0.00849 (0.0282)	0.0064 (0.0276)	0.00633 (0.0498)	0.000602 (0.00148)	0.164 (0.238)	0.144 (0.234)	0.112 (0.297)	0.0658 (0.0541)
4th Order	0.00493 (0.0161)	0.00154 (0.00329)	0.000543 (0.000917)	0.000298 (0.000955)	0.121 (0.161)	0.0849 (0.0805)	0.0535 (0.0399)	0.0406 (0.0374)
6th Order	0.00302 (0.00668)	0.00152 (0.00377)	0.000498 (0.000975)	0.000149 (0.000187)	0.101 (0.0959)	0.0804 (0.0816)	0.0487 (0.0375)	0.0314 (0.021)
8th Order	0.00306 (0.0074)	0.00159 (0.00424)	0.000499 (0.00113)	0.000147 (0.000199)	0.0972 (0.0978)	0.0805 (0.0856)	0.0471 (0.0396)	0.03 (0.021)
10th Order	0.00259 (0.0061)	0.00067 (0.001)	0.000387 (0.000745)	0.000104 (8.47e-05)	0.0901 (0.0916)	0.0573 (0.0438)	0.0408 (0.028)	0.0239 (0.00973)
Dirichlet	0.000835 (0.00173)	0.000447 (0.00105)	0.00149 (0.0113)	6.95e-05 (0.000101)	0.0557 (0.0357)	0.0413 (0.0281)	0.0426 (0.0785)	0.0177 (0.0108)
WISE	1.12 (11.1)	0.014 (0.0471)	0.00865 (0.058)	0.0014 (0.00423)	0.786 (6.14)	0.216 (0.38)	0.15 (0.35)	0.093 (0.112)

Standard Deviations of MSE Simulation Estimates in Parentheses

Table 3: Monte Carlo Simulation Results: (100 Sims): Derivative of 4-Dimensional Density

	Mean MSE			Mean Max. Dev.		
	N=100	N=500	N=1000	N=100	N=500	N=1000
2nd Order	0.00257 (0.012)	0.000207 (0.000958)	0.000113 (0.000652)	0.0805 (0.221)	0.0332 (0.0524)	0.0268 (0.0457)
4th Order	8.43e-05 (0.000147)	2.64e-05 (4.01e-05)	1.7e-05 (2.13e-05)	0.0264 (0.0196)	0.0179 (0.0099)	0.015 (0.00805)
6th Order	6.05e-05 (9.25e-05)	2.15e-05 (2.8e-05)	1.37e-05 (1.9e-05)	0.0225 (0.0144)	0.0161 (0.00893)	0.0132 (0.00674)
8th Order	5.58e-05 (8.59e-05)	1.82e-05 (2.84e-05)	1.56e-05 (6.45e-05)	0.0213 (0.0132)	0.0147 (0.0086)	0.0122 (0.00866)
10th Order	3.23e-05 (8.62e-06)	2.25e-05 (4.41e-06)	2.04e-05 (3.13e-06)	0.017 (0.00269)	0.0161 (0.00174)	0.0157 (0.00118)
Dirichlet	2.21e-05 (2.06e-05)	8.12e-06 (4.92e-06)	5.51e-06 (3.58e-06)	0.0139 (0.00416)	0.0101 (0.00256)	0.00869 (0.0024)
WISE	0.452 (3.48)	0.00493 (0.0233)	0.0153 (0.146)	1.01 (4.57)	0.164 (0.424)	0.139 (0.799)

Standard Deviations of MSE Simulation Estimates in Parentheses

Table 4: Monte Carlo Simulation Results: (100 Sims): Derivative of 8-Dimensional Density

	Mean MSE			Mean Max. Dev.		
	N=100	N=500	N=1000	N=100	N=500	N=1000
2nd Order	2.42e-05 (0.00013)	3.38e-08 (5.54e-08)	1.64e-08 (3.55e-09)	0.00859 (0.0297)	0.00102 (0.000702)	0.000851 (0.000111)
4th Order	3.56e-08 (6.13e-08)	1.3e-08 (3.37e-09)	1.04e-08 (1.78e-09)	0.000715 (0.000418)	0.000705 (0.000117)	0.000729 (0.000111)
6th Order	2.04e-08 (0.000186)	1.14e-08 (0.000114)	8.84e-09 (9.81e-05)	0.00061 (1.31e-08)	0.000679 (2.77e-09)	0.000686 (1.58e-09)
8th Order	1.68e-08 (0.000159)	1.07e-08 (0.000111)	8.22e-09 (9.53e-05)	0.000584 (6.8e-09)	0.00067 (2.26e-09)	0.00067 (1.53e-09)
10th Order	2.25e-08 (7.16e-09)	2.34e-08 (3.07e-09)	2.29e-08 (2.86e-09)	0.000659 (0.000179)	0.000894 (0.000145)	0.000979 (9.82e-05)
Dirichlet	1.71e-08 (5.64e-09)	1.17e-08 (1.87e-09)	8.65e-09 (1.42e-09)	0.000603 (0.000169)	0.000713 (0.000125)	0.000705 (9.01e-05)
WISE	122 (1.21e+03)	39.1 (314)	0.397 (1.36)	8.94 (78)	17.8 (97.7)	4.67 (13.9)

Standard Deviations of MSE Simulation Estimates in Parentheses

Table 5: Derivative Density Size Comparisons

Dimension	$\max_{x_i} \partial f(x_i)/\partial x_k $	$\frac{1}{n} \sum_i \frac{\partial f(x_i)}{\partial x_k}$
1	0.2420	0.1628
2	0.1148	0.0520
4	0.0253	0.0062
8	0.0011	0.00009

Table 6: Ratio of Best Average Maximum Deviation to Maximum True Derivative Density Height (Dirichlet Kernel)

Dimension	$N = 100$	$N = 500$	$N = 1000$	$N = 5000$
1	0.4835	0.3405	0.2202	0.1517
2	0.4852	0.3598	0.3554	0.1542
4	0.5494	0.3992	0.2727	-
8	0.5309	0.6091	0.6091	-

Table 7: Ratio of Best Average Maximum Deviation to Average True Derivative Density Height (Dirichlet Kernel)t

Dimension	$N = 100$	$N = 500$	$N = 1000$	$N = 5000$
1	0.7187	0.5061	0.3274	0.2254
2	1.0712	0.7942	0.7846	0.3404
4	2.2419	1.6290	1.1129	-
8	6.1855	7.0963	7.0963	-

5.3 Figures

Figure 1: Derivative of Univariate Normal Density and Estimations, $N = 500$

5.4 References

- Bearse, Peter and Paul Rilstone, P. (2008). “Higher Order Bias Reduction of Kernel Density and Density Derivative Estimators at Boundary Points,” 7th Annual Advances in Econometrics Conference, Baton Rouge, Louisiana.
- Cacoullos, Theophilos (1966). “Estimation of a Multivariate Density,” *Annals of the Institute of Statistical Mathematics* (Tokyo), 18:2, 179-189.
- Chacon, J.E., Duong, T., and Wand, M.P. (2010). “Asymptotics for General Multivariate Kernel Density Derivative Estimators.” *Statistica Sinica*, 21, 807-840.
- Hansen, Bruce (2005). “Exact Mean Integrated Squared Error of Higher Order Kernel Estimators,” *Econometric Theory*, 21: 1031-1057
- Hardle, Wolfgang, J.S. Marron and M.P. Wand (1990), “Bandwidth Choice for Density Derivatives,” *Journal of the Royal Statistical Society, Series B (Methodological)* 52:1 223-232.
- Li, Qi, Racine, Jeffrey (2007). *Nonparametric Econometrics*. Princeton University Press, Princeton, NJ.
- Loader, Clive R. (1999). “Bandwidth Selection: Classical or Plug-In?” *The Annals of Statistics* 27:2 415-438.
- Marron, J.S. (1994). “Visual Understanding of Higher-Order Kernels,” *Journal of Computational and Graphical Statistics*, 3:4 447-458.
- Marron, J.S. and Hardle, W. (1986). “Random Approximations to Some Measures of Accuracy in Nonparametric Curve Estimation,” *Journal of Multivariate Analysis* 20, 91-113.
- Marron, J.S. and M.P. Wand (1992). “Exact mean integrated squared error,” *Annals of Statistics*, 20, 712-736.
- Turlach, Berwin A. (1993). “Bandwidth Selection in Kernel Density Estimation: A Review,” Discussion Paper 9307, Institut für Statistik und Ökonometrie, Humbolt-Universität zu Berlin, Berlin, Germany.
- Wand and Schucany (1990). “Gaussian-Based Kernels,” *The Canadian Journal of Statistics* 18:3 197-204.
- Wu, Tiee-Jian (1997). “Root n Bandwidth Selectors for Kernel Estimation of Density Derivatives,” *Journal of the American Statistical Association* 92:438 536-547.
- Zhang, Xibin, Maxwell L. King, and Rob J. Hyndman (2006). “A Bayesian approach to bandwidth selection for multivariate kernel density estimation,” *Computational Statistics & Data Analysis* 50, 3009-3031.