BIOMETRICS 52, 1283–1294
December 1996

Studying School Size Effects in Line Transect Sampling Using the Kernel Method

Song Xi Chen

CSIRO Division of Fisheries & Biometrics Unit, and
School of Statistics, La Trobe University, Victoria 3083, Australia

SUMMARY

When conducting line transect sampling to estimate the abundance of a clustered wildlife population, detection of a school depends not only on the perpendicular distance of the school to the transect line, but also on the size of school. Larger size schools are easier to detect than smaller schools. Thus, a bivariate detection function with distance and size as covariates should be considered. This paper considers using the kernel smoothing method to fit the bivariate line transect data in order to estimate both abundance and the mean school size. Two kernel estimators are studied: the fixed kernel estimator, which uses the same smoothing bandwidth for all data points, and the adaptive kernel estimator, which allows the bandwidth to vary across the data points.

1. Introduction

Line transect aerial surveys of commercially valuable southern bluefin tuna have been conducted over the Great Australian Bight in summer, when the tuna tend to form schools and stay on the surface. Surface tuna schools are detected by two experienced spotters on-board a light aircraft flying along randomly allocated transect lines. A satellite-based Global Positioning System (GPS) is used to measure the perpendicular distance from a detected school to the transect line. While flying above a detected school, the spotters give independent estimates of the size of the school (in tonnes), as the survey is more interested in the number of tonnes of tuna per unit area (i.e., the biomass density) than in schools per unit area. The spotters' estimates of the size are based on many years' commercial spotting and learning from the catches of fishing boats. The aim of the survey is to estimate the abundance of the tuna on the sea surface. This estimation serves as a relative abundance index as long as the surveys are conducted in a consistent way from year to year.

A tuna school of large size is more easily detected than a school of small size, given the same perpendicular sighting distance. This is the so-called "size effect." To correct this size effect, a bivariate detection function $g(x, s)$ should be considered. This is the probability of detecting a school given that the perpendicular sighting distance is x and the school size is s. Drummer and McDonald (1987) proposed parametric models for $g(x, s)$ by inserting the size covariate into the distance-only detection function. In their recent book on distance sampling, Buckland et al. (1993) suggest regressing the school size s_i (or log school size) on $\hat{g}(x_i)$—the estimated distance-only detection function at x_i. The nonparametric Fourier series (FS) method was used by Quang (1991) to correct the size effect. In our aerial survey, the survey plane flies about 800 kilometers per survey day between inshore and the continental shelf, and usually travels through different weather conditions in terms of wind speed, cloud cover, glare, etc. The sighting function $g(x, s)$ is subject to changes in the weather. So nonparametric estimates, which are robust against changing g, are sought.

With the aerial survey as background, this paper aims to develop nonparametric kernel estimates for the biomass density and the mean school size of a clustered population by taking account of the

Key words: Aerial survey; Biomass density; Bootstrap; Confidence intervals; Kernel method; Size-biased sampling.

size effect. Unlike the parametric line transect method, the kernel method does not need to assume a parametric detection function. Therefore, it is very much a data-oriented modeling approach, and is robust against changing environmental survey conditions. Compared with the Fourier method, the kernel method does not require an explicit truncation width in calculations, and produces real probability density functions which are nonnegative and have an integral value of one. Another advantage of the kernel method is that, by using a multivariate kernel, it can be generalized to analyze data of three or four dimensions if other environmental variables are found to be significant and are included in the detection function.

Section 2 introduces some notation and outlines the problem. Section 3 gives a general description of a kernel estimator. Section 4 describes two kernel estimators, the fixed and adaptive kernel estimators. Section 5 shows bootstrap algorithms to estimate the variance and confidence intervals for the biomass density and the mean school size. In Section 6 a data set from the tuna aerial survey is used to illustrate the method. Section 7 presents some simulation results. A general discussion is given in Section 8.

2. Notation and Outline

Assume the school size s is randomly distributed with the mean school size $\mu_s = \mathrm{E}(s)$. The total biomass density D_1 is defined as $N\mu_s/A$, where N is the total number of schools in the survey area A. To estimate the density D_1 and the mean school size μ_s, randomly allocated transect lines of total length L are traversed by observers to detect schools with maximum detection width w on both sides. Suppose that n schools are detected independently, with the perpendicular sighting distance and school size being recorded as $(x_1, s_1), \ldots, (x_n, s_n)$. We assume that $g(0, s) = 1$ for any $s > 0$, which means certain detection of a school on the transect whatever its size. Let $f(x, s)$ be the joint probability density function (p.d.f.) from which the sample is drawn. We assume that f satisfies the shoulder condition

$$\left.\frac{\partial f(x,s)}{\partial x}\right|_{x=0} = 0 \quad \text{and} \quad \left.\frac{\partial f(x,s)}{\partial s}\right|_{s=0} = 0. \tag{2.1}$$

Let p be the probability of detecting a school in the survey area. As the transect lines are randomly allocated, regardless of whether the schools are uniformly distributed or aggregated in the survey area, we have $\mathrm{E}(n) = Np$.

Following the derivations given by Drummer and McDonald (1987, p. 15, figures) and Quang (1991),

$$D_1 = (2L)^{-1}\,\mathrm{E}(n)\beta(0),$$

where $\beta(0) = \int_0^\infty sf(0, s)\,ds$. Let $\hat{\beta}(0)$ be an estimator for $\beta(0)$. Then, an estimator for D_1 is

$$\hat{D}_1 = (2L)^{-1}\,n\hat{\beta}(0). \tag{2.2}$$

Let $D_0 = N/2Lw = (2L)^{-1}\mathrm{E}(n)f_x(0)$ be the number of schools per unit area, where f_x is the marginal probability density function with respect to the sighting distance x. A generic estimator for D_0 has the form $\hat{D}_0 = (2L)^{-1}n\hat{f}_x(0)$. Various estimators for D_0 have been given by using different estimators for $f_x(0)$. See Seber (1982) and Buckland et al. (1993) for comprehensive reviews, and Burnham, Anderson, and Laake (1980), Buckland (1992), and Chen (1996) for nonparametric estimators.

Due to the size effect, the average size of the detected schools, $n^{-1}\Sigma_{i=1}^n s_i$, overestimates the mean school size μ_s. Since $\mu_s = D_1/D_0 = \beta(0)/f_x(0)$, μ_s should be estimated by

$$\hat{\mu}_s = \hat{\beta}(0)/\hat{f}_x(0). \tag{2.3}$$

From (2.2) and (2.3), we see that the critical part of establishing estimators for D_1 and μ_s is to estimate $\beta(0)$, since various estimators for $f_x(0)$ are available. In this paper we use the kernel method to estimate $\beta(0)$ by replacing $f(0, s)$ with a kernel estimator in $\beta(0) = \int_0^\infty sf(0, s)\,ds$.

3. Kernel Estimator for $\beta(0)$

A kernel estimator of the two dimensional p.d.f. f can be viewed as smoothing the histogram of two dimensional data. However, instead of counting the number of observations falling into each rectangular bin, we weight each data point by a two dimensional smoothing function centered at the data point. For an independent sample of $(x_1, s_1), \ldots, (x_n, s_n)$ drawn from f, a general kernel estimator of f at (x, s) is defined as

$$\hat{f}(x,s) = \frac{1}{n}\sum_{i=1}^{n}(h_{1i}h_{2i})^{-1}W\left(\frac{x-x_i}{h_{1i}}, \frac{s-s_i}{h_{2i}}\right), \tag{3.1}$$

where W is a kernel function that determines the shapes of the "bumps" centered at each data point, and h_{1i} and h_{2i} are the smoothing bandwidths controlling the amount of smoothing in each of the dimensions at data point (x_i, s_i).

There are basically two methods for choosing the bandwidths. One is to apply the same amount of smoothing at all data points, which means $h_{1i} = h_1$ and $h_{2i} = h_2$ for all i and some h_1 and h_2. The other is to vary the bandwidths at each data point. We shall discuss these two approaches in the next section. Before that, we use the general h_{1i} and h_{2i} in our derivation.

A simple way of defining the kernel W is to multiply a univariate kernel K with itself, that is

$$W\left(\frac{x-x_i}{h_{1i}}, \frac{s-s_i}{h_{2i}}\right) = K\left(\frac{x-x_i}{h_{1i}}\right)K\left(\frac{s-s_i}{h_{2i}}\right). \tag{3.2}$$

Usually the univariate kernel K is itself a p.d.f., as listed in Silverman (1986, p. 43). As Silverman (1986) reported, there is not much to choose between various kernels, as they all contribute about the same amount to the mean integrated square error. Therefore, we use only the product kernel (3.2) with $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ (the Gaussian kernel).

Some modifications have to be made before we apply the kernel estimator (3.1) to our line transect data. The problem arises because all the distances are nonnegative while the sizes are positive, that is $x_i \geq 0$ and $s_i > 0$. So $\hat{f}(x, s)$ should satisfy

$$\hat{f}(x,s) = 0 \quad \text{if} \quad x < 0 \quad \text{or} \quad s \leq 0. \tag{3.3}$$

To make the kernel estimator satisfy (3.3), we reflect the sample twice by replacing each data pair (x_i, s_i) by $(x_i, s_i), (-x_i, s_i), (x_i, -s_i)$ and $(-x_i, -s_i)$, which generalizes a proposal by Buckland (1992) for univariate sighting data. Thus, we obtain an extended sample

$$\begin{aligned} S_1 = \{&(x_1, s_1), (-x_1, s_1), (x_1, -s_1), (-x_1, -s_1), \ldots, \\ &(x_n, s_n), (-x_n, s_n), (x_n, -s_n), (-x_n, -s_n)\}. \end{aligned}$$

The reflection of x_i occurs because animal schools are sighted on both sides of the transect. If the "signed perpendicular sighting distances," where the signs indicate on which side of the transect the schools were detected, are recorded in the original sample, we need only to reflect the s_i. The data reflection makes the resulting kernel estimate asymptotically unbiased. The real argument behind this data reflection is that the p.d.f. f in line transect sampling usually has a wide shoulder near $x = 0$ and a flat valley near $s = 0$, as assumed by (2.1).

Applying the kernel estimator to the extended sample S_1, we have

$$\begin{aligned} \hat{f}(x,s) = \sum_{i=1}^{n}\frac{1}{nh_{1i}h_{2i}}\Bigg[&K\left(\frac{x-x_i}{h_{1i}}\right)\left\{K\left(\frac{s-s_i}{h_{2i}}\right) + K\left(\frac{s+s_i}{h_{2i}}\right)\right\} \\ &+ K\left(\frac{x+x_i}{h_{1i}}\right)\left\{K\left(\frac{s-s_i}{h_{2i}}\right) + K\left(\frac{s+s_i}{h_{2i}}\right)\right\}\Bigg] \end{aligned} \tag{3.4}$$

for $x \geq 0$ and $s > 0$, and $\hat{f}(x, s) = 0$ for $x < 0$ or $s \leq 0$. Since K is symmetric,

$$\hat{f}(0,s) = \sum_{i=1}^{n}\frac{2}{nh_{1i}h_{2i}}K\left(\frac{x_i}{h_{1i}}\right)\left\{K\left(\frac{s-s_i}{h_{2i}}\right) + K\left(\frac{s+s_i}{h_{2i}}\right)\right\}.$$

Then the kernel estimator for $\beta(0)$ is given by

$$\begin{aligned} \hat{\beta}(0) &= \int_0^{+\infty} s\hat{f}(0,s)\,ds = \sum_{i=1}^{n}\frac{2}{nh_{1i}h_{2i}}K\left(\frac{x_i}{h_{1i}}\right)\int_0^{+\infty} s\left\{K\left(\frac{s-s_i}{h_{2i}}\right) + K\left(\frac{s+s_i}{h_{2i}}\right)\right\}ds \\ &= \sum_{i=1}^{n}\frac{4}{nh_{1i}}K\left(\frac{x_i}{h_{1i}}\right)\left\{s_i\int_0^{s_i/h_{2i}} K(u)\,du + h_{2i}\int_{s_i/h_{2i}}^{+\infty} uK(u)\,du\right\}. \end{aligned}$$

As the Gaussian kernel is used, let ϕ and Φ denote the density and the distribution functions for an N(0, 1) random variable, respectively, then

$$\hat{\beta}(0) = \sum_{i=1}^{n} \frac{4}{nh_{1i}} \phi(x_i/h_{1i}) \big[s_i\{\Phi(s_i/h_{2i}) - 1/2\} + h_{2i}\phi(s_i/h_{2i})\big]. \tag{3.5}$$

Since h_{2i} is usually relatively small when L is large, we see $\hat{\beta}(0)$ is basically a weighted average of the school sizes. This reflects the meaning of $\beta(0) = \mathrm{E}\{sf(0 \mid s)\}$, which is the expected school size weighted by the conditional probability density, given the school size, as noted in Quang (1991).

4. Fixed and Adaptive Kernel Estimators

Two kernel estimators for D_1 are introduced based on two approaches to select the smoothing bandwidths h_{1i} and h_{2i} at (x_i, s_i). The first one is the fixed kernel method, which uses the same bandwidths for smoothing at every data point. The second one is the adaptive kernel method, which allows the bandwidths to vary from one data point to another. Both methods are conditional on the sample size n.

4.1 *Fixed Kernel Estimator*

The fixed kernel method uses the same bandwidths at all data points, that is $h_{1i} = h_1$ and $h_{2i} = h_2$ for $1 \le i \le n$. The common bandwidths h_1 and h_2 are determined by minimizing the mean integrated squared error (MISE) of $\hat{f}$, that is

$$\mathrm{MISE}(\hat{f};\, h_1, h_2) = \int \mathrm{E}\{\hat{f}(x,s) - f(x,s)\}^2 \, dx\, ds.$$

The least-squares cross-validation (LSCV) and the reference to a standard distribution are two popular methods for finding the optimal h_1 and h_2 that minimize the above MISE. General descriptions of the two methods are given in Silverman (1986); those related to line transect sampling are available in Chen (1996).

While the LSCV method provides generally robust estimates, it is computationally involved. In contrast, the reference to a standard distribution method is simpler and performs reasonably well if the assumed reference distribution is not too far away from the real underlying distribution. If f is close to a bivariate normal distribution, according to Scott (1992, p. 151),

$$h_1 = \sigma_x (1 - \rho^2)^{5/2} (1 + \rho^2/2)^{-1/6} n^{-1/6}$$

and

$$h_2 = \sigma_s (1 - \rho^2)^{5/2} (1 + \rho^2/2)^{-1/6} n^{-1/6}, \tag{4.1}$$

where σ_x and σ_s are the standard deviations of x and s, respectively, and ρ is the correlation coefficient between x and s. For line transect data, ρ^2 is usually small. Thus, one can simply choose $h_1 = \sigma_x n^{-1/6}$ and $h_2 = \sigma_s n^{-1/6}$. However, we use (4.1) for the sake of generality. For practical use, we just replace σ_x, σ_s, and ρ by their sample estimates in (4.1). Only the original sample, not the reflected sample, is used in computing the above sample estimates. According to (3.5), the fixed kernel estimator for $\beta(0)$ is

$$\hat{\beta}_f(0) = \frac{4}{nh_1} \sum_{i=1}^{n} \phi(x_i/h_1) \big[s_i\{\Phi(s_i/h_2) - 1/2\} + h_2\phi(s_i/h_2)\big].$$

And the fixed kernel estimators for D_1 and μ_s are

$$\hat{D}_{1f} = (2L)^{-1} n \hat{\beta}_f(0) \quad \text{and} \quad \hat{\mu}_{sf} = \hat{\beta}_f(0)/\hat{f}_{xf}(0),$$

where $\hat{f}_{xf}(0)$ is a fixed kernel estimator for $f_x(0)$ and is available in Chen (1996).

A derivation deferred to Appendix 1 shows that

$$\mathrm{E}(\hat{D}_{1f}) = D_1 + bL^{-1/3} + O(L^{-2/3}), \tag{4.2}$$

where b is some constant. Thus, $\hat{D}_1$ is an asymptotically unbiased estimator of D_1. Its dominant bias term is of order $L^{-1/3}$. When L is not too large, a bias-corrected estimator could be proposed by estimating the coefficient b. However, it would involve estimating the second order partial derivatives of f. We shall show shortly that the adaptive kernel method makes this explicit bias correction unnecessary.

The bandwidths given by (4.1) are based on the global MISE criterion which aims to give a good fit for the entire density surface. However, in the estimation of $\beta(0)$, it is sufficient to fit $f(0,s)$, which is just a "slice" of the entire density surface. Therefore, a local bandwidth selection may be more appropriate. This leads us to consider the following adaptive kernel method.

4.2 *Adaptive Kernel Estimator*

The adaptive kernel method (Abramson, 1982; Silverman, 1986) uses different bandwidths in different parts of the density surface; larger bandwidths are used in low density areas, and smaller ones in high density areas. To decide whether a data point is in a high or low density area, a pilot-fixed kernel density estimate, say $\tilde{f}$, is computed. Then, a local smoothing weight at the ith data point, say λ_i, is defined as $\lambda_i = \{\tilde{f}(x_i, s_i)/G\}^{-1/2}$, where G is the geometric mean of $\tilde{f}(x_i, s_i)$. Clearly, G is just a scaling factor to make the geometric mean of λ_i equal to 1. One way of choosing the bandwidths at (x_i, s_i) is

$$h_{1i} = \lambda_i h_1 \quad \text{and} \quad h_{2i} = \lambda_i h_2, \tag{4.3}$$

where h_1 and h_2 are some fixed kernel bandwidths based on a criterion related to the global fit of the density surface.

An adaptive kernel estimator for $f(0,s)$ has the form

$$\hat{f}(0,s) = 2n^{-1} \sum_{i=1}^{n} \lambda_i^{-2} h_1^{-1} h_2^{-1} K\left(\frac{x_i}{\lambda_i h_1}\right) \left\{ K\left(\frac{s - s_i}{\lambda_i h_2}\right) + K\left(\frac{s + s_i}{\lambda_i h_2}\right) \right\}.$$

It has been shown that the adaptive kernel estimate is relatively insensitive to the pilot estimate $\tilde{f}$ (Silverman, 1986, p. 101). Therefore, $\tilde{f}$ can be constructed by using just the bandwidth given in (4.1). After the λ_i are obtained, LSCV could be used to determine h_1 and h_2 in (4.3). Silverman (1986) suggested using (4.1) again as an "ad hoc" way of choosing h_1 and h_2. Simulation in Section 7 shows that this choice is satisfactory.

The adaptive kernel estimator for $\beta(0)$ is

$$\hat{\beta}_a(0) = \frac{4}{n} \sum_{i=1}^{n} \lambda_i^{-1} h_1^{-1} \phi\{x_i/(\lambda_i h_1)\} \Big[s_i \big[\Phi\{s_i/(\lambda_i h_2)\} - 1/2\big] + \lambda_i h_2 \phi\{s_i/(\lambda_i h_2)\} \Big]. \tag{4.4}$$

The corresponding adaptive estimator for D_1 is

$$\hat{D}_{1a} = (2L)^{-1} n \hat{\beta}(0).$$

A derivation deferred to Appendix 2 shows that

$$\mathrm{E}(\hat{D}_{1a}) = D_1 + o(L^{-1/3}). \tag{4.5}$$

Comparing (4.2) and (4.5), we see that the asymptotic bias of $\hat{D}_{1a}$ is only slightly smaller than its fixed kernel counterpart $\hat{D}_{1f}$. The derivation is based on the following latest result on adaptive kernal estimation: if f has an exponentail tail, which is very typical for f in a line transect survey, by extending Theorem 2.2 of Hall, Hu, and Marron (1995) to multivariate situations, we may have

$$\mathrm{E}\{\hat{f}(0,s)|n\} = f(0,s) + O\left(\left\{ \frac{h_1}{\log(h_1)} \right\}^2 + \left\{ \frac{h_2}{\log(h_2)} \right\}^2 \right) + O(h_1^4 + h_2^4). \tag{4.6}$$

The above result is a correction to some early results which claimed that the bias was $O(h_1^4 + h_2^4)$ and ignored the first bias term in (4.6). As explained by Terrell and Scott (1992), the first bias term in (4.6) is contributed by the tails of f. This is because if (x_i, s_i) is in the tail area of f, $\tilde{f}(x_i, s_i)$ will have a very small value and h_i a very big value. The value of h_i can be so big that the bias term is not $O(h_1^4 + h_2^4)$.

However, as observed by Terrell and Scott (1992) for small sample sizes, the influence of the tail appears to be negligible and the bias looks to be $O(h_1^4 + h_2^4)$ in many cases. They reported that for a normal density function f, the adaptive estimator for f has significantly smaller mean integrated squared error than its fixed kernel counterpart when $n < 500$. As the sample sizes from a line transect survey are not likely to be large, the adaptive method will still be advantageous. This is confirmed by our simulation study in Section 7.

For large samples, the adaptive kernel estimator in its current form may not have any real advantages over the fixed kernel estimator due to the tail effect. However, the tail effect can be eliminated by either clipping the pilot density estimate $\tilde{f}$ away from zero or truncating the kernel functions. Interested readers should referr to Abramson (1982), Terrell and Scott (1992), and Hall et al. (1995) for modified adaptive kernel estimation.

To obtain an adaptive estimator for the mean school size $\mu_s = \beta(0)/f_x(0)$, we have to construct an adaptive kernel estimator for $f_x(0)$ based on the univariate sighting distances $x_1, \ldots, x_n$. Similar to the bivariate adaptive kernel estimation, a pilot estimator $f_x^\dagger$ for the univariate p.d.f. f_x is constructed with a bandwidth $h = \hat{\sigma}_x n^{-1/5}$. The local smoothing weight ν_i, which is the equivalent of λ_i, is defined as $\nu_i = \{f_x^\dagger(x_i)/q\}^{-1/2}$, where q is the geometrical mean of $f_x^\dagger(x_i)$. An adaptive estimator for $f_x(0)$ is

$$\hat{f}_{xa}(0) = n^{-1} \sum_{i=1}^{n} \nu_i^{-1} h^{-1} K(\frac{x_i}{\nu_i h}),$$

where the same Gaussian kernel K is used here. Then, the adaptive kernel estimator for μ_s is $\hat{\mu}_{sa} = \hat{\beta}_a(0)/\hat{f}_{xa}(0)$.

5. Variance Estimation and Confidence Intervals

Let $\hat{\beta}(0)$ and $\hat{D}_{1k}$ denote either the fixed or adaptive kernel estimators for $\beta(0)$ and D_1, respectively. Because $\mathrm{E}\{n\hat{\beta}(0)\} \approx \mathrm{E}(n)\mathrm{E}\{\hat{\beta}(0)\}$, the correlation between n and $\hat{\beta}(0)$ is very small (Buckland et al., 1993, p. 53). If the wildlife population is uniformly distributed, n is basically Poisson distributed, that is $\mathrm{var}(n) = \mathrm{E}(n)$. Then, an estimate for $\mathrm{var}(\hat{D}_{1k})$ is

$$\widehat{\mathrm{var}}(\hat{D}_{1k}) = \hat{D}_{1k}^2 \left[\frac{\widehat{\mathrm{var}}\{\hat{\beta}(0)\}}{\hat{\beta}^2(0)} + n^{-1}\right], \tag{5.1}$$

where $\widehat{\mathrm{var}}\{\hat{\beta}(0)\}$ estimates the variance of $\hat{\beta}(0)$.

If the wildlife population tends to aggregate, n is no longer Poisson distributed. Usually data show that $\mathrm{var}(n) = a\mathrm{E}(n)$ for some $a > 1$. In this case, (5.1) becomes

$$\widehat{\mathrm{var}}(\hat{D}_{1k}) = \hat{D}_{1k}^2 \left[\frac{\widehat{\mathrm{var}}\{\hat{\beta}(0)\}}{\hat{\beta}^2(0)} + \widehat{cv}^2(n)\right], \tag{5.2}$$

where $\widehat{cv}(n)$ estimates the coefficient of variation of n. To estimate $cv(n)$, we need replicate line information on the number of sightings made from each transect line and its length. The formula given in Buckland et al. (1993, p. 90) can be used to compute $\widehat{cv}(n)$.

To use (5.1) or (5.2), we need to estimate $\mathrm{var}\{\hat{\beta}(0)\}$. But, an analytic expression for $\mathrm{var}\{\hat{\beta}(0)\}$ is difficult to obtain. This is because $\hat{\beta}(0)$ is not an average of independent and identically distributed random variables, due to the bandwidths being estimated from the data. To avoid this difficulty, we suggest using a bootstrap method to estimate $\mathrm{var}\{\hat{\beta}(0)\}$. One primary motivation for the bootstrap, as indicated by Efron (1982), is to estimate the variance of a statistic by Monte Carlo simulation when its analytic solution is difficult to obtain. We give a bootstrap algorithm for estimating $\mathrm{var}\{\hat{\beta}_a(0)\}$ only; that for $\mathrm{var}\{\hat{\beta}_f(0)\}$ is easier and can be worked out in a similar way.

Step 1. Generate B independent bootstrap resamples of the original sighting sample $\{(x_i, s_i)\}_{i=1}^n$. Denote the bth resample as $\{(x_i^{\star b}, s_i^{\star b})\}_{i=1}^n$ for $b = 1, \ldots, B$. Let $\sigma_x^{\star b}$, $\sigma_y^{\star b}$, and $\rho^{\star b}$ be estimates of σ_x, σ_y, and ρ, respectively, from the bth resample. Put $\gamma^{\star b} = (1 - \rho^{\star b^2})^{5/2} (1 + \rho^{\star b^2}/2)^{-1/6}$.

Step 2. According to (4.1), use $h_1^{\star b} = \sigma_x^{\star b} \gamma^{\star b} n^{-1/6}$ and $h_2^{\star b} = \sigma_y^{\star b} \gamma^{\star b} n^{-1/6}$ to construct a pilot kernel density estimate $\tilde{f}^{\star b}$ and compute the weight $\lambda_i^{\star b}$. Then, choose the adaptive bandwidths for smoothing at $(x_i^{\star b}, s_i^{\star b})$ to be $h_{1i}^{\star b} = \lambda_i^{\star b} h_1^{\star b}$ and $h_{2i}^{\star b} = \lambda_i^{\star b} h_2^{\star b}$.

Step 3. Calculate the adaptive estimate $\hat{\beta}_a^{\star b}(0)$ for $\beta(0)$ for the bth resample according to (4.5) with the bandwidths obtained in step 2.

Step 4. Repeat steps 2 and 3 for $b = 1, \ldots, B$, and obtain $\{\hat{\beta}_a^{\star b}(0)\}_{b=1}^B$.

The bootstrap estimate of $\text{var}\{\hat{\beta}_a(0)\}$ is

$$\widehat{\text{var}}\{\hat{\beta}_a(0)\} = \frac{1}{B-1}\sum_{b=1}^{B}\{\hat{\beta}_a^{\star b}(0) - \overline{\hat{\beta}_a}(0)\}^2,$$

where $\overline{\hat{\beta}_a}(0) = B^{-1}\Sigma_{b=1}^{B}\hat{\beta}_a^{\star b}(0)$. Substitute the above bootstrap estimate for $\widehat{\text{var}}\{\hat{\beta}_a(0)\}$ in (5.1) or (5.2) to obtain $\widehat{\text{var}}(\hat{D}_{1a})$.

The derivation given in Appendix 3 shows that $\hat{D}_{1a}$ is asymptotically normally distributed. Thus, an α-level confidence interval for D_1 can be constructed as

$$\left(\hat{D}_{1a} - z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{D}_{1a})},\ \hat{D}_{1a} + z_{\alpha/2}\sqrt{\widehat{\text{var}}(\hat{D}_{1a})}\right),$$

where $z_{\alpha/2}$ is the $\alpha/2$ upper percentile point of the standard normal distribution.

Because $\hat{\mu}_{sa}$ is the ratio of $\hat{\beta}_a(0)$ to $\hat{f}_{xa}(0)$, it seems that the bootstrap approach is the only reliable way to estimate the variance of $\hat{\mu}_{sa}$. The delta method could be used to approximate the variance, but often has a large bias. A bootstrap procedure for calculating $\widehat{\text{var}}(\hat{\mu}_{sa})$ can be set up by slightly modifying the bootstrap procedure for evaluating $\widehat{\text{var}}\{\hat{\beta}(0)\}$. We only need to include the computation of the adaptive kernel estimate for $f_x(0)$ for each of the bootstrap resamples.

6. An Example

In this section we apply the kernel estimators developed in earlier sections to a data set from the aerial survey of southern bluefin tuna. The data set, which is shown in Figure 1, was collected in January 1993 with sample size $n = 119$ and total search length $L = 2782.83$ miles. The sample mean school size was 115.50 tonnes. Due to the large scale of the school size, a log transform was used to rescale the size data in Figure 1.

We see a basically monotonic decrease in the probability of detection with respect to the distance and a single mode of school size as shown by Figure 1(1) and Figure 1(2) respectively. Both figures show a wide shoulder near $x = 0$ and a flat valley near $s = 0$. Figure 1(3) shows that no small schools were detected when the distances were larger than 10 miles. This is a clear indication of the size effect. It is worth mentioning that the survey may be subject to violation of the assumption $g(0, s) = 1$. However, the histogram in Figure 1(1) shows that the violation would not be a severe one.

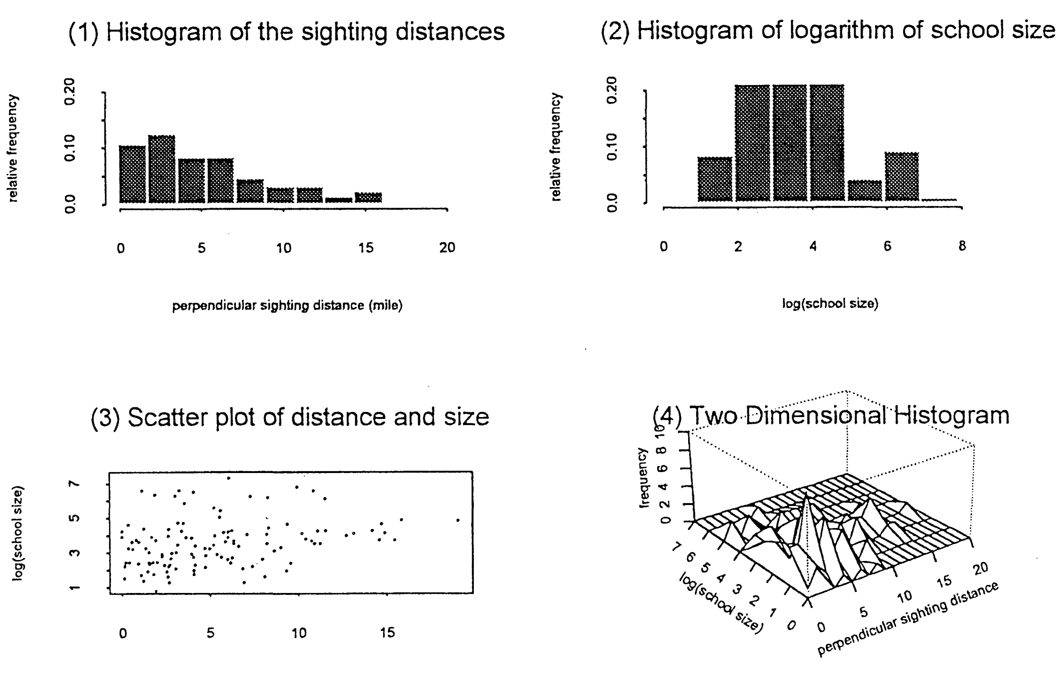

Figure 1. The aerial survey data set of southern bluefin tuna.

There were 17 replicate transect lines. The number of sightings on each transect line enabled us to estimate $cv(n)$. We have $\widehat{cv}^2(n) = 0.053$, which is much larger than $1/119 = 0.0084$ under the Poisson model. Therefore, (5.2) is used to compute the variance of the kernel estimates for D_1.

Using (4.1), we have the two bandwidths $h_1 = 1.86$ and $h_2 = 21.64$ for constructing the fixed kernel estimate for $\beta(0)$, and a bandwidth $h = 1.59$ to estimate $f_x(0)$ by reference to a normal distribution. The above h_1, h_2, and h are also used to construct pilot estimates for the two dimensional p.d.f. f and the one dimensional distance-only p.d.f. f_x. The resulting fixed and adaptive kernel estimates and 95% confidence intervals for D_1 and μ_s are displayed in Table 1. We also include results from the FS method and the parametric model (7.1) of Drummer and McDonald (1987) for comparison.

The results show that the fixed kernel and the parametric methods gave similar point estimates for D_1 and μ_s. They were all smaller than the adaptive and FS estimates. The smaller estimates given by the fixed kernel estimators may be due to their bias as revealed both by the analysis given in Section 4 and the simulation results presented in Section 7. The FS estimates have larger CVs and wider confidence intervals than the adaptive kernel method; for example, the FS estimate of school size has an estimated CV of 46%, compared to 29% for the adaptive estimate.

7. Simulation Results

In this section we present some simulation results designed to investigate the performance of the kernel estimators and confidence intervals proposed in the previous sections. We generated $N = 300$ random points within a $L \times 2w$ rectangular "survey region" with $w = 10$. We chose the transect length L such that $D_1 = 1.0$. Each point simulates the position of a school. As the detected school size has a usually skewed distribution on $[0, \infty)$, each school size was randomly generated from the χ_5^2 distribution. Thus, the mean school size $\mu_s = 5$. To detect each simulated school we used the following exponential power series detection function, proposed by Drummer and McDonald (1987):

$$g(x, s) = \exp\{-(bx/s^c)^a\}. \tag{7.1}$$

It is easy to see that (7.1) was set up by inserting the size covariate s into a distance-only parametric detection function, where c is a parameter controlling the size effect and a is a shape parameter. We fixed $b = 0.5$ and varied the shape and size parameters $a = 1.5, 2.0$, and 2.5 and $c = 0.2$ and 0.6. The simulation results shown in this section were all based on 1000 simulations and 499 bootstrap resamples.

Table 2 gives point estimates for D_1 and μ_s together with their standard errors and mean squared errors (MSE) by using the fixed kernel, the adaptive kernel, and the FS methods. Tables 3 and 4 give coverages and lengths of the bootstrap confidence intervals for both kernel estimators for D_1 and μ_s, respectively. Results for the bootstrap FS confidence intervals are also contained in Table 3.

In summary, we observe that:

(1) the adaptive kernel method produced quite satisfactory point estimates for D_1; it had much smaller bias than the fixed kernel estimates, which may be due to the not very large sample sizes used in the simulation, and the tail effect on the bias was still negligible;

Table 1

Estimates of tuna biomass density D_1 and the mean school size μ_s, together with their standard errors (Std. error) and 95% confidence intervals. The subscripts f, a, FS, and p on an estimate are the fixed kernel, adaptive kernel, Fourier series, and Drummer and McDonald's parametric estimates respectively.

	Estimate	Std. error	95% confidence interval
$\hat{D}_{1f}$	177.2	80.9	(15.3, 339.1)
$\hat{D}_{1a}$	225.0	85.3	(57.8, 392.1)
$\hat{D}_{1FS}$	247.0	97.3	(56.2, 437.8)
$\hat{D}_{1p}$	187.8	71.1	(48.4, 327.2)
$\hat{\mu}_{sf}$	71.9	27.5	(16.3, 124.2)
$\hat{\mu}_{sa}$	89.5	25.9	(38.7, 140.4)
$\hat{\mu}_{sFS}$	91.1	42.7	(7.41, 174.8)
$\hat{\mu}_{sp}$	75.6	18.0	(40.3, 110.9)

Table 2

Point estimates, standard error (Std. error) and mean square error (MSE) for density D_1 and μ_s with exponential series detection function $g(x, s) = \exp\{(bx/s^c)^a\}$, $b = 0.5$, and $w = 10$. The subscripts f, a, and FS denote the fixed kernel, the adaptive kernel, and the FS estimates, respectively. The true density $D_1 = 1.0$ and $\mu_s = 5.0$.

	$a = 1.5$		$a = 2.0$		$a = 2.5$	
	$c = 0.2$	$c = 0.6$	$c = 0.2$	$c = 0.6$	$c = 0.2$	$c = 0.6$
$\hat{D}_{1f}$	0.854	0.880	0.914	0.925	0.944	0.950
Std. error	0.172	0.127	0.176	0.128	0.175	0.128
MSE	0.051	0.031	0.038	0.022	0.034	0.019
$\hat{D}_{1a}$	0.941	0.964	0.987	0.999	1.001	1.014
Std. error	0.197	0.148	0.199	0.148	0.197	0.147
MSE	0.043	0.023	0.040	0.022	0.039	0.022
$\hat{D}_{FS}$	0.945	0.949	1.003	0.988	1.029	1.010
Std. error	0.224	0.171	0.213	0.163	0.205	0.155
MSE	0.053	0.032	0.045	0.027	0.043	0.024
$\hat{\mu}_{sf}$	5.10	5.29	5.09	5.26	5.08	5.23
Std. error	0.502	0.395	0.502	0.391	0.498	0.384
MSE	0.262	0.240	0.260	0.220	0.254	0.200
$\hat{\mu}_{sa}$	4.95	5.09	4.98	5.10	5.01	5.05
Std. error	0.523	0.422	0.525	0.423	0.522	0.418
MSE	0.276	0.178	0.276	0.189	0.272	0.179
$\hat{\mu}_{sFS}$	4.96	5.05	4.99	5.07	5.02	5.06
Std. error	0.747	0.649	0.662	0.618	0.68	0.590
MSE	0.557	0.424	0.438	0.387	0.463	0.352
Ave. n	72.4	127.1	70.9	129.0	71.0	130.9

(2) the two kernel estimates had much lower MSE than the FS estimates; the MSE of the fixed kernel estimates were less than those of the adaptive kernel estimates for μ_s when the school size parameter c was small (0.2), and for D_1 when f had a larger shoulder ($a = 2.5$); the opposite was observed for μ_s when c was large (0.6) and for D_1 when the shoulder was small ($a = 1.5$).

(3) the adaptive confidence intervals for both D_1 and μ_s had better coverages and were shorter than their fixed kernel and FS counterparts;

(4) the fixed kernel estimators gave estimates with smaller variance which explained the observations that they had shorter confidence intervals and smaller MSE in some cases mentioned in (2);

(5) improvements in the kernel estimates and confidence intervals for D_1 are evident when the shape parameter a increases, which increases the smoothness of f near $x = 0$; while for μ_s, the size parameter c plays a more important role than the shape parameter a.

Table 3

Coverages and lengths of the two kernel and FS confidence intervals for density D based on a bootstrap variance estimation with 0.95% nominal coverage. The average sample sizes used are those given in Table 2.

	$a = 1.5$		$a = 2.0$		$a = 2.5$	
	$c = 0.2$	$c = 0.6$	$c = 0.2$	$c = 0.6$	$c = 0.2$	$c = 0.6$
Fixed kernel	0.877	0.895	0.924	0.937	0.942	0.953
Length	0.771	0.594	0.798	0.603	0.805	0.604
Adap. kernel	0.914	0.931	0.942	0.965	0.956	0.960
Length	0.830	0.625	0.855	0.633	0.860	0.631
FS	0.900	0.910	0.931	0.987	0.985	0.987
Length	0.886	0.707	0.870	0.770	0.990	0.768

Table 4

Coverages and lengths of the two kernel confidence intervals for the mean school size μ_s based on a bootstrap variance estimation with 0.95% nominal coverage. The average sample sizes used are those given in Table 2.

	$a = 1.5$		$a = 2.0$		$a = 2.5$	
	$c = 0.2$	$c = 0.6$	$c = 0.2$	$c = 0.6$	$c = 0.2$	$c = 0.6$
Fixed kernel	0.928	0.932	0.927	0.933	0.931	0.936
Length	2.02	1.59	2.04	1.58	2.05	1.56
Kernel	0.942	0.942	0.947	0.939	0.951	0.948
Length	2.08	1.62	2.09	1.62	2.10	1.60

It is interesting to see that the bootstrap technique substantially improves the coverage of the FS confidence interval, compared with the nonbootstrap results reported in Quang (1991) and reproduced by this author. This again confirms that the bootstrap offers more accuracy in the variance estimation.

8. Discussion

We have seen that kernel smoothing is a useful nonparametric tool for analyzing line transect data when both the sighting distance and school size are factors determining the detection of cluster schools. The adaptive kernel method produces sensible estimates and confidence intervals for both D_1 and μ_s. The fixed kernel method gives estimates with smaller variance and shorter confidence intervals for μ_s than those produced by its adaptive kernel counterpart. But, it also produces a larger bias in both the point estimate and the confidence level. The adaptive estimator is subject to the tail effect when the sample size is large. After balancing these findings, we would recommend that the adaptive kernel method be used for inference on both D_1 and μ_s when the sample size is not too large. However, when the sample size is large or the variation of the estimate is of concern, the fixed kernel estimate should be used.

The kernel estimates generally require the p.d.f. f of the data to have a wide shoulder near $x = 0$ and a flat valley near $s = 0$. When f is not very smooth near $x = 0$, for example when $c = 1.5$ in the simulation presented in Section 7, the kernel methods produce biased estimates and confidence intervals with coverage less than nominal, as does the FS method. This is because when a highly smoothed Gaussian kernel or cos( ) for the FS method is used to model a p.d.f. that is not very smooth near $x = 0$, some bias is inevitable. To remedy this, a less smoothed kernel may be used instead. Further research is needed in this area.

To make the method developed in this paper of immediate use for practitioners, computer software has been developed and can be obtained by contacting the author.

Acknowledgements

The author would like to thank Drs B. M. Brown and T. Polacheck for interesting discussions, and two referees and the Associate Editor for constructive comments which led to substantial improvement of the paper. The author is grateful to one of the referees, who has brought the latest results on adaptive kernel methods to the author's attention.

Résumé

Quand on effectue un échantillonnage le long d'un transect pour estimer l'abondance d'une population sauvage vivant en groupes, la détection d'un groupe dépend non seulement de la distance perpandiculaire du groupe au transect, mais aussi de la taille du groupe. Les groupes de tailles les plus importantes sont plus faciles à détecter que les plus petits. Alors on peut envisager une fonction de détection à deux variables, la distance et la taille. Cet article utilise une méthode de lissage par noyau pour ajuster les données bivariées du transect afin d'estimer à la fois l'abondance et la taille moyenne des groupes. On étudie deux estimateurs par noyau: l'estimateur à noyau fixe, qui utilise la même largeur de fenêtre pour tous les points; et l'estimateur à noyau adaptatif qui permet de faire varier la largeur de la fenêtre suivant les points.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates—A square root law. *The Annals of Statistics* **10,** 1217–1223.

Buckland, S. T. (1992). Fitting density functions with polynomials. *Applied Statistics* **41,** 63–76.

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 Monographs* **72.**

Chen, S. X. (1996). Kernel estimates for density of a biological population using line transect sampling. *Applied Statistics* **45,** 135–150.

Drummer, T. D. and McDonald, L. L. (1987). Size bias in line transect sampling. *Biometrics* **44,** 13–21.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: SIAM.

Hall, P. G., Hu, T. C., and Marron, J. S. (1995). Improved variable window kernel estimates of probability densities. *The Annals of Statistics* **23,** 1–11.

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation.* Orlando, Florida: Academic Press.

Quang, P. X. (1991). A nonparametric approach to size-based line transect sampling. *Biometrics* **47,** 269–279.

Quang, P. X. (1993). Nonparametric estimators for variable circular plot surveys. *Biometrics* **49,** 837–852.

Scott, D. W. (1992). *Multivariate Density Estimation.* New York: Wiley.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance.* London: Griffin.

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

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

Terrell, G. R. and Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics* **20,** 1236–1265.

Received March 1995; *revised October* 1995 *and February* 1996; *accepted April* 1996.

Appendix

1. Derivation of (4.2)

Let $\sigma_k^2 = \int t^2 K(t)\,dt$, $R(K) = \int K(t)^2\,dt$, and f_{11} and f_{22} be the second partial derivatives of f with respect to x and s, respectively. We first derive the mean of $\hat{f}(0,s)$ conditioning on the sample size n. According to standard kernel theory, for example Scott (1992, p. 150),

$$\mathrm{E}\{\hat{f}(0,s)\,|\,n\} = f(0,s) + \frac{1}{2}\,\sigma_K^2\{h_1^2 f_{11}(0,s) + h_2^2 f_{22}(0,s)\} + O(n^{-2/3}). \tag{A.1}$$

We assume: (1) all the conditions in the Fubini Theorem hold which allow changing the order of integrations; (2) $T_i = \int_0^\infty s f_{ii}(0,s)\,ds < \infty$ for $i = 1$ and 2; (3) $\int_0^\infty s f_{iijj}(0,s)\,ds$ are bounded for the fourth-order partial derivatives $f_{iijj}(0,s)$ where $i,j = 1$ or 2. Then, a similar derivation to Prakasa Rao (1983, p. 45) may show that

$$\begin{aligned}\mathrm{E}\{\hat{\beta}_f(0)\,|\,n\} &= \int_0^\infty s\mathrm{E}\{\hat{f}(0,s)\,|\,n\}\,ds \\ &= \beta(0) + \frac{1}{2}\,\sigma_K^2(h_1^2 T_1 + h_2^2 T_2) + O(n^{-2/3}).\end{aligned}$$

With h_1 and h_2 given by (4.1) and defining $t(\rho) = (1-\rho^2)^5\,(1+\rho^2/2)^{-1/3}$,

$$\begin{aligned}\mathrm{E}(\hat{D}_{1f}) &= \mathrm{E}\{n\hat{\beta}_f(0)/2L\} \\ &= D_1 + \frac{1}{4}\,\sigma_K^2 t(\rho)\,(\sigma_x^2 T_1 + \sigma_s^2 T_2)\mathrm{E}(n^{2/3})/L + O\{\mathrm{E}(n^{1/3})/L\}.\end{aligned} \tag{A.2}$$

It may be shown that $p = 1/\{f(0)w\}$ and $\mathrm{E}(n^l) \to (Np)^l = (2LD_1 p/\mu_s)^l$ for $l = 1/3$ and $2/3$, respectively. Therefore,

$$\mathrm{E}(\hat{D}_{1f}) = D_1 + bL^{-1/3} + O(L^{-2/3}),$$

where $b = 1/4\,\sigma_K^2(\sigma_x^2 T_1 + \sigma_y^2 T_2)\,(1-\rho^2)^5\,(1+\rho^2/2)^{-1/3}\big[2D_1/\{f_x(0)w\mu_s\}\big]^{2/3}$.

2. Derivation of (4.5)

The derivation of (4.5) follows the route shown in Appendix 1. It may be shown that the adaptive kernel estimator $\hat{f}(0,s)$ for $f(0,s)$ has a bias which is a smaller order than h_1^2 and h_2^2, by extending Theorems 2.1 and 2.2 of Hall et al. (1995) to multivariate situations. In particular, if f has roughly an exponential tail, which is a typical form for f from a line transect survey,

$$\mathrm{E}\{\hat{f}(0,s)\,|\,n\} = f(0,s) + O\left(\left\{\frac{h_1}{\log(h_1)}\right\}^2 + \left\{\frac{h_2}{\log(h_2)}\right\}^2\right) + O(h_1^4 + h_2^4).$$

Since $h_i = O(n^{-1/6})$ according to (4.2), we have

$$\mathrm{E}\{\hat{f}(0,s)\,|\,n\} = f(0,s) + o(n^{-1/3}).$$

Then, by following the same derivation of (A.2), we have

$$\mathrm{E}(\hat{D}_{1a}) = D_1 + o(L^{-1/3}).$$

3. Asymptotic Normality of D_{1a}

Remember $\hat{D}_{1a} = (2L)^{-1} n\hat{\beta}_a(0)$ and $\hat{\beta}_a(0) = n^{-1}\Sigma_{i=1}^n W_i$, where

$$W_i = 4\lambda_i^{-1}h_1^{-1}\phi\{x_i/(\lambda_i h_1)\}\Big[s_i\big[\Phi\{s_i/(\lambda_i h_2)\} - 1/2\big] + \lambda_i h_2 \phi\{s_i/(\lambda_i h_2)\}\Big]. \qquad \text{(A.3)}$$

We treat G in $\lambda_i = \{\tilde{f}(x_i, s_i)/G\}^{-1/2}$ as a constant since it is just a scaling factor. From the theory of kernel estimation, under the condition that f has bounded second derivatives,

$$\lambda_i = \lambda_{i0} + O_p(n^{-5/12}) \quad \text{uniformly for } i = 1, \ldots, n, \qquad \text{(A.4)}$$

where $\lambda_{i0} = \{f(x_i, s_i)/G\}^{-1/2}$. Define h_{i0} for $i = 1$ and 2 by replacing the sample estimates of σ_x, σ_s, and ρ with their true values in h_i. Then, it can be shown that, for $i = 1$ and 2,

$$h_i = h_{i0} + O_p(n^{-7/12}), \qquad \text{(A.5)}$$

provided the sample estimates are $\sqrt{n}$ consistent. Substituting (A.4) and (A.5) into (A.3), and noting that Φ and ϕ have bounded derivatives, we have

$$W_i = W_{i0} + O_p(n^{-5/12}) \quad \text{uniformly for } i = 1, \ldots, n, \qquad \text{(A.6)}$$

where

$$W_{i0} = 4\lambda_{i0}^{-1}h_{10}^{-1}\phi\{x_i/(\lambda_{i0} h_{10})\}\Big[s_i\big[\Phi\{s_i/(\lambda_{i0} h_{20})\} - 1/2\big] + \lambda_{i0} h_2 \phi\{s_i/(\lambda_{i0} h_{20})\}\Big].$$

The uniformity in (A.6) ensures that $n^{-1}\Sigma_{i=1}^n (W_i - W_{i0}) = O_p(n^{-5/12})$. Since $n \to \infty$ when $L \to \infty$, we have

$$n^{-1}\sum_{i=1}^{n}(W_i - W_{i0}) \to 0 \quad \text{in probability as } L \to \infty. \qquad \text{(A.7)}$$

We assume there is a sequence of positive constants $\{a_L\}$ tending to ∞ as $L \to \infty$ such that $n/a_L \to c$ in probability for some positive constant c. If n is Poisson or binomially distributed, then condition (1) holds. For instance if n is binomial, then $a_L = 2WD_0pL$ and $c = 1$.

As $\{W_{i0}\}_{i=1}^n$ are independent and identically distributed random variables and as n satisfies the above condition, we have $\Sigma_{i=1}^n W_{i0}/\sqrt{n}$ is asymptotically normally distributed by using the central limit theorem on a random number of summands as given in Serfling (1980, p. 32). As $\sqrt{n}/2L \to \sqrt{\mathrm{E}(n)}/2L$ in probability, the Slutsky theorem gives the asymptotic normality for $(2L)^{-1}n^{-1}\Sigma_{i=1}^n W_{i0}$. Thus, the asymptotic normality for $D_{1a} = (2L)^{-1}n^{-1}\Sigma_{i=1}^n W_i$ is obtained from (A.7), again using the Slutsky theorem. The last part of the proof is similar to that of Lemma 3 of Quang (1993) for a point transect estimator.