LOCAL BANDWIDTH SELECTION FOR KERNEL ESTIMATION OF POPULATION DENSITIES WITH LINE TRANSECT SAMPLING

Patrick D. Gerard
Experimental Statistics Unit
Mississippi State University,
Mississippi 39762

William R. Schucany
Department of Statistical Science
Southern Methodist University
Dallas, TX 75275-0332

*Revised version to
appear Biometrics
(March?) 1999*

Local Bandwidth Selection for Kernel Estimation of Population Densities with Line Transect Sampling

Patrick D. Gerard

Experimental Statistics Unit, Box 9653, Mississippi State University, MS 39762

William R. Schucany

Department of Statistical Science, Southern Methodist University, Dallas TX 75275

SUMMARY

Seber (1986) suggested an approach to population density estimation using kernel estimates of the probability density of detection distances in line transect sampling. Chen (1996) and others have employed cross validation to choose a global bandwidth for the kernel estimator or suggested adaptive kernel estimation (Chen 1997). Since estimation of the density is required at only a single point, we propose a local bandwidth selection procedure that is a modification of the method of Schucany (1995) for nonparametric regression. We report on simulation results comparing the proposed method with cross validation and adaptive estimation. The local bandwidths produce estimates with mean squares that are half the size of the others. Consistency results are also provided.

key words : detectability, nonparametric, survey, ecology

1. Introduction

Sampling to estimate population density is a common practice among biologists. For example, the populations of a large number of species of mussels are declining (Williams et al. 1993) which has lead to efforts to quantify and monitor the population densities of the involved species. One of the more frequently employed sampling techniques is line transect sampling (Burnham et al. 1980, Buckland et al. 1993). In line transect sampling, the distances from the objects of interest to randomly selected transect lines are used. The result is to estimate population density by $\hat{D} = \dfrac{\hat{f}(0)n}{2L}$, where n is the number of objects seen, L is the length of line traversed, and $\hat{f}(0)$ is an estimate of the underlying probability density function of perpendicular sighting distances evaluated on the transect line (i.e. distance = 0).

Numerous authors have suggested parametric estimates of $f(0)$ (Polluck 1978, Buckland 1985), while Burnham et al. (1980) suggested using Fourier series methods. Seber (1986) seems to be the first to suggest using kernel density estimation (see Silverman 1986) techniques to obtain a nonparametric estimate of $f(0)$. Kernel density estimates require the specification of a smoothing parameter or bandwidth that governs the smoothness of the resulting estimator. Figure 1 depicts a kernel estimate of the probability density of the sighting distances of mussels using a arbitrary global bandwidth of 20 cm. The estimated density at zero is used in the formula above to estimate population density. The goal in bandwidth selection is to select a bandwidth that balances bias and variance. The typical approach minimizes some estimate of mean squared error (MSE).

[Insert Figure 1 here]

Figure 1. Estimated Kernel Density of Sighting Distances of Mussel Line Transect Sightings

Using a Global Bandwidth of 20 cm.
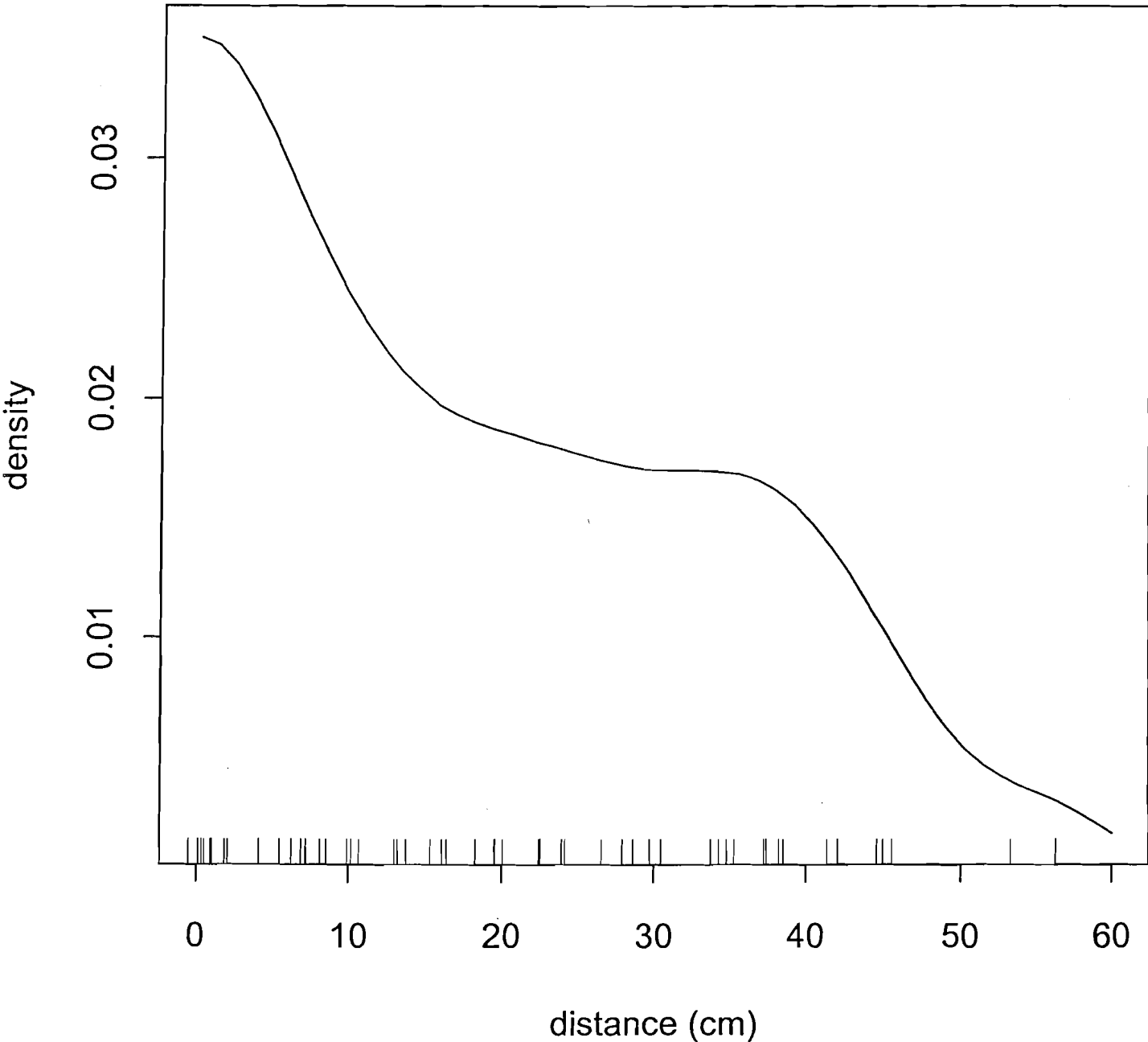
One method of bandwidth selection in line transect sampling suggested by Chen (1996), is least squares cross validation (LSCV). It is a global estimate, using the same bandwidth at each point of estimation. Because it is global in nature, it does not allow for different amounts of smoothing at different estimation points. Specifically, in line transect sampling where an estimate of the probability density of distances is only required at a single point, the bandwidth is influenced by the shape of the density at other locations.

Another method of bandwidth estimation, suggested to overcome some of the difficulties of LSCV, is adaptive bandwidth estimation. A form of the adaptive method is advocated by Chen (1997) in line transect sampling. Adaptive bandwidth selection allows a different bandwidth for each observation, thus allowing different amounts of smoothing at different points of estimation. Implementation typically requires both the use of a pilot estimate of the density and cross validation.

Yet another method of bandwidth selection that is especially pertinent to line transect sampling is local bandwidth selection. In local bandwidth selection, a different bandwidth is used at each point of estimation. Hence, in line transect sampling, only one bandwidth for estimation at zero needs to be specified. However, consistent estimates of unknown quantities are required to plug into an expression for the bandwidth that minimizes local asymptotic MSE.

In Section 3, we present a local bandwidth estimator similar to the one developed for kernel regression by Schucany (1995). This modification is particularly useful in line transect sampling. We show by simulation results in Section 4 that the MSE of the resulting population density estimates are smaller than those of LSCV and adaptive bandwidth estimation by a factor of two or more. We also apply each method in the estimation of the population density of a

common species of mussel (*Actinonaias ligametina*). Consistency results for the local bandwidth selection estimators are provided in a technical appendix.

2. Kernel Estimation of Population Density

To estimate the probability density, $f(t)$, of a random variable without specifying the form of the underlying density Silverman (1986) and Wand and Jones (1995) describe the kernel estimator

$$\hat{f}(t;h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - t}{h}\right),$$

where $x_1,\ldots,x_n$ are a random sample, $K(\cdot)$ is a weight function, and h is a smoothing parameter. Typically, $K(\cdot)$ is taken to be a symmetric, univariate probability density function, like the normal density or the quadratic density, $K(u) = .75\left(1 - u^2\right)I(u)_{\{-1,1\}}$, which has been shown to have good asymptotic properties (Epanechnikov 1969). Choice of the kernel function has been shown to be less critical than selection of the smoothing parameter (Silverman 1986). A large bandwidth will result in a more biased but less variable estimate and a small bandwidth will result in a more variable but less biased estimate. Hence, a reasonable goal is the selection of a bandwidth that minimizes MSE.

When estimating population density using distances from transect lines, the estimate $\hat{D}$, is

$$\hat{D} = \frac{\hat{f}(0;h)n}{2L},$$

where n is the realized number of objects sighted, L is the length of the transect line, and $\hat{f}(0;h)$ is the kernel estimate of the density of sighting distances from the transect line.

4

One feature of kernel estimates is their increased bias near bounds on the domain of the data. It has been suggested by Silverman (1986) that one way to account for this boundary bias is to adjust the estimator near the edges resulting in

$$\hat{f}(t;h) = \frac{1}{nh}\sum_{i=1}^{n}\left(K\left(\frac{x_i - t}{h}\right) + K\left(\frac{-x_i - t}{h}\right)\right).$$ (2.1)

This adjusted estimator has little or no effect away from the edge of the data. For estimation at zero, as with distance data,

$$\hat{f}(0;h) = \frac{2}{nh}\sum_{i=1}^{n}K\left(\frac{x_i}{h}\right),$$ (2.2)

because $K(\cdot)$ is symmetric about zero.

Chen (1996) suggested least squares cross validation (LSCV) to determine h. Essentially, LSCV minimizes an estimate of mean integrated squared error, $\int\left(\hat{f}(x) - f(x)\right)^2 dx$, by finding the value of h that minimizes

$$M(h) = \int\hat{f}(x;h)^2 dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(x_i;h),$$

where $\hat{f}_{-i}(x_i;h)$ is a kernel estimate as in (2.1) without the ith data value. Obviously, LSCV yields an estimate of bandwidth that satisfies a global criteria even though estimation at only one point is required for line transect data. Furthermore, LSCV is quite variable and converges slowly to the asymptotically optimal value (Hall and Marron 1987, Park and Marron 1990). We will include this $\hat{f}\left(0;\hat{h}_{cv}\right)$ for (2.2) in our simulation study.

Chen (1997) also suggested, in a somewhat different setting, that adaptive bandwidth selection be used. An adaptive density estimate, $\hat{f}_a(t;\mathbf{h})$, analogous to (2.1) is

$$\hat{f}_a(t;\mathbf{h}) = \sum_{i=1}^{n} \frac{1}{nh_i} \left(K\left(\frac{x_i - t}{h_i}\right) + K\left(\frac{-x_i - t}{h_i}\right) \right).$$

It should be noted that a different bandwidth is used for each data value. The theory for this method of bandwidth estimation suggests that reducing the order of the bias is possible (Abramson 1982). Chen suggested the implementation of Silverman (1986) which requires a pilot bandwidth, usually taken from a reference parametric family of distributions. For the normal distribution using the quadratic kernel, this pilot bandwidth is $h_r = 2.345n^{-1/5}s_x$, where s_x is the usual sample standard deviation of the sighting distances. The adaptive bandwidths are then $h_i = h_a\lambda_i$, where $\lambda_i = \left(\dfrac{\hat{f}(x_i;h_r)}{g} \right)^{-1/2}$, $\hat{f}$ uses (2.1) employing the reference bandwidth, and g is the geometric mean of those density estimates. The value of h_a is chosen using cross validation and hence inherits some of the problems inherent with cross validation. In principle, the adaptive bandwidths allow for differential smoothing depending on how much the density is changing at each observation. The estimator $\hat{f}_a(0;\hat{\mathbf{h}})$ also appears in Section 4.

In the next section we propose a local bandwidth selection procedure that allows a different bandwidth to be specified at each point of estimation. Therefore in the present application, only one bandwidth needs to be estimated.

3. Local Bandwidth Selection

Local bandwidths allow one to change the amount of smoothing based on the degree that the underlying density is changing. Hence a local density estimate is

$$\hat{f}_l(t;h(t)) = \frac{1}{nh(t)} \sum_{i=1}^{n} K\left(\frac{x_i - t}{h(t)}\right).$$

Therefore, the localized version of (2.2) is

$$\hat{f}_l(0; h(0)) = \frac{2}{nh(0)} \sum_{i=1}^{n} K\left(\frac{x_i}{h(0)}\right).$$

Since in this line transect application it is clear that only a bandwidth at zero is required, the dependence of the bandwidth on the point of estimation will be suppressed. The necessary assumptions for consistency of this estimator are that, even though it is not fixed in advance, $n \to \infty$ and $h \to 0$ such that $nh \to \infty$ and continuity of $f''(0)$. In the line transect setting it is also typically assumed that $f'(0)=0$ as well. Under these assumptions, the expected value and variance at the boundary of the positive distances are

$$E\left(\hat{f}_l(0; h)|n\right) = f(0) + k_2^* f''(0)h^2 + o\left(h^2\right)$$

and

$$\text{var}\left(\hat{f}_l(0; h)|n\right) = \frac{4f(0)Q^*}{nh} + o\left(\frac{1}{nh}\right),$$

where $k_2^* = \int_0^{\infty} u^2 K(u)\,du$ and $Q^* = \int_0^{\infty} K^2(u)\,du$. Hence the asymptotic MSE is

$$AMSE\left(\hat{f}_l(0; h)|n\right) = \left(k_2^* f''(0)\right)^2 h^4 + \frac{4f(0)Q^*}{nh}. \tag{3.1}$$

The bandwidth that minimizes (3.1) is found by differentiation to be

$$h_{opt} = \left[\frac{f(0)Q^*}{n\left(k_2^* f''(0)\right)^2}\right]^{1/5} = \left[\frac{A}{B}\right]^{1/5}, \tag{3.2}$$

where $A = \dfrac{f(0)Q^*}{n}$ and $B = \left(k_2^* f''(0)\right)^2$. To obtain a data driven bandwidth, estimation of A and B is required because each contains unknown quantities. If consistent estimates of both can

be obtained, then they can be plugged into (3.2) to estimate h_{opt}. We will modify the local bandwidth proposal of Schucany (1995) for nonparametric regression here in the density estimation setting. Estimation of B is motivated by the form of the asymptotic squared bias, $bias^2 \approx Bh^4$.

The form of this relationship is that of a regression equation through the origin with squared bias as the dependent variable, B as the unknown regression coefficient, and h^4 as the independent variable. If the bias were known for a grid of bandwidths, then B could be approximated using least squares. Because the bias is not known, it must be estimated. A reasonable estimate is

$$\hat{bias}_j = \hat{b}_j = k_2^* \hat{f}''(0)h_j^2, \qquad (3.3)$$

where $\hat{f}''(0)$ is a fourth-order kernel estimator for the second derivative, for example

$K_4(u) = \dfrac{105}{16}\left(-1 + 6u^2 - 5u^4\right)I(u)_{\{-1,1\}}$. The resulting estimator is

$$\hat{f}''(0) = \frac{2}{nh^3}\sum_{i=1}^{n} K_4\left(\frac{x_i}{h}\right),$$

with the factor of two required for bias correction of edge effects analogous to (2.2). If we form a grid of trial bandwidths, $h_1,\ldots,h_q$, such that $h_j = C_j n^{-\rho}$, with $|C_j|<\infty$ and $\rho>0$, then the resulting least squares estimate of B is

$$\hat{B} = \frac{\sum\limits_{j=1}^{q} \hat{b}_j^2 h_j^4}{\sum\limits_{j=1}^{q} h_j^8} = B + O\left(n^{-\rho\gamma}\right) + O_p\left(n^{-\frac{1-5\rho}{2}}\right),$$

where γ is the Lipschitz constant of $f''(\cdot)$. Details are provided in Theorem 1 in the Appendix.

Similarly estimation of A follows from $\mathrm{var} \approx \dfrac{A}{h}$. If the variance were known for a grid

of bandwidths, then A could be estimated by least squares. As before these variances can be

estimated. One estimate of A involves a simple plug-in estimate of variance,

$$\widehat{\mathrm{var}}_{1j} = \hat{v}_{1j} = \frac{\hat{f}\big(0;h_j\big)Q^*}{nh_j},$$

where $\hat{f}\big(0;h_j\big)$ is the kernel estimate (2.2). Hence, for a grid of bandwidths, $h_{11},\dots,h_{1q}$, such that

$h_{1j} = D_j n^{-\delta}$, $|D_j|<\infty$ and $\delta>0$, the least squares solution is

$$\hat{A}_1 = \frac{\displaystyle\sum_{j=1}^{q} \hat{v}_{1j} \frac{1}{h_{1j}}}{\displaystyle\sum_{j=1}^{q} \frac{1}{h_{1j}^2}} = A + O\big(n^{-1-2\delta}\big) + O_p\!\left(n^{\frac{\delta-3}{2}}\right).$$

Details are provided in Theorem 2 in the Appendix.

An alternative estimate of A follows by noting that

$$\hat{f}(0;h) = \frac{2}{nh}\sum_{i=1}^{n} K\!\left(\frac{x_i}{h}\right) = \frac{2}{nh}\sum_{i=1}^{n} y_i ,$$

where $y_i = K\!\left(\dfrac{x_i}{h}\right)$. A natural estimate of variance is suggested the fact that for each fixed h

$$\mathrm{var}\big(\hat{f}(0;h)\big) = \frac{4}{n^2 h^2}\,\mathrm{var}\!\left(\sum_{i=1}^{n} y_i\right) = \frac{4}{nh^2}\,\mathrm{var}\big(y_i\big).$$

The y_i are independent and identically distributed because the x_i are. Thus var(y_i) can be

estimated by the usual sample variance of the y_i's, s_y^2. The resulting unbiased estimate of the

variance is $\hat{v}_{2j} = \dfrac{4}{nh_j^2} s_y^2$. Once again, using least squares for a grid of bandwidths,

$h_{2j} = E_j n^{-\theta}$, for $j=1,\ldots,q$, $|E_j|<\infty$, and $\theta>0$, the resulting estimate is

$$\hat{A}_2 = \frac{\displaystyle\sum_{j=1}^{q} \hat{v}_{2j} \frac{1}{h_{2j}}}{\displaystyle\sum_{j=1}^{q} \frac{1}{h_{2j}^2}} = A + o\left(n^{-1}\right) + O_p\left(n^{\theta-3/2}\right).$$

Details are provided in Theorem 3 in the Appendix. In each case, the resulting bandwidth

estimate is of the form $\left[\dfrac{\hat{A}}{\hat{B}}\right]^{1/5}$.

It should be noted that $\rho<1/5$ is required for consistency of $\hat{B}$. This parallels the results

of Schucany (1995). Additionally, $\delta<3$ is required for consistency of $\hat{A}_1$, and $\theta<3/2$ for

consistency of $\hat{A}_2$. In the simulation study reported in the next section, different grids are used

for the three methods.

4. A Monte Carlo Comparison of Density Estimators

To evaluate the performance of the four estimators proposed in the previous section, a

simulation study analogous to that of Chen (1996) was conducted. A constant population density

of .15 was maintained in all of the simulation runs. The effective width of the transect area was

10 and N=200, 300, and 400 objects were randomly generated at distances uniformly distributed

in the interval [0,10]. The exponential power series detection function, $g(x) = \exp\left(-(bx)^a\right)$,

was used with b=0.5 and a=1.5, 2.0, and 2.5. For each simulation run, 5000 independent

replications were generated.

For each replication, estimators of population density were calculated with the kernel method employing LSCV and the adaptive method described in Section 2. Additionally, two local bandwidth estimators were employed, one using $\hat{A}_1$ and the other $\hat{A}_2$, both with $\hat{B}$ as described in Section 3.

The grid values used in the simulation study for $\hat{A}_1$, $\hat{A}_2$, and $\hat{B}$ were developed from the asymptotic properties of the estimators as well as previous experience with nonparametric regression (Schucany 1995, Gerard and Schucany 1997) in which seven equally spaced grid values produce satisfactory results. The grid values for $\hat{B}$ correspond to $\gamma=1$ which results in $\rho=1/7$ as an optimal rate; hence the grid values are set to be $(j/7)n^{-1/7} \times (\text{range of data})$ for $j=1,...,7$. Similarly, the grid for $\hat{A}_1$ is $(j/7)n^{-1/5} \times (\text{range of data})$ for $j=1,...,7$ and for $\hat{A}_2$ is $(j/7)n^{-1/2} \times (\text{range of data})$ for $j=1,...,7$. The quadratic kernel from Section 2 was used for K in all four estimators.

The results of the simulation study are summarized in Table 1. The number of objects sighted, n, was averaged over all replications of each run. The averages range from 35 for $N=200$ to 72 for $N=400$. The average population density and MSE are calculated for each method. The two local estimation methods, though slightly more biased in some cases, have MSE's that are significantly smaller, by a factor of two or more, compared to the adaptive and LSCV estimators. Additionally, local estimation has using variance estimator $\hat{A}_1$ has significantly smaller MSE than using $\hat{A}_2$, based on s_y^2. However, the differences seen were typically less than 10%. To examine these comparisons, we calculated the efficiencies of the LSCV method relative to the adaptive method (the average MSE for the adaptive method divided by the average MSE for the LSCV method) and the local method with variance estimate

$\hat{A}_1$ relative to the adaptive method (the average MSE for the adaptive method divided by the average MSE for the local method). These relative efficiencies are plotted in Figure 2. The maximum standard error of the points are estimated to be .184 for the dotted lines and .046 for the solid lines, using standard approximation techniques for ratios of dependent averages. In the next section, each of these methods is used to estimate the population density of a common species of mussels (*Actinonaias ligamentina*).

[Insert Table 1 here]

[Insert Figure 2 here]

5. Application to a Population of Mussels

The population density of many species of mussels has been declining the past 30 years (Williams et al. 1993), which has motivated close monitoring of population trends. As part of that effort, line transect sampling was used to estimate the density of a common mussel (*Actinonaisa ligamentina*) in French Creek, which is located in the Allegheny River Basin in Pennsylvania. Distances from a line 42 meters in length were recorded for 53 mussels. The four methods evaluated in the simulation study were used to estimate the underlying probability density of sighting distances. These are displayed in Figure 3. The bandwidths estimated at 0 were used for the entire curves for the local estimators. For this data set, all of the methods yielded similar results. The resulting population density estimates are 1.81 and 1.71 mussels/m^2 for the local methods, 1.77 mussels/m^2 for LSCV, and 1.76 mussels/m^2 for the adaptive method.

[Insert Figure 3 here]

6. Summary and Discussion

Using the kernel method to estimate population density requires that a bandwidth be specified for estimating the probability density of sighting distances. Because estimation is only

Table 1. Summary of Simulation Study Averages (Standard Errors) Based on 5000 Replications

a	N	n	LSCV			Adaptive		local ($\hat{A}_1$)			local ($\hat{A}_2$)		
			h	bias[1]	MSE[2]	bias	MSE	h	bias	MSE	h	bias	MSE
1.5	200	36	1.38 (.08)	-1.1 (.01)	32.1 (1.5)	-.44 (.08)	31.0 (1.5)	1.55 (.005)	-2.0 (.04)	13.5 (.23)	1.48 (.006)	-.20 (.04)	14.1 (.25)
	300	54	1.25 (.09)	-1.2 (.008)	21.4 (.87)	-.54 (.06)	20.2 (.86)	1.46 (.004)	-2.0 (.04)	10.8 (.18)	1.41 (.005)	-1.9 (.04)	11.0 (.19)
	400	72	1.16 (.11)	-1.1 (.007)	16.2 (.67)	-.41 (.05)	14.3 (.54)	1.38 (.003)	-1.8 (.03)	8.6 (.14)	1.34 (.004)	-1.8 (.03)	8.7 (.14)
2	200	36	1.32 (.08)	-.60 (.009)	29.9 (1.4)	.21 (.08)	30.5 (1.5)	1.33 (.005)	-.91 (.05)	12.9 (.26)	1.25 (.006)	-1.0 (.05)	14.3 (.31)
	300	53	1.20 (.09)	-.57 (.008)	23.7 (1.5)	.18 (.07)	21.4 (.98)	1.28 (.004)	-.84 (.04)	9.1 (.18)	1.21 (.005)	-.87 (.04)	9.8 (.21)
	400	71	1.11 (.11)	-.55 (.007)	14.7 (.58)	.22 (.05)	14.5 (.56)	1.24 (.004)	-.83 (.03)	6.8 (.13)	1.19 (.004)	-.83 (.04)	7.1 (.14)
2.5	200	35	1.28 (.08)	-.33 (.009)	30.1 (1.4)	.42 (.08)	32.4 (1.6)	1.22 (.005)	-.50 (.05)	14.2 (.33)	1.13 (.006)	-.64 (.06)	16.6 (.41)
	300	53	1.19 (.09)	-.33 (.008)	21.4 (1.3)	.49 (.07)	23.3 (1.8)	1.20 (.004)	-.43 (.04)	9.6 (.20)	1.13 (.005)	-.50 (.05)	10.7 (.25)
	400	71	1.10 (.11)	-.28 (.007)	16.9 (.89)	.49 (.06)	18.6 (1.0)	1.18 (.004)	-.38 (.04)	7.3 (.16)	1.13 (.005)	-.42 (.04)	8.1 (.19)

[1] bias × 10²

[2] MSE × 10⁴

Figure 2. Relative Efficiency of LSCV to adaptive (solid) and adaptive to local (dashed) using

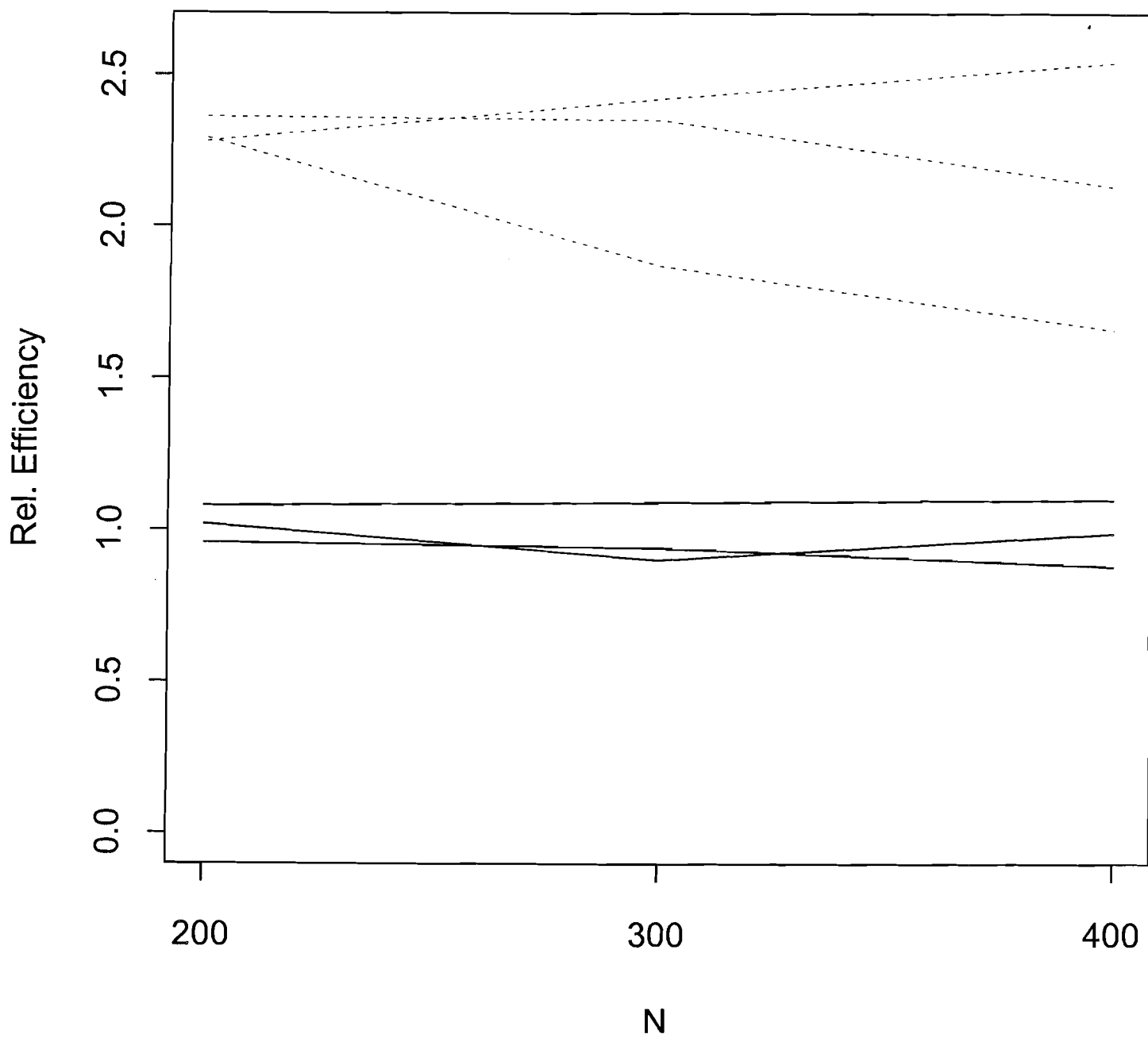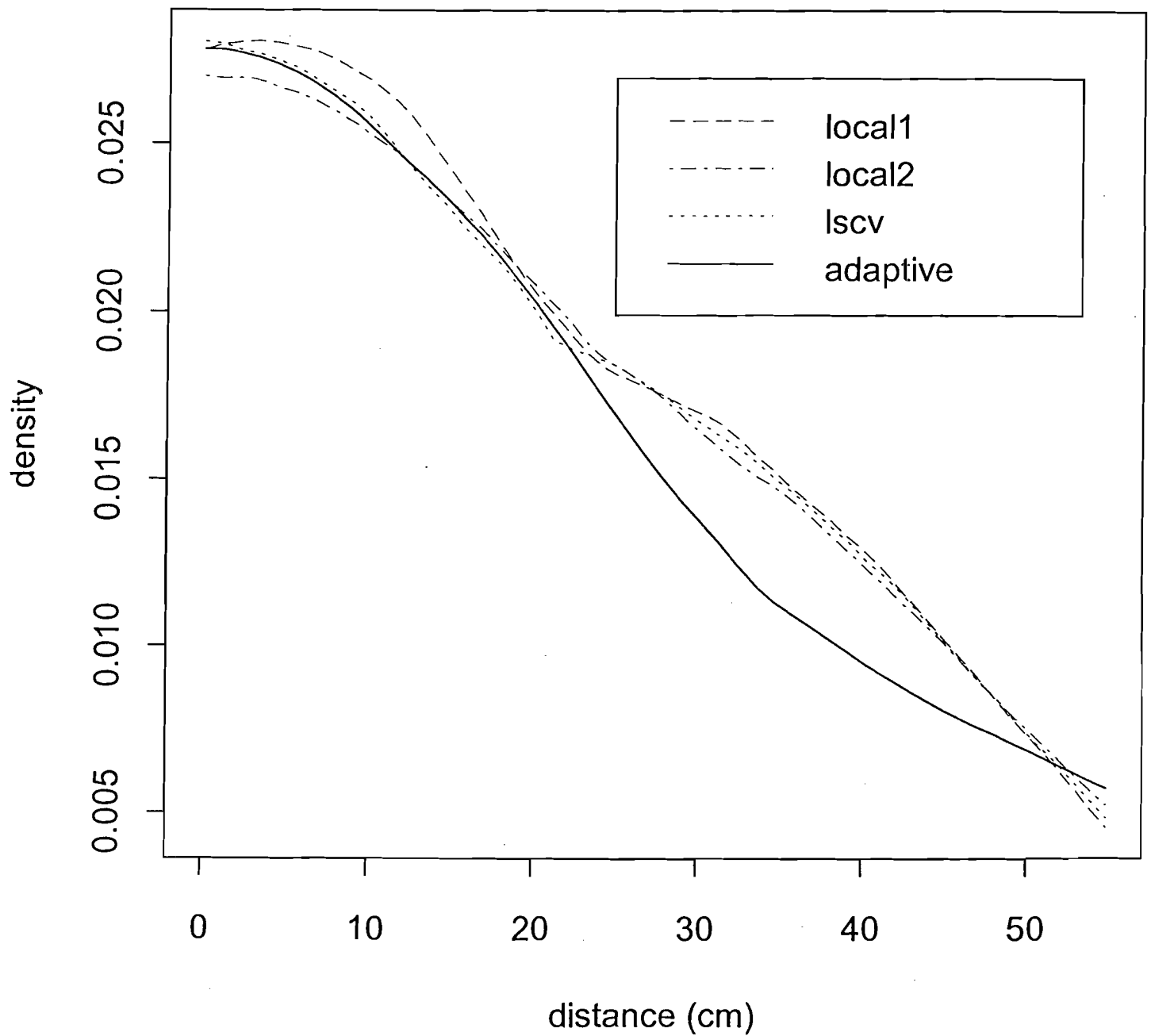variance estimate $\hat{A}_1$ for the values of a and N in Table 1.

Figure 3. Kernel Density Estimates of Sighting Distances for Mussel Line Transect Data. Bandwidth Selection is done locally using $\hat{A}_1$ (local1) and $\hat{A}_2$ (local2), globally using Least Squares Cross Validation (lscv), and Adaptively (adaptive).

required at a single point, excessive influence of data far from the point of estimation should be avoided. With that in mind, two new methods of estimating a local bandwidth are described and found in a simulation to have smaller MSE than methods previously advocated. It also appears that the method of variance estimation involving the simple plug-in principle outperforms regression on unbiased variance estimators. One drawback to local estimation is that it does not necessarily integrate to one. This problem is obviated by the fact that estimation of the entire density is not required.

Another benefit of these local methods is their computational efficiency. Unless one resorts to complicated binning routines, the number of operations required for LSCV and the adaptive methods is $O\left(n^2\right)$, but is $O(n)$ for the local methods proposed here.

7. Acknowledgements

Appendix

Theorem 1

Suppose that f'' is Lipschitz continuous of order γ and $f'(0)=0$. Conditionally on n, as $n\to\infty$ and $h\to 0$ such that $nh\to\infty$, then for a kernel function K supported on $[-1,1]$

$$\hat{B} = \frac{\sum_{j=1}^{q}\hat{b}_j^2 h_j^4}{\sum_{j=1}^{q} h_j^8} = B + O\left(n^{-\rho\gamma}\right) + O_p\left(n^{-\frac{1-5\rho}{2}}\right) .$$

13

if $h_j = C_j n^{-\rho}$ for $j = 1, \ldots, q$ with $\rho > 0$ and $|C_j| < \infty$.

Proof:

From (3.3),

$$\hat{b}_j = \left(k_2^* \hat{f}''(0) \right) h_j^2 ,$$

where $\hat{f}''(0) = \dfrac{2}{nh_j^3} \displaystyle\sum_{i=1}^{n} K_4 \left(\dfrac{x_i}{h_j} \right)$ for $K_4()$ a fourth-order kernel function for estimating second

derivatives. From standard Taylor series arguments,

$$E\left(\hat{f}''(0) \right) = \frac{2}{h_j^3} \int_0^\infty K_4 \left(\frac{x}{h_j} \right) f(x) dx$$

$$= \frac{2}{h_j^2} \int_0^1 K_4(u) \left(f(0) + f'(0)uh_j + \frac{1}{2} f''(\xi) u^2 h_j^2 \right) du,$$

with $0 < \xi < uh_j$ and $u = x/h_j$. This yields

$$E\left(\hat{f}''(0) \right) = \frac{2}{h_j^2} \int_0^1 K_4(u) \left(f(0) + f'(0)uh_j + \frac{1}{2} u^2 h_j^2 \left(f''(0) + O(\xi^\gamma) \right) \right) du,$$

$$= f''(0) + O\left(h_j^\gamma \right).$$

Again using standard arguments,

$$\text{var}\left(\hat{f}''(0) \right) = \frac{4}{h_j} \text{var}\left(\frac{1}{h_j^3} K_4 \left(\frac{x}{h_j} \right) \right) = O\left(\frac{1}{nh_j^5} \right).$$

Hence, $\hat{f}''(0) = f''(0) + O\left(h^\gamma \right) + O_p\left(\left(nh^5 \right)^{-1/2} \right)$. Therefore, substituting $h_j = C_j n^{-\rho}$ yields

14

$$\hat{B} = \frac{\sum_{j=1}^{q} \hat{b}_j^2 h_j^4}{\sum_{j=1}^{q} h_j^8} = \frac{\sum_{j=1}^{q} \left(k_2^* \hat{f}''(0) \right)^2 h_j^8}{\sum_{j=1}^{q} h_j^8} = \frac{\sum_{j=1}^{q} \left(k_2^* \left(f''(0) + O\left(h^\gamma \right) + O_p\left(\left(nh_j^5 \right)^{-1/2} \right) \right) \right)^2 h_j^8}{\sum_{j=1}^{q} h_j^8}$$

$$= B + O\left(n^{-\rho\gamma} \right) + O_p\left(n^{-\frac{1-5\rho}{2}} \right) \cdot \|$$

Theorem 2

If the conditions of Theorem 1 hold except that $h_{1j} = D_j n^{-\delta}$, for $j = 1, \ldots, q$ with $\delta > 0$ and $|D_j| < \infty$, then

$$\hat{A}_1 = \frac{\sum_{j=1}^{q} \hat{v}_{1j} \frac{1}{h_{1j}}}{\sum_{j=1}^{q} \frac{1}{h_{1j}^2}} = A + O\left(n^{-1-2\delta} \right) + O_p\left(n^{\frac{\delta-3}{2}} \right).$$

Proof:

The result follows from $\hat{v}_{1j} = \frac{\hat{f}(0) Q^*}{nh_{1j}}$, with $\hat{f}(0)$ from (2.3). Again using standard

arguments, $\hat{f}(0) = f(0) + O\left(h_{1j}^2 \right) + O_p\left(\left(nh_{1j} \right)^{-1/2} \right)$. Using $h_{1j} = D_j n^{-\delta}$ yields

$$\hat{A}_1 = \frac{\sum_{j=1}^{q} \left(\frac{Q^* \left(f(0) + O\left(h_{1j}^2 \right) + O_p\left(\left(nh_{1j} \right)^{-1/2} \right) \right)}{nh_{1j}} \right) \frac{1}{h_{1j}}}{\sum_{j=1}^{q} \frac{1}{h_{1j}^2}}$$

$$= A + O\left(n^{-1-2\delta} \right) + O_p\left(n^{\frac{\delta-3}{2}} \right) \cdot \|$$

Theorem 3

If the conditions of Theorem 1 hold except $h_{2j}=E_j n^{-\theta}$, for $j=1,\ldots,q$ with $\theta>0$ and $|E_j|<\infty$,

then

$$\hat{A}_2 = \frac{\sum_{j=1}^{q} \hat{v}_{2j} \frac{1}{h_{2j}}}{\sum_{j=1}^{q} \frac{1}{h_{2j}^2}} = A + o\left(n^{-1}\right) + O_p\left(n^{\theta-3/2}\right).$$

Proof:

Because

$$\hat{f}(0) = \frac{2}{nh} \sum_{i=1}^{n} K\left(\frac{x_i}{h}\right) = \frac{2}{nh} \sum_{i=1}^{n} y_i \text{ , where } y_i = K\left(\frac{x_i}{h}\right) \text{ , then}$$

$$\operatorname{var}\left(\hat{f}(0)\right) = \frac{4}{n^2 h^2} \operatorname{var}\left(\sum_{i=1}^{n} y_i\right) = \frac{4}{nh^2} \operatorname{var}(y_i).$$

An unbiased estimate is then $\hat{v}_2 = \frac{4}{nh^2} s_y^2$, where s_y^2 is the sample variance of the y's. If follows

that (Serfling 1980)

$$\operatorname{var}(\hat{v}_2) = \frac{1}{n^2 h^4} O\left(n^{-1}\right) = O\left(\frac{1}{n^3 h^4}\right)$$

because $\operatorname{var}\left(s_y^2\right) = O\left(n^{-1}\right)$. Hence,

$$\frac{1}{nh^2} s_y^2 = \operatorname{var}\left(\hat{f}(0)\right) + O_p\left(n^{-3/2}h^{-2}\right) = \frac{A}{h} + o\left(\frac{1}{nh}\right) + O_p\left(n^{-3/2}h^{-2}\right). \text{ Therefore, using}$$

least squares with $h_{2j}=|E_j|n^{-\theta}$ yields

$$\hat{A}_2 = \frac{\displaystyle\sum_{j=1}^{q} \hat{v}_{2j} \frac{1}{h_{2j}}}{\displaystyle\sum_{j=1}^{q} \frac{1}{h_{2j}^2}} = \frac{\displaystyle\sum_{j=1}^{q} \left(\frac{A}{h_{2j}} + o\left(\frac{1}{n h_{2j}} \right) + O_p\left(n^{-3/2} h_{2j}^2 \right) \right) \frac{1}{h_{2j}}}{\displaystyle\sum_{j=1}^{q} \frac{1}{h_{2j}^2}} = A + o\left(n^{-1} \right) + O_p\left(n^{\theta - 3/2} \right). \parallel$$

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates - a square root rule. *Annals of Statistics* **10**, 1217-1223.

Buckland, S. T. (1985). Perpendicular distance models for line transect sampling. *Biometrics* **41**, 177-185.

Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (1993). *Distance Sampling*. London : Chapman and Hall.

Burnham, K. P., Anderson, D. R., and Laake, J. L. (1980). *Estimation of Density from Line Transect Sampling of Biological Populations*. Wildlife Monograph Number 72.

Chen, S. X. (1996). A kernel estimate for the density of a biological population by using line transect sampling. *Applied Statistics* **45**, 135-150.

Chen, S. X. (1997). Studying school size effects in line transect sampling using the kernel method. *Biometrics* **52**, 1283-1294.

Epanechnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory of Probability and its Applications* **14**, 153-158.

Gerard, P. D. and Schucany, W. R. (1997) Methodology for nonparametric regression from independent sources. *Computational Statistics and Data Analysis* **25**, 287-304.

Hall, P. and Marron, J. S. (1987). Extent to which least-squares cross validation minimizes integrated squared error in nonparametric density estimation. *Probability Theory and Related Fields* **74**, 567-581.

Park, B. U. and Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association* **85**, 66-72.

Polluck, K. H. (1978). A family of density estimators for line transect sampling. *Biometrics* **34**,

475-478.

Seber, G. A. F. (1986). A review of estimating animal abundance. *Biometrics* **42**, 267-292.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York : Wiley.

Schucany, W. R. (1995). Adaptive bandwidth selection for kernel regression. *Journal of the American Statistical Association* **90**, 535-540.

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

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

Williams, J. D., Warren, M. L., Cummings, K. S., Harris, J. L., and Neves, R. J. Conservation Status of Freshwater Mussels of the United States and Canada. *Fisheries* 18, 6-22.