Appl. Statist. (1996)
45, *No.* 2, *pp.* 135–150

A Kernel Estimate for the Density of a Biological Population by Using Line Transect Sampling

By SONG XI CHEN†

La Trobe University, Bundoora, Australia

[Received February 1994. Final revision June 1995]

SUMMARY

Motivated by line transect aerial surveys of Southern Bluefin Tuna in the sea, a nonparametric kernel method is explored for estimating the density $D = N/A$ of a biological population where N is the unknown population size and A is the area occupied by the population. The kernel estimator is based on explicitly modelling the probability density function of the perpendicular sighting distances without any assumptions on the form of a detection function. The kernel estimates are shown to be asymptotically unbiased and robust estimates for D, satisfying the robustness criteria suggested by Burnham and co-workers. A new kernel-type confidence interval for D is also proposed. A simulation study shows that the kernel confidence intervals have better coverage than those of the Fourier series method. A tuna data set is analysed; the kernel method yields reasonable estimates of abundance and is robust against the changing detection function during a line transect survey.

Keywords: Aerial survey; Confidence intervals; Density estimate; Fourier series estimate; Kernel smoothing; Line transect sampling

1. Introduction

This paper was motivated by the need to develop a relative abundance index for a commercially valuable fishery species—Southern Bluefin Tuna. Aerial surveys using line transect sampling were conducted over the Great Australian Bight in summer to detect the tuna schools on the sea surface. Two experienced tuna spotters, one of whom was also the pilot, sat in the front of a light plane; this, together with the special structure of the plane, largely eliminated the 'blind strip' problem experienced in other aerial surveys (Pollock and Kendall, 1987). The survey plane had an auto-pilot to allow the pilot–spotter to concentrate on spotting during the survey. The plane was also equipped with a satellite-based global position system (GPS). The GPS records the position (latitude and longitude) of the aircraft when a school is first detected, the position where the plane breaks the transect to close the school and the exact position of the school. These gave us fairly accurate measurements of the sighting distance needed for using the line transect method. The use of the GPS also avoids the rounding error associated with measuring the sighting angles. Since only surface schools can be detected from the air, acoustic tracking and archival tagging were incorporated to investigate the surfacing behaviour of the tuna; these

†*Address for correspondence*: School of Statistics, La Trobe University, Bundoora, Victoria 3083, Australia.
E-mail: stasc@saturn.latrobe.edu.au

0035–9254/96/45135

techniques gave the proportion of the time that the tuna stay on the surface. This paper considers only using the line transect method to estimate the surface density of the tuna. Nevertheless, the results in this paper are applicable to similar line transect surveys.

Line transect sampling has been used for many decades to estimate wildlife abundance. Comprehensive reviews of the method can be found in Seber (1982) and Buckland *et al.* (1993). Let $D = N/A$ be the density of a population, where N is the unknown population size and A is the area occupied by the population. For the tuna aerial survey, N is the total number and D the density of tuna schools on the sea surface. Dividing D by the surfacing probability obtained from related acoustic and archival tagging will give the total abundance in the entire survey ocean. To estimate D, randomly allocated non-overlapping transect lines within the area are traversed by observers for a distance L. Each object sighted from the transect is counted and its perpendicular distance to the transect is measured by various means. Let w be the maximum posible perpendicular distance from a detected school to the transect lines. Assume that n objects are detected independently, with perpendicular distances $x_1, x_2, \ldots, x_n$. Let f be the probability density function from which the sample $x_1, x_2, \ldots, x_n$ is drawn and g be the conditional probability function of detecting an object given that the object is at a perpendicular distance x from the transect line; g is commonly called the detection function.

Suppose that $g(0) = 1$, which means certain detection of an object on the transect line. This was a reasonable assumption for the aerial survey as the two spotters at the front of the plane could cover the area immediately ahead of them. This is also reasonably supported by the real data, which show a monotonic declining histogram of the perpendicular sighting distance from 0 to w. Since the transect lines are allocated randomly, the expected number of objects seen is $E(n) = Np$, where p is the probability of sighting an object from the transect lines. From Seber (1982), p. 29 and following feature,

$$D = \frac{E(n)\, f(0)}{2L}.$$

Let $\hat{f}(0)$ be an estimate for $f(0)$; then a general estimate for D is

$$\hat{D} = \frac{n\, \hat{f}(0)}{2L}. \tag{1.1}$$

Various parametric estimates $\hat{f}(0)$ have been proposed by assuming parametric models for the detection function $g(x)$ (see Seber (1982) and Buckland *et al.* (1993)). A nonparametric Fourier series estimate for $f(x)$ was given by Burnham *et al.* (1980), and a Hermite polynomial model was proposed by Buckland (1985).

In the aerial survey, the plane flew about 800 km each survey day between inshore and the edge of the continental shelf. The weather conditions differed in terms of wind speed, cloud cover, glare etc. Obviously, the sighting function $g(x)$ can change from one time to another. Also, parametric forms for $g(x)$ can be difficult to assign. So, nonparametric estimates, which are robust against changing $g(x)$, were more appropriate to the survey. The nonparametric (or 'semiparametric') methods of both Burnham *et al.* (1980) and Buckland (1992) provide robust estimates for $f(0)$. However, the estimates of $f(x)$ given by the two methods are not necessarily probability density function themselves, and severe undercoverage of confidence intervals for D using the Fourier series method was reported by Burnham *et al.* (1980) and Quang (1990). This motivated the author to look at another non-parametric method—the kernel method.

The kernel method was introduced by Fix and Hodges (1951) as a way of freeing discriminant analysis from rigid distributional assumptions. Since then it has been developed and used in many areas of statistical applications. A full description of the kernel method can be found in Silverman (1986). The kernel method was first used in line transect sampling by Buckland (1992) in a comparison with the Hermite polynomial model. He reported that

> 'the Hermite polynomial and the kernel estimates are very similar, raising the question of whether the computationally more expensive polynomial method is worthwhile'

(Buckland (1992), p. 69). However, as he noted, the kernel method could be sensitive to the choice of window width and

> 'a narrow window is adversely affected by rounding in the data . . . leading to higher values for $\hat{f}(0)$'.

Rounding was not a problem in the aerial survey, as the use of the GPS eliminated the need for measuring sighting angles. Quang (1993) developed kernel estimators for variable circular plot surveys. He used the kernel method to estimate the derivative of a probability density function. Although there are technical similarities between Quang's approach and ours, the differences are also large because of the differences between variable circular plot surveys and line transect surveys in terms of survey design, detection pattern and the form of estimate. Also, using the kernel method to estimate the density derivative is different from estimating the density itself: the former requires a larger smoothing window width as it is much noisier than smoothing the density itself: see Härdle *et al.* (1990). Therefore, a study of the kernel method for line transect surveys is very much needed.

This paper shows that the kernel method provides a sensible estimate for D, and that the kernel estimate is asymptotically unbiased, and both 'model' and 'pooling' robust under the criteria defined by Burnham *et al.* (1980). A kernel-type confidence interval for D is proposed to improve the coverage of the confidence interval constructed by the Fourier series method. A simulation study shows that the kernel confidence interval has better coverage.

Section 2 gives the kernel estimates for D. Section 3 describes two methods of choosing the smoothing window width. Section 4 calculates the mean and variance of the kernel estimate. Section 5 introduces a kernel-type confidence interval for D. Section 6 presents an analysis for a data set from the tuna aerial survey. Section 7 shows some results from a simulation study. Section 8 gives a general discussion of the kernel and some related methods.

2. Kernel Estimates of $f(0)$ and D

The kernel estimate of a probability density function can be viewed as a smoothed version of a histogram. The difference is that, instead of counting the number of data values falling into bins, we weight each data point by a smooth function K centred at the data point. Precisely, we assume a sample $x_1, x_2, \ldots, x_n$ drawn from a distribution with probability density function $f(x)$. Then, a kernel estimate of $f(x)$ is defined as

$$\hat{f}_k(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{2.1}$$

where K is the kernel that determines the shape of the bumps centred at each data point and h is the window width that controls the smoothness of the bumps. It can be shown that $\hat{f}_k(x)$ is a probability density function if the kernel K is itself a probability density function. Furthermore, $\hat{f}_k(x)$ preserves all the continuity and differentiability properties of the kernel K. The most widely used kernel is the Gaussian kernel

$$K(x) = \frac{1}{\surd(2\pi)} \exp\left(-\frac{x^2}{2}\right),$$

which is both symmetric and differentiable.

Some modifications have to be made before we can apply equation (2.1) to line transect data. Usually all the distances x_i are non-negative, with no record of the side of the transect on which a detection is made. This implies that

$$f(x) = 0 \qquad \text{if } x \leqslant 0 \text{ or } x \geqslant w. \tag{2.2}$$

However, the kernel density estimate $\hat{f}_k(x)$ given by equation (2.1) does not necessarily satisfy condition (2.2). As suggested by Buckland (1992), to make it satisfy condition (2.2), we replace each data value x_i with x_i and its reflection $-x_i$. Applying equation (2.1) on the extended sample $x_1, -x_1, \ldots, x_n, -x_n$, a kernel estimate for $f(x)$ is obtained as

$$\hat{f}_k(x) = \begin{cases} \dfrac{1}{nh}\sum_{i=1}^{n} \left\{K\left(\dfrac{x - x_i}{h}\right) + K\left(\dfrac{x + x_i}{h}\right)\right\} & \text{if } 0 \leqslant x \leqslant w, \\ 0 \quad \text{otherwise.} \end{cases} \tag{2.3}$$

Even the Gaussian kernel K has support on $(-\infty, \infty)$; however, when $|x \pm x_i| > 4h$ $K\{(x \pm x_i)/h\}$ are negligibly small. Therefore, by choosing w sufficiently large that $w \geqslant \max(x_i) + 4h$ the above $\hat{f}_k(x)$ decays to 0 for $x \geqslant w$. Therefore, condition (2.2) is satisfied by the kernel estimate (2.3). Correspondingly, a kernel estimate for population density D can be defined as

$$\hat{D}_k = \frac{n\,\hat{f}_k(0)}{2L}. \tag{2.4}$$

By condition (2.3) and assuming that K is a symmetric kernel,

$$\hat{f}_k(0) = \frac{2}{nh} \sum_{i=1}^{n} K\left(\frac{x_i}{h}\right). \tag{2.5}$$

The kernel estimate (2.4) satisfies the 'model' and 'pooling robust' criteria given by Burnham *et al.* (1980), p. 44 and following feature. It is 'model robust' because a general and flexible model for $f(x)$ is assumed without specifying a particular form for it. Some simple algebra shows that it is also pooling robust. If the kernel K is symmetric and differentiable, e.g. the Gaussian kernel, then $\hat{f}_k(x)$ in condition (2.3) satisfies $\hat{f}_k^{(1)}(0) = 0$, which resembles the 'shoulder condition' suggested by Burnham *et al.* (1980). (We denote $q^{(l)}(x)$ as the lth derivative of a function $q(x)$.) The shoulder condition implies 100% detection in some region near the track lines, which is reasonable for most line transect surveys, including the aerial survey. From equation (2.5) we see that the sighting width w does not appear in the kernel estimate $\hat{f}_k(0)$. This implies that the kernel method automatically chooses a suitable truncation width w from the data and the value of $\hat{f}_k(0)$ does not explicitly depend on w. By contrast, w is directly used in the Fourier series method.

To investigate the performance of the proposed kernel estimate, we assume the following conditions:

(a) $f(x)$ has the second-order derivative and $f^{(1)}(0) = 0$;
(b) $g(0) = 1$;
(c) $E(n) = Np$ and $\text{var}(n) = \gamma\, E(n)$, where p is the probability of sighting an object from the transect lines and γ is some constant.

Condition (a) implies that the shoulder condition $g^{(1)}(0) = 0$ as g is proportional to f. The requirement of the second derivative is used to develop a bias correction for $\hat{D}_k$ in Section 4. Condition (b) assumes that 100% detection can be achieved on the transect lines. In condition (c), if $\gamma \leqslant 1$ then the sample size n is distributed very much like the binomial distribution where $\text{var}(n) = (1-p)\, E(n)$, or the Poisson distribution where $\text{var}(n) = E(n)$. In practice γ is commonly larger than 1.

From Seber (1982), the sighting probability p has the form

$$p = w^{-1} \int_0^w g(x)\, \mathrm{d}x = \{w\, f(0)\}^{-1}.$$

By the definition of D, $N = AD = 2LwD$. So we have

$$N \to \infty \qquad \text{if and only if } L \to \infty.$$

Also, from condition (b), $E(n) = Np = 2wLD\{w\, f(0)\}^{-1} = 2LD\, f^{-1}(0)$. We shall use these formulae repeatedly throughout this paper.

3. Choosing *K* and *h*

To construct the kernel estimate, we must choose kernel K and window width h. A widely used criterion is to choose K and h to minimize the *mean integrated squared error* (MISE) of $\hat{f}_k$, defined as

$$\text{MISE}(\hat{f}_k) = \int_0^\infty E\{\hat{f}_k(x) - f(x)\}^2\, \mathrm{d}x. \tag{3.1}$$

There is not much to choose between the various kernels as they all contribute very similar amounts to the MISE (see Silverman (1986)). We use the Gaussian kernel in this paper as it is a density function, has enough derivatives and allows the estimate to meet the shoulder condition. Hence, we concentrate on choosing h. The window width h controls the smoothness of the fitted density curve. A larger h gives a smoother estimate with smaller variance and larger bias. A smaller h produces a rougher estimate with larger variance but, in the absence of rounding errors, smaller bias. The optimal window width defined by minimizing MISE (3.1) is a compromise between the variance and bias of the estimate.

According to Silverman (1986), p. 40, the optimal h has the form

$$h_n = k_2^{-2/5}\left\{\int_{-\infty}^{\infty} K(t)^2\,dt\right\}^{1/5}\left\{\frac{1}{2}\int_0^{\infty} f^{(2)}(t)^2\,dt\right\}^{-1/5} n^{-1/5}, \tag{3.2}$$

where $k_2 = \int_{-\infty}^{\infty} t^2 K^2(t)\,dt$. As the Gaussian kernel is used, then $k_2 = 1$ and $h_n = C_0 n^{-1/5}$ where

$$C_0 = (4\pi)^{-1/10}\left\{\frac{1}{2}\int_0^{\infty} f^{(2)}(t)^2\,dt\right\}^{-1/5}. \tag{3.3}$$

The subscript n in h_n emphasizes that the above optimal h is obtained by conditioning on the sample size n. Thus, h_n is a random variable itself. It must be estimated from the data because of its dependence on an unknown derivative of $f(x)$.

Silverman (1986) suggested a reference to a standard distribution method to estimate the optimal h_n. The idea is to assign a standard family of distributions for f to obtain a value for the term $\int_0^{\infty} f^{(2)}(t)^2\,dt$ in equation (3.2) or (3.3). For instance, we may assume a normal detection function $g(x) = \exp(-x^2/2\sigma^2)$. This implies

$$f(x) = 2\{\sigma\surd(2\pi)\}^{-1}\exp(-x^2/2\sigma^2),$$

which gives us

$$\frac{1}{2}\int_0^{\infty} f^{(2)}(t)^2\,dt = \tfrac{3}{8}\pi^{-1/2}\sigma^{-5}.$$

Substituting it into equation (3.3) we have

$$h_n = 1.06\sigma n^{-1/5}. \tag{3.4}$$

To use equation (3.4) in practice, we replace σ with the sample standard deviation. This reference to a standard distribution method is easy to calculate and performs quite well when the true underlying distribution is close to the distribution assumed. Buckland (1992) used equation (3.4) for the deer data and reported very similar results to those obtained by the Hermite polynomial method. However, when the true $f(x)$ is not close to the assumed reference distribution, the result can be misleading.

Another approach for estimating the optimal h_n is the least-squares cross-validation (LSCV) method. From equation (3.1) we see that

$$\text{MISE}(\hat{f}_k; h) = E\left\{\int \hat{f}_k(x)^2\,dx - 2\int \hat{f}_k(x)\,f(x)\,dx\right\} + \int f^2(x)\,dx.$$

Since the last term in this equation does not depend on h, h_n actually minimizes the first term on the right-hand side, say $R(\hat{f}_k; h)$. The idea of LSCV is to calculate an unbiased estimate

$$M_0(h) = \int \hat{f}_k(x)^2\, dx - 2 \sum_{i=1}^{n} \hat{f}_{-ik}(x_i)$$

of $R(\hat{f}_k; h)$ for each h, where $\hat{f}_{-ik}$ is the kernel estimate for f constructed from all the data values except x_i and $-x_i$, i.e.

$$\hat{f}_{-ik}(x) = (n-1)^{-1} h^{-1} \sum_{j \neq i} \left\{ K\left(\frac{x - x_j}{h}\right) + K\left(\frac{x + x_j}{h}\right) \right\}.$$

Then we find the h that minimizes $M_0(h)$ over a grid of h-values.

A strong justification for using LSCV comes from Stone (1984), who showed that the window width obtained from LSCV converges to its optimal counterpart with probability 1 as the sample size $n \to \infty$ under some mild conditions on the kernel K and f. LSCV is a purely data-driven method that makes no assumption on the form of $f(x)$. Compared with the reference to a standard distribution method, LSCV generally provides more robust estimates of the optimal window width. However, it is computationally more involved. Computer software developed by the author is available to assist applied practitioners using LSCV in line transect sampling.

Silverman (1986) warned of a potential problem with LSCV when the data are discretized. However, in the aerial survey the perpendicular distance can be measured continuously, as the GPS records the positions of detection and schools fairly accurately. LSCV can, therefore, produce a meaningful window width for the survey and for similar surveys.

4. Mean and Variance of $\hat{D}_k$

After choosing the window width h, we derive the mean and variance for $\hat{D}_k$ in this section. For simple notation and without affecting the results, our calculations are carried out only for $h = h_n = C_0 n^{-1/5}$, with C_0 given theoretically by equation (3.3). In practice, after determining h by either the reference or LSCV method, C_0 can be approximated by $hn^{1/5}$. We use notations $O(v)$ and $o(v)$ to represent terms at the same and smaller order of magnitudes of v respectively, as explained by Serfling (1980), p. 1.

We start by computing the mean of $\hat{D}_k$. As the Gaussian kernel has been used, from Silverman (1986), p. 39,

$$E\{\hat{f}_k(0)|n\} = f(0) + \tfrac{1}{2} f^{(2)}(0) h_n^2 + O(h_n^3). \tag{4.1}$$

Thus,

$$E(\hat{D}_k) - D = \frac{E[n\, E\{\hat{f}_k(0)|n\}] - E(n) f(0)}{2L} = b_k + O\left\{\frac{E(nh_n^3)}{L}\right\},$$

where

$$b_k = \tfrac{1}{4} f^{(2)}(0)\, E(nh_n^2)/L = \tfrac{1}{4} f^{(2)}(0)\, C_0^2\, E(n^{3/5})/L.$$

It may be shown from Serfling (1980), p. 14 and following feature, that $E(n/N)^{3/5} \to p^{3/5}$ on the basis of the fact that $n/N \to p$ in probability. Also, since $N = 2wDL$ and

$p = \{w f(0)\}^{-1}$, then

$$b_k = \tfrac{1}{4} f^{(2)}(0)\, C_0^2\{2D/f(0)\}^{3/5} L^{-2/5} + o(L^{-2/5}).$$

We may similarly show that $E(nh_n^3)/L = O(L^{-3/5})$. Therefore,

$$E(\hat{D}_k) = D + b_k + o(L^{-2/5}), \tag{4.2}$$

which indicates that, like all other nonparametric estimates, $\hat{D}_k$ is also subject to some bias, the bias here being of the order $L^{-2/5}$. However, it can be reduced by increasing the survey effort L. Thus, $\hat{D}_k$ is an asymptotically unbiased estimate of D. We can also reduce the bias by explicitly correcting $\hat{D}_k$. From equation (4.2), by estimating the dominant bias term b_k, a bias-corrected estimate

$$\hat{D}_{kc} = \hat{D}_k - \tfrac{1}{2}\hat{f}^{(2)}(0)\, h_n^2 n L^{-1}$$

can be proposed where $\hat{f}_2^{(2)}(0)$ is an estimate of $f^{(2)}(0)$. To estimate $\hat{f}_2^{(2)}(0)$, we could simply take the second derivative of $\hat{f}_k(x)$ given by equation (2.3). However, as the kernel estimates for the density derivative are usually much noisier than that for $f(0)$, a larger window width should be used. Härdle *et al.* (1989) suggested the use of LSCV to choose the window width for estimating density derivatives. They showed that the bandwidth for estimating the kth derivative under the MISE criterion is of order $n^{-1/(2k+5)}$, which is a larger order than that of estimating $f(x)$ to reduce the noise in the derivative estimation. A simple rule of thumb for choosing the window width to estimate $f^{(2)}(0)$ is

$$h_1 = h_n n^{1/5-1/9},$$

where h_n is the bandwidth for estimating $f(x)$. Accordingly,

$$\hat{f}_k^{(2)}(0) = 2n^{-1} h_1^{-3} \sum K^{(2)}(x_i/h_1),$$

where $K^{(2)}$ is the second derivative of K. It can be readily shown that $E(\hat{D}_{kc}) = D + o(L^{-2/5})$, which means that the bias of $\hat{D}_{kc}$ is smaller than that of $\hat{D}_k$.

Next we look at the variance of $\hat{D}_k$. By the argument of conditional expectation,

$$\mathrm{var}(\hat{D}_k) = (2L)^{-2}\, E[n^2\, \mathrm{var}\{\hat{f}_k(0)|n\}] + (2L)^{-2}\, \mathrm{var}[n\, E\{\hat{f}_k(0)|n\}].$$

Again from Silverman (1986), p. 39 and following feature,

$$\mathrm{var}\{\hat{f}_k(0)|n\} = 4(nh_n)^{-1} f(0) \int_0^\infty K(t)^2\, \mathrm{d}t + O(n^{-1}). \tag{4.3}$$

Thus, we have from equation (4.1) and the condition $\mathrm{var}(n) = \gamma\, E(n) = O(L)$ in condition (c) at the end of Section 2 that

$$\operatorname{var}(\hat{D}_k) = L^{-2}\left\{f(0)\int_0^\infty K(t)^2\,dt\,E(h_n^{-1}n) + \tfrac{1}{4}f(0)^2\operatorname{var}(n)\right\} + O(L^{-1})$$
$$= L^{-2}f(0)C_0^{-1}p^{6/5}N^{6/5}\int_0^\infty K(t)^2\,dt + O(L^{-1})$$
$$= \tfrac{1}{2}(4\pi)^{-1/2}(2D)^{6/5}f(0)^{-1/5}C_0^{-1}L^{-4/5} + O(L^{-1}). \tag{4.4}$$

This implies that $\operatorname{var}(\hat{D}_k)$ is of order $L^{-4/5}$ and can be reduced by increasing the survey effort L.

To estimate the variance of $\hat{D}_k$ practically, the following analogue of equation (3.3) of Buckland *et al.* (1993), p. 53, should be used:

$$\widehat{\operatorname{var}}(\hat{D}_k) = \hat{D}_k^2[\widehat{\operatorname{cv}}^2(n) + \widehat{\operatorname{cv}}^2\{\hat{f}_k(0)\}], \tag{4.5}$$

where $\widehat{\operatorname{cv}}(n)$ and $\widehat{\operatorname{cv}}\{\hat{f}_k(0)\}$ are the estimates for the coefficient of variation of n and $\hat{f}(0)$ respectively. If the target population is not aggregated and n is basically Poisson distributed, then $\widehat{\operatorname{cv}}^2(n) = 1/n$. When the target population is aggregated, n is no longer Poisson distributed. Usually data show that $\operatorname{var}(n) = \gamma E(n)$ for some $\gamma > 1$. In this case, the variation of n is better assessed by replication of the random transects and recording the number of sightings made in each transect and the transect length. Interested readers should consult Buckland *et al.* (1993) for more detail.

After obtaining the mean and variance in equations (4.2) and (4.4), the coefficient of variation of $\hat{D}_k$ is

$$\operatorname{cv}^2(\hat{D}_k) = (4\pi)^{-1/2}C_0^{-1}\{2D_k/f_k(0)\}^{1/5}D_k^{-1}L^{-4/5} + O(L^{-1}).$$

The $\operatorname{cv}^2(\hat{D}_k)$ can be approximated by

$$\operatorname{cv}^2(\hat{D}_k) \approx (4\pi)^{-1/2}\hat{C}_0^{-1}\{2\hat{D}_k/\hat{f}_k(0)\}^{1/5}\hat{D}_k^{-1}L^{-4/5}, \tag{4.6}$$

where $\hat{C}_0 = h_n n^{1/5}$, since $h_n = C_0 n^{-1/5}$. Formula (4.6) can be used to determine the length of the transect line L to achieve a certain level of coefficient of variation.

5. Confidence Intervals

In this section we construct confidence intervals for D based on a result on the limiting distribution of $\hat{D}_k$. We show in theorem 1 that $L^{2/5}(\hat{D}_k - D)$ is asymptotically normally distributed, which allows us to construct confidence intervals for D. The proof of theorem 1 is deferred until Appendix A.

Theorem 1. Assume conditions (a) and (b) at the end of Section 2 and that n is either Poisson or binomially distributed, and $h = \beta L^{-1/5}$ for some constant β. Then,

$$L^{2/5}(\hat{D}_k - D) \to N(\mu, \upsilon^2) \qquad \text{in distribution as } L \to \infty$$

where $\mu = f^{(2)}(0)\,f(0)^{-1}D\beta^2$ and $\upsilon^2 = (\beta\sqrt{\pi})^{-1}D$.

Theorem 1 implies that $\hat{D}_k$ is asymptotically normally distributed with mean $D + \mu L^{-2/5}$ and variance $\upsilon^2 L^{-4/5}$. The limiting normal distribution takes account of the bias of $\hat{D}_k$, which is reflected in the secondary term $\mu L^{-2/5}$ for the mean. The theorem is proved for a non-random h-value $h = \beta L^{-1/5}$. When h is a random value, e.g. $h = h_n = C_0 n^{-1/5}$, as recommended in Section 3, similar results can be attained by using a Lindeberg–Feller type of central limit theorem for a triangular array of random variables. However, to save space, the proof is not supplied here, but a very similar proof was given by Quang (1993) for circular plot surveys. Notice that $n = 2DL/f(0) + o_p(L)$. Therefore, when $h = h_n = C_0 n^{-1/5}$, we have

$$h = C_0\{2D/f(0)\}^{-1/5} L^{-1/5} + o_p(L^{-1/5}),$$

implying $\beta = C_0\{2D/f(0)\}^{-1/5}$, $\mu = b_k L^{2/5}$ and $v^2 = \mathrm{var}(\hat{D}_k) L^{4/5}$ where b_k is the dominant bias term of $\hat{D}_k$. In practice we need not evaluate C_0, as only β is directly involved in the expressions for μ and v^2, and β can be approximated as $h_n L^{1/5}$.

From theorem 1 a confidence interval I_α with nominal coverage α for D may be constructed:

$$I_\alpha = (\hat{D}_k - L^{-2/5}(\hat{\mu} + \hat{v} z_{(1+\alpha)/2}),\ \hat{D}_k - L^{-2/5}(\hat{\mu} - \hat{v} z_{(1+\alpha)/2})), \tag{5.1}$$

where $z_{(1+\alpha)/2}$ is the $(1+\alpha)/2$ upper percentile of the standard normal distribution and $\hat{\mu}$ is the estimate of μ and $\hat{v}_2^2$ of v^2 by replacing $D, f(0)$ and $f^{(2)}(0)$ with $\hat{D}_k, \hat{f}_k(0)$ and $\hat{f}_k^{(2)}(0)$ respectively. Theorem 1 assures that

$$\Pr(D \in I_\alpha) \rightarrow \alpha, \qquad \text{as } L \rightarrow \infty.$$

6. Analysing Tuna Data

The annual aerial surveys for Southern Bluefin Tuna since 1991 have been conducted from January to March, when the tuna tend to stay on the sea surface and the weather conditions are generally good. The survey area is from longitude 128° E to 135° E and from inshore to the 800 m contour of the sea-floor. It is divided into five parallel north–south blocks such that one block can be surveyed in a survey day (from 10 a.m. to about 6 p.m.). The data obtained from the two preliminary surveys in 1991 and 1992 revealed that most of the tuna sightings were made inshore and on the edge of the continental shelf. Subsequent surveys therefore used a transect design of random long (200–400 km) north–south lines connected by zigzags inshore and at the shelf edge. These zigzags increased the survey effort in the two strata. Fig. 1 shows the transects of one replicate of the five blocks in the 1993 survey. The rectangular transects in the most eastern block were across the hot spots in the region.

Two types of abundance measure are required for the survey. One is the value of D (the number of the tuna schools per unit area); the other is the so-called biomass density D_1 (tons of the tuna per unit area on the surface). We consider estimating only D in this paper. Estimating D_1 is related to the problem of correcting school size bias as discussed by Drummer and McDonald (1987). Use of the kernel method to correct the school size bias is discussed in Chen (1994). The aim of the survey is to accumulate time series of $(\hat{D}, \hat{D}_1)$ such that a real abundance trend can be identified and a relative index of the tuna abundance can be developed.

We applied the kernel method to a data set from the 1993 survey. The data set consists of pooled perpendicular distances from all transects for schools whose size is less than 50 tons to eliminate the size effect on detection. The sample size $n = 162$ and survey effort $L = 10361$ miles.

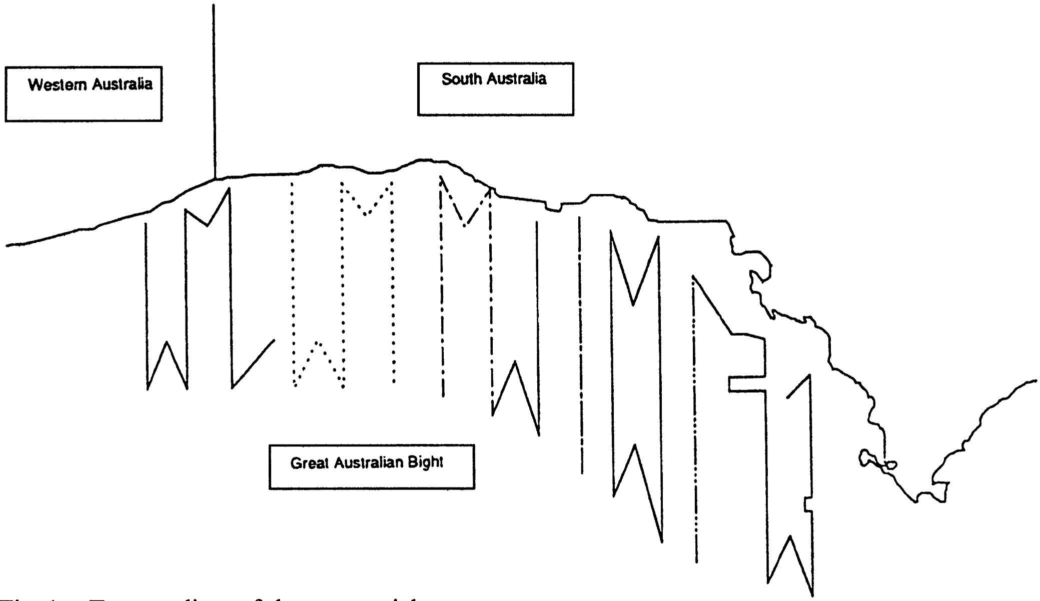

Fig. 1. Transect lines of the tuna aerial survey

We used both the reference and the LSCV methods to determine the window width h and to compute the kernel estimate $\hat{D}_k$ and the bias-corrected kernel estimate $\hat{D}_{kc}$, and confidence interval for D. The reference method gave $h = 1.37$ and the LSCV method $h = 1.66$. The larger window width from LSCV was because the data were more spread out and had a wider shoulder than in a normal distribution. For comparison, Fourier series estimates with truncation width $w = 16$ and $w = 9.6$ were also calculated. No data are truncated for $w = 16$. By choosing $w = 9.6$ the largest 5% of the data were truncated. The estimated curves of $f(x)$ by using the kernel method and the Fourier series method for the complete and truncated data set are given in Fig. 2. The kernel and Fourier series estimates for $f(0)$ and $\hat{D}$ together with the standard error of $\hat{D}$ are summarized in Table 1. The standard errors for the kernel estimates were calculated according to equation (4.5) with $\widehat{\text{cv}}^2(n) = 1/n$ under a Poisson assumption. Those for the two Fourier series estimates were obtained similarly.

Fig. 2 shows that the kernel curves fitted both the untruncated and the truncated

TABLE 1
Kernel and Fourier series estimates for $f(0)$ and D for the tuna data

	h	$\hat{f}_k(0)$	$\hat{f}_{kc}(0)$	$\hat{D}_k$	$\hat{D}_{kc}$	*Standard error of* $\hat{D}_k$
Reference	1.37	0.182	0.186	0.000143	0.000147	0.000047
LSCV	1.65	0.176	0.181	0.000139	0.000143	0.000042
Fourier series with $w = 16$	3†	0.186	NA‡	0.000147	NA	0.000030
Fourier series with $w = 9.6$	1†	0.179	NA‡	0.000141	NA	0.000021

†The number of terms used in the Fourier series.
‡NA — the bias-corrected estimates were not available for the Fourier method.

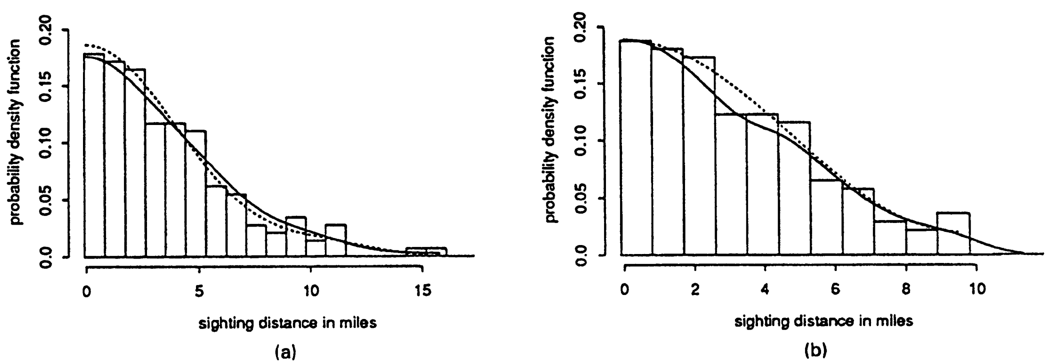

Fig. 2. Histograms of the tuna data set and the kernel (———) and Fourier series (- - - -) estimates for $f(x)$ with truncation widths (a) $w = 16$ and (b) $w = 9.6$

data quite well. The Fourier series curve for untruncated data ($w = 16$) tended to overestimate $f(x)$ near the origin. After data truncation with $w = 9.6$ the Fourier series curve seemed to give a good fit near 0. In contrast, different truncation widths had little effect on the kernel estimates for $f(0)$, as the kernel method need not be assigned an explicit truncation width in its calculation. In Fig. 2(b), the kernel curve split beyond the truncation width $w = 9.6$ to make it integrate to 1. The purpose of choosing $w = 9.6$ was only to improve the Fourier series estimate for $f(0)$. Plotting the kernel curve in Fig. 2(b) was for demonstration only as the kernel curve for untruncated data is the curve that the inference should use. One referee has suggested, alternatively, a kernel curve for $w = 9.6$ by truncating the kernel curve for untruncated data and then rescaling it so that it integrates to 1.

Table 1 shows that the reference method gives a larger point estimate and standard error than its LSCV counterparts, owing to a smaller h-value used. For the Fourier series method, as the truncation width w decreased the number of terms used in the Fourier series also decreased. This, in turn, reduced the standard error of the Fourier abundance estimates as fewer terms were used in the series. This relationship was consistent with the findings in Buckland (1982, 1985). Consistent with Fig. 2, the Fourier method produced larger estimates for $f(0)$ and D than did their kernel counterparts, but with smaller variance. The latter resulted in narrower confidence intervals for D by the Fourier method. This may be why the Fourier confidence intervals have an undercoverage problem as reported by Burnham *et al.* (1980) and Quang (1990). The 95% confidence intervals for D were (0.0001, 0.000 194) for the reference method, (0.000 101, 0.000 185) for LSCV and (0.000 117, 0.000 177) and (0.000 120, 0.000 163) for the Fourier method with $w = 16$ and $w = 9.6$ respectively. An advantage of the kernel method is that it automatically chooses a truncation width from the data, and its estimate for $f(0)$ does not depend on the truncation width as much as the Fourier series method does.

7. Simulation Results

We now present some simulation results designed to provide empirical outcomes to the theory developed in the early sections of this paper. The aims of the simulation study were to investigate the bias and variance of the kernel estimates for D and coverages of the confidence intervals given in equation (5.1). The results were based on 5000 simulations using the random number generator supplied by Press *et al.* (1989). In each simulation, we generated N uniform random points, which simulated the positions of a biological population, within a rectangular area with length L and width $2w$. Hence the density $D = N/2Lw$ was known. The exponential power series detection function

$$g(x) = \exp\{-(bx)^a\}$$

was used to detect the simulated populations, where a and b are the shape and scale parameters respectively. We fixed $D = 0.15$, $b = 0.5$ and $w = 10$, and chose the shape parameter $a = 1.5$, 2.0 and 2.5 and the population size $N = 200$ and $N = 300$. The window width h_n was determined by the LSCV method. The nominal coverage level considered was $\alpha = 0.95$. The Fourier series estimates and confidence intervals were also calculated for comparison.

Each entry in Tables 2 and 3 was based on the average of the 5000 simulations. Table 2 shows point estimates for D. The estimators considered are the kernel estimate $\hat{D}_k$, the bias-corrected kernel estimate $\hat{D}_{kc}$ and the Fourier series estimate $\hat{D}_{FS}$ of Burnham *et al.* (1980). The average window width and the number of terms used in the Fourier series estimates are also supplied, together with average sample sizes. We observe that first the bias-corrected estimate $\hat{D}_{kc}$ substantially reduced the bias associated with their uncorrected counterparts. Secondly, $\hat{D}_{kc}$ performed well for the case where $f(x)$ had a very smooth shoulder near the origin when $a = 2$ and $a = 2.5$, and were biased downwards when $f(x)$ was less smoothed at the origin when $a = 1.5$. The latter was not surprising as the requirement of $f^{(1)}_{(0)} = 0$ was violated. Third the performance of the Fourier series estimates was just the opposite of those of the kernel method, which was less biased when $f(x)$ was less smoothed and overestimated when $f(x)$ was smooth at 0. Finally, from $N = 200$ to $N = 300$ with increasing sample size the standard error of each estimate was reduced; the window width became smaller and the truncation terms of the Fourier series were increased.

TABLE 2
Point estimates and their standard errors (in parentheses) for $D = 0.15$†

	Estimates when $a = 1.5$		*Estimates when $a = 2$*		*Estimates when $a = 2.5$*	
	$N = 200$	$N = 300$	$N = 200$	$N = 300$	$N = 200$	$N = 300$
$\hat{D}_k$	0.123	0.126	0.131	0.133	0.136	0.138
	(0.033)	(0.028)	(0.033)	(0.027)	(0.032)	(0.027)
$\hat{D}_{kc}$	0.135	0.136	0.143	0.144	0.147	0.148
	(0.037)	(0.031)	(0.036)	(0.030)	(0.036)	(0.030)
$\hat{D}_{FS}$	0.145	0.146	0.154	0.153	0.159	0.158
	(0.041)	(0.034)	(0.039)	(0.033)	(0.037)	(0.031)
Average h	0.933	0.832	0.852	0.773	0.804	0.733
Average Fourier series terms	3.89	4.24	3.83	5.06	3.69	3.85
Average n	36.2	54.1	35.5	53.1	35.5	53.1

†The exponential series detection function $g(x) = \exp\{-(bx)^a\}$ was used with $b = 0.5$ and $w = 10$. The subscripts k, kc and FS imply the kernel, bias-corrected kernel and the Fourier series estimates respectively.

TABLE 3
Coverages and lengths of 95% kernel and Fourier series confidence intervals for $D = 0.15$

	Estimates when $a = 1.5$		*Estimates when $a = 2$*		*Estimates when $a = 2.5$*	
	$N = 200$	*$N = 300$*	*$N = 200$*	*$N = 300$*	*$N = 200$*	*$N = 300$*
Kernel length	0.907	0.925	0.942	0.957	0.957	0.965
	0.143	0.123	0.147	0.126	0.153	0.135
Fourier series length	0.815	0.818	0.887	0.894	0.919	0.921
	0.115	0.098	0.117	0.098	0.117	0.097

Table 3 gives the coverages of the kernel and Fourier series confidence intervals. It shows that the kernel confidence intervals had better coverage than their Fourier series counterparts for almost every case considered, where the latter suffered severe undercoverage, as previously reported by others. The lengths of the Fourier series confidence intervals were too narrow. This is due to an underestimate of the standard error by the Fourier method, which is also revealed by comparing a quarter-length of the 95% Fourier-type confidence intervals with the standard error of the point estimates shown in Table 2. It should be noted that the bootstrap has been proposed by Buckland *et al.* (1993) in conjunction with the Fourier series method to improve its standard error estimate and coverage. The advantages of the kernel method is that it can achieve reasonably good coverage without resorting to the bootstrap.

8. Discussion

We have shown in the previous sections that the kernel method is a useful technique in analysing data from line transect sampling. Unlike the parametric line transect method, it need not assume a parametric form for the detection function. Therefore, it is very much a data-oriented modelling approach and is robust against a changing detection function during a line transect survey. The price paid for this simplicity and robustness is that we must choose the smoothing window width h. Two ways of choosing h have been described. In fact, the window width h plays the same role as the number of terms used in the Fourier series: both control the smoothness of the curves fitted. Compared with the Fourier method, the estimate of $f(x)$ by the kernel method is a real probability density function. It does not require an explicit truncation width in calculation. Furthermore, the simulation study reported in the previous section showed that the kernel-based confidence interval (5.1) had better coverage than its Fourier series counterpart.

As Buckland (1992) pointed out, the kernel method cannot be applied to data subjected to severe rounding error. This usually occurs when the perpendicular sighting distance is derived from the sighting angle, which is frequently rounded to the nearest 5° or 10°. Consequently, more zero perpendicular distances are recorded. This results in a smaller window width h and an inflated estimate for $f(0)$. The kernel method discussed in this paper cannot be used directly on grouped data. However, a 'binned' kernel method (Silverman, 1982) has been developed to reduce the computational burden by binning the data into bins or groups. It will be interesting to see how the idea leads to kernel density estimations for grouped data and how they perform in the line transect sampling. If the binned kernel method works well for the grouped data, the problem of rounding error at the origin can be tackled by grouping the data. More research is certainly needed in these areas.

Acknowledgements

The author is grateful to Dr B. M. Brown and Dr T. Polacheck for interesting discussions, and to Dr S. T. Buckland, a referee, the Associate Editor and the Editor for comments which led to substantial improvements in the paper.

Appendix A: Proof of Theorem 1

We give the proof of theorem 1 only for where n is Poisson distributed with mean Np, since the case of n being binomial can be handled similarly.

Let $\phi_h(t)$ denote the common characteristic function of $y_i = h^{-1} K(x_i/h)$ for $i = 1, \ldots, n$. By the Taylor series expansion of $\phi_h(t)$,

$$\phi_h(t/L^{1-\alpha}) = 1 + E(y_1)L^{-(1-\alpha)}(it) - \tfrac{1}{2} E(y_1^2)L^{-2(1-\alpha)}t^2 + O(L^{-3(1-\alpha)}t^3) \tag{A.1}$$

for any $0 \leqslant \alpha \leqslant 1$. From Silverman (1986), p. 39, it may be shown that

$$E(y_1) = f(0) + \tfrac{1}{2} f^{(2)}(0)h^2 + O(h^3),$$

$$E(y_1^2) = f(0)h^{-1} \int_{-\infty}^{\infty} K(t)^2 \, dt + O(h) = (4\pi)^{-1/2} f(0)h^{-1} + O(h),$$

by using the Gaussian kernel. Substitute the above equations into equation (A.1); as $h = \beta L^{-1/5}$, we end up with

$$\phi_h(t/L^{1-\alpha}) = 1 + \{f(0) + \tfrac{1}{2} f^{(2)}(0)h^2\}L^{-(1-\alpha)}(it) - \tfrac{1}{2}(4\pi)^{-1/2} f(0)h^{-1}L^{-2(1-\alpha)}t^2 + O(\Delta), \tag{A.2}$$

where $\Delta = L^{-3/5-(1-\alpha)}t - L^{-1/5-2(1-\alpha)}t^2 + L^{-3(1-\alpha)}t^3$. Since n is Poisson distributed, from Feller (1957), p. 269, the characteristic function of $\Sigma_{i=1}^n y_i$ has the form $\exp\{-Np + Np\, \phi_h(t)\}$. Moreover, let $\Pi(t)$ be the characteristic function of

$$L^\alpha(\hat{D}_k - D) = L^{-(1-\alpha)}\Big\{\sum y_i - E(n)\, f(0)\Big\}.$$

Then

$$\Pi(t) = \exp\{-L^{-(1-\alpha)}\, E(n)\, f(0)it - Np + Np\, \phi_h(t/L^{1-\alpha})\}.$$

From equation (A.2) and $E(n) = Np$,

$$\begin{aligned}\log \Pi(t) &= \tfrac{1}{2}Np\{f^{(2)}(0)\, h^2 L^{-(1-\alpha)}(it) - (4\pi)^{-1/2} f(0)h^{-1}L^{-2(1-\alpha)}t^2\} + O(\Delta)\\ &= D f(0)^{-1}\{f^{(2)}(0)\beta^2 L^{\alpha-2/5}(it) - (4\pi)^{-1/2} f(0)\beta^{-1}L^{2\alpha-4/5}t^2\} + O(\Delta).\end{aligned}$$

Put $\mu = f^{(2)}(0)\, f(0)^{-1} D\beta^2$ and $v^2 = (\beta\sqrt{\pi})^{-1}D$. Then,

$$\log \Pi(t) = \mu L^{\alpha-2/5}(it) - v^2 L^{2\alpha-4/5}t^2/2 + O(\Delta).$$

If $\alpha = 2/5$ and $L \to \infty$, then

$$\log \Pi(t) \to \mu(it) - v^2t^2/2.$$

This implies that $L^\alpha(\hat{D}_k - D)$ converges in distribution to a normal random variable with mean μ and variance v^2 if $\alpha = 2/5$. Thus, theorem 1 is proved.

References

Buckland, S. T. (1982) A note on the Fourier series model for analysing line transect data. *Biometrics*, **38**, 469–477.
——(1985) Perpendicular distance models for line transect sampling. *Biometrics*, **41**, 177–196.
——(1992) Fitting density functions with polynomials. *Appl. Statist.*, **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. *Wldlf. Monogr.*, **72**.
Chen, S. X. (1994) Studying school size effects in line transect sampling using the kernel method. *Research Report*. CSIRO Division of Fisheries, Hobart.
Drummer, T. D. and McDonald, L. L. (1987) Size bias in line transect sampling. *Biometrics*, **44**, 13–21.
Feller, W. (1957) *An Introduction to Probability Theory and Its Applications*. New York: Wiley.
Fix, E. and Hodges, J. L. (1951) Discriminatory analysis, nonparametric estimation: consistency properties. *Report 4, Project 21-49-004*. US Air Force School of Aviation Medicine, Randolph Field.
Härdle, W., Marron, J. S. and Wand, M. P. (1990) Bandwidth choice for density derivatives. *J. R. Statist. Soc.* B, **52**, 223–232.
Pollock, K. H. and Kendall, W. L. (1987) Visibility bias in aerial surveys: a review of estimation procedures. *J. Wldlf. Mangmnt*, **51**, 502–509.
Press, W. H., Flannery, B. F., Teukolsky, S. A. and Vetterling, W. T. (1989) *Numerical Recipes: the Art of Scientific Computing*. Cambridge: Cambridge University Press.
Quang, P. X. (1990) Confidence intervals for densities in line transect sampling. *Biometrics*, **46**, 459–472.
——(1993) Nonparametric estimators for variable circular plot surveys. *Biometrics*, **49**, 837–852.
Seber, G. A. F. (1982) *The Estimation of Animal Abundance*, 2nd edn. London: Griffin.
——(1992) A review of estimating animal abundance II. *Int. Statist. Rev.*, **60**, 129–166.
Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
Silverman, B. W. (1982) Algorithm AS 176: Kernel density estimation using the fast Fourier transform. *Appl. Statist.*, **31**, 93–99.
——(1986) *Density Estimation*. London: Chapman and Hall.
Stone, C. J. (1984) An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, **12**, 1285–1297.