Environmental and Ecological Statistics **7**, 285–299, 2000

Animal abundance estimation in independent observer line transect surveys

SONG XI CHEN

Department of Statistics and Applied Probability, National University of Singapore, Singapore 117543

Received March 1998; Revised February 1999

The theory of conventional line transect surveys is based on an essential assumption that 100% detection of animals right on the transect lines can be achieved. When this assumption fails, independent observer line transect surveys are used. This paper proposes a general approach, based on a conditional likelihood, which can be carried out either parametrically or nonparametrically, to estimate the abundance of non-clustered biological populations using data collected from independent observer line transect surveys. A nonparametric estimator is specifically proposed which combines the conditional likelihood and the kernel smoothing method. It has the advantage that it allows the data themselves to dictate the form of the detection function, free of any subjective choice. The bias and the variance of the nonparametric estimator are given. Its asymptotic normality is established which enables construction of confidence intervals. A simulation study shows that the proposed estimator has good empirical performance, and the confidence intervals have good coverage accuracy.

Keywords: conditional likelihood, line transect survey, kernel smoothing

1. Introduction

Line transect sampling has been used for estimating the abundance of biological populations for decades. It has been a practical approach for wildlife management. The population density D of a non-clustered biological population is defined as $D = N/A$, where N is the unknown population size and A is the area occupied by the population. To estimate D, an observer traverses a distance L along randomly generated non-overlapping transect lines within the survey area. Each object sighted from the transect lines is counted and its perpendicular distance (also called sighting distance) from the line is measured. Comprehensive reviews on line transect survey are available in Seber (1982) and Buckland *et al.* (1993).

An essential assumption in the theory of traditional line transect surveys is that 100% detection of animal objects is achieved on the transect lines. Let $g(x)$, commonly called the detection function, be the conditional probability of detecting an object given that the

1352-8505 © 2000 Kluwer Academic Publishers

object is at perpendicular distance x from the transect line. Thus, the assumption means $g(0) = 1$. This has been regarded as a reasonable assumption in line transect surveys. However, it has been revealed that the assumption may not be true all of the time, as shown by experimental survey conducted by Laake (1978) and whale surveys reported by Butterworth and Borchers (1988), Schweder (1990) and others.

To obtain information about $g(0)$ and to estimate animal abundance when $g(0) < 1$, independent observer line transect surveys have been proposed (Butterworth and Borchers, 1988). Two observers are employed to detect animals independently of each other. A third person acts as a coordinator to identify which objects are detected by both of the observers, and which by only one of them. This relates to the method of mark-recapture, as the objects sighted by both can be viewed as recaptures. It is based on this relationship that Alpizar-Jara and Pollock (1996) used variations of the Lincoln-Petersen estimators for line transect surveys by grouping the sighting distances into categories. Borchers, Zucchini and Fewster (1998) proposed mark-recapture models for line transect data; and Borchers, Buckland, Goedhart *et al*. (1998) proposed Horvitz-Thompson type estimators. Parametric assumptions for the detection functions were made by the above authors.

In this paper, a different approach, largely independent of the existing mark-recapture methods, is considered. It involves first an estimation of the effective search width by a conditional likelihood based on the detection information of the two observers. Then, estimation of the probability density function of the sighting distances is used to finalize the effective search width estimation. A general framework for estimating D is proposed, which is a generalized form of the Petersen estimator. This framework has the flexibility to be used for both parametrical and nonparametrical estimation. It also has the ability to measure the amount of non-uniform detection through a key quantity α. A nonparametric estimator based on the kernel smoothing method is specifically studied as it allows the data themselves to dictate the shape and the form of the detection functions, free of any subjective choice.

The paper is structured as follows. Section 2 introduces some notations and outlines the problem. A general estimator for D is given in Section 3. Section 4 introduces a nonparametric estimator based on the kernel method. Its properties are studied in Section 5. Confidence intervals are proposed in Section 6. Section 7 discusses how to choose the smoothing bandwidth. Laake (1978)'s stake data is analyzed in Section 8. Section 9 reports a simulation study, followed by a general discussion in Section 10.

2. Notations and outlines

Let $X_1, X_2, \ldots, X_n$ be the perpendicular sighting distances obtained from n sightings from a conventional single observer line transect survey, and f be the common probability density function of the sighting distances. It is convenient to use signed sighting distances for the nonparametric estimator proposed in this paper. Signed sighting distances are natural in line transect surveys as sightings are made on both sides of the transect lines. Let w be the maximum perpendicular distance on either side of the transect line, and $g(x)$ be the detection function. Let P_0 be the probability of detecting an animal. Then, as shown in Seber (1982), $P_0 = \int_{-w}^{w} g(x)dx/(2w)$. If the transect lines are allocated randomly,

$$E(n) = NP_0 = LD \int_{-w}^{w} g(x)dx. \tag{2.1}$$

Then, a classical estimator for D is

$$\hat{D} = \frac{n}{L\hat{\mu}} \tag{2.2}$$

where $\hat{\mu}$ is an estimator for the effective search width $\mu = \int_{-w}^{w} g(x)dx$.

A well known relationship between the density f and the detection function is

$$f(x) = g(x)/\mu. \tag{2.3}$$

Under the assumption that $g(0) = 1$, the estimator in (2.2) becomes

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

In this case, the estimation of D becomes an estimation of $f(0)$. If $g(0) < 1$, the estimator in (2.4) has severe bias and should not be used. However, the original estimator (2.2) can still be used, and the estimation becomes primarily one for μ.

To gain information about μ under $g(0) < 1$, two independent observers are employed to make independent detection. Suppose n_1 and n_2 animal objects are detected by the two observers respectively, and n_{11} objects are detected by both. Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be the perpendicular sighting distances by the two observers respectively. Let g_i and f_i, $i = 1$ and 2, be the detection and the probability density functions of the ith observer respectively.

Let $n_{10} = n_1 - n_{11}$ be the number of sightings by the first observer only, and $n_{01} = n_2 - n_{11}$ be that by the second observer only. Then $n = n_{10} + n_{01} + n_{11}$ is the total number of sightings by both observers. Define

$$\mu_i = \int_{-w}^{w} g_i(x)dx$$

to be the effective search width for the i-th observer. Let $g(x) = g_1(x) + g_2(x) - g_1(x)g_2(x)$ be the joint detection function. Then the joint effective search width $\mu = \mu_1 + \mu_2 - \alpha\mu_1\mu_2$ where

$$\alpha = \int_{-w}^{w} f_1(x)f_2(x)dx.$$

3. General estimators for μ and D

The approach proposed in this paper first estimates μ by the conditional likelihood method outlined below, and then substitutes the estimate back to (2.2) to obtain estimator for D.

Conditional on n, (n_{10}, n_{01}, n_{11}) is distributed as a multinomial distribution

$$Mult(n, P_{10}, P_{01}, P_{11})$$

where

$$P_{10} = (\mu_1 - \alpha\mu_1\mu_2)/\mu, \quad P_{01} = (\mu_2 - \alpha\mu_1\mu_2)/\mu, \quad \text{and} \quad P_{11} = \alpha\mu_1\mu_2/\mu$$

are the conditional probabilities. Using the well known conditional likelihood method, the conditional likelihood for μ_1, μ_2 and α is

$$L(\mu_1, \mu_2, \alpha|n) = C\left(\frac{\mu_1 - \alpha\mu_1\mu_2}{\mu}\right)^{n_{10}}\left(\frac{\mu_2 - \alpha\mu_1\mu_2}{\mu}\right)^{n_{01}}\left(\frac{\alpha\mu_1\mu_2}{\mu}\right)^{n_{11}} \tag{3.1}$$

where C is a constant free of μ_1, μ_2 and α. The above conditional likelihood is part of the full likelihood given in Borchers, Zucchini and Fewster (1998). The sighting distances do not appear in the above conditional likelihood, but will be utilized separately in the estimation of α later. This treatment of the sighting distances provides two advantages. One is that the computation is largely simplified as the full likelihood is not used. The other is that the estimation can be carried out nonparametrically.

Given α, the values of μ_1 and μ_2 which maximize the conditional likelihood (3.1) are, respectively,

$$\mu_1 = n_{11}/(\alpha n_2) \quad \text{and} \quad \mu_2 = n_{11}/(\alpha n_1).$$

Thus

$$\mu = \frac{1}{\alpha}\frac{nn_{11}}{n_1 n_2}. \tag{3.2}$$

To develop an estimator for μ, α has to be estimated first. Notice that

$$\alpha = \int_{-w}^{w} f_1(x)f_2(x)dx = E\{f_1(X_2)\} = E\{f_2(X_1)\}$$

where X_1 and X_2 represent the sighting distance of the first and the second observers respectively. Using the method of moments, a general estimator for α is either

$$\hat{\alpha}_1 = n_2^{-1}\sum_{j=1}^{n_2}\hat{f}_1(X_{2j}) \quad \text{or} \quad \hat{\alpha}_2 = n_1^{-1}\sum_{j=1}^{n_1}\hat{f}_2(X_{1j}) \tag{3.3}$$

where $\hat{f}_1$ and $\hat{f}_2$ are probability density estimators for f_1 and f_2 respectively. If the kernel density estimation is used to estimate f_1 and f_2, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are equivalent to each other as shown in the next section.

Substitute the method of moments estimator $\hat{\alpha}$ into (3.2), an estimator for μ is

$$\hat{\mu} = \frac{1}{\hat{\alpha}}\frac{nn_{11}}{n_1 n_2} \tag{3.4}$$

and that for D is

$$\hat{D} = \frac{n}{L\hat{\mu}} = \frac{\hat{\alpha}}{L}\frac{n_1 n_2}{n_{11}}. \tag{3.5}$$

Note that $n_1 n_2/n_{11}$ is just the Petersen estimator for N used in mark-recapture experiments by assuming uniform probability of detecting animals. In a line transect survey, the detection probability is not uniform and depends on the perpendicular distance from the transect lines. This non-uniform detection is taken care of by α. The key of this paper is to estimate $\alpha = \int f_1(x)f_2(x)dx$. Other authors (Butterworth and Borchers, 1988) proposed scheme of estimation involving $\int g_1(x)g_2(x)dx$, which can not be formulated in a simple form like (3.3) and requires estimation of two detection functions.

It remains only to specify the probability density estimator $\hat{f}_1$ used in (3.3). The main attraction of the proposed approach is that both the parametric and nonparametric estimators for $\hat{\alpha}$ can be used. A parametric estimator assumes parametric forms for the detection functions g_1 and g_2, and α is estimated parametrically. An outline of the parametric implementation is discussed in Section 10. A nonparametric estimator, for instance the kernel or the Fourier series estimators, requires fewer assumptions about the detection process.

4. A kernel estimator

In this section a nonparametric estimator for α, which is based on kernel estimation for f_1, is considered. The kernel method has been used in line transect surveys by Buckland (1992), Chen (1996a) and (1996b), and Mack and Quang (1998), and in point transect surveys by Quang (1993). A standard kernel density estimator for f_1 is

$$\hat{f}_1(x) = (n_1 h)^{-1} \sum_{i=1}^{n_1} K\left(\frac{x - X_{1i}}{h}\right) \tag{4.1}$$

where K is a kernel function and h is a smoothing bandwidth which determines the amount of smoothing used. The kernel function K is usually a symmetric probability density function itself.

An estimator for α based on the kernel density estimator (4.1) is

$$\hat{\alpha} = n_2^{-1} \sum_{j=1}^{n_2} \hat{f}_1(X_{2j}) = (n_1 n_2 h)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K\left(\frac{X_{2j} - X_{1i}}{h}\right). \tag{4.2}$$

When only the absolute sighting distances are available, the reflected kernel density estimator for f_1 used in Buckland (1992) and Chen (1996a) should be used, and

$$\hat{f}_1(x) = (n_1 h)^{-1} \sum_{1}^{n_1} \left\{ K\left(\frac{x - X_{1i}}{h}\right) + K\left(\frac{x + X_{1i}}{h}\right) \right\}.$$

The corresponding estimator for α is

$$\hat{\alpha} = n_2^{-1} \sum_{j=1}^{n_2} \hat{f}_1(X_{2j}) = (n_1 n_2 h)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left\{ K\left(\frac{X_{2j} - X_{1i}}{h}\right) + K\left(\frac{X_{2j} + X_{1i}}{h}\right) \right\}.$$

5. Properties of $\hat{\alpha}$ and $\hat{D}$

In evaluating the performance of $\hat{\alpha}$ and $\hat{D}$ we assume that

(i) K is a positive symmetric kernel; (ii) $h \to 0$ as $L \to \infty$; and (iii) f_i have bounded third derivative and $\int_{-w}^{w} |f_i(x)| dx < \infty$ for $i = 1$ and 2. (5.1)

We first study the properties of $\hat{\alpha}$. Conditioning on n_1, and from the standard results on kernel density estimation (for instance those given in Silverman, 1986),

$$E\{\hat{f}_1(x)|n_1\} = f_1(x) + \frac{1}{2}h^2\sigma_k^2 f_1''(x) + O(h^3) \quad \text{and} \tag{5.2}$$

$$\text{Var}\{\hat{f}_1(x)|n_1\} = (n_1 h)^{-1} R(K) f_1(x)\{1 + o(1)\} \tag{5.3}$$

where $\sigma_k^2 = \int t^2 K(t)dt$ and $R(K) = \int K^2(t)dt$.

Let $\widetilde{X}_2 = (X_{21}, \ldots, X_{2n_2})$ be the second sighting sample. Using the conditional expectation and (5.2)

$$\begin{aligned}
E(\hat{\alpha}) &= E\{E(\hat{\alpha}|\widetilde{X}_2, n_2)\} = E\left[n_2^{-1}\sum_{j=1}^{n_2} E\{\hat{f}_1(X_{2j})|\widetilde{X}_2, n_2\}\right] \\
&= E\left[n_2^{-1}\sum_{j=1}^{n_2}\{f_1(X_{2j}) + \frac{1}{2}h^2\sigma_k^2 f_1''(X_{2j}) + O(h^3)\}\right] \\
&= \alpha + \frac{1}{2}h^2\sigma_k^2\int_{-w}^{w} f_1''(x) f_2(x)dx + O(h^3).
\end{aligned} \tag{5.4}$$

So, $\hat{\alpha}$ is an asymptotically unbiased estimator for α as $L \to \infty$.

By conditional expectation

$$\text{Var}(\hat{\alpha}) = E\{\text{Var}(\hat{\alpha}|\widetilde{X}_2, n_2, n_1)\} + \text{Var}\{E(\hat{\alpha}|\widetilde{X}_2, n_2, n_1)\}. \tag{5.5}$$

Using (5.3)

$$\begin{aligned}
\text{Var}(\hat{\alpha}|\widetilde{X}_2, n_2, n_1) &= \text{Var}\{n_2^{-1}\sum_{j=1}^{n_2}\hat{f}_1(X_{2j})|\widetilde{X}_2, n_2, n_1\} \\
&= n_2^{-2}\sum_{j=1}^{n_2}\text{Var}\{\hat{f}_1(X_{2j})|\widetilde{X}_2, n_1\} \\
&= (n_1 n_2^2 h)^{-1}\sum_{j=1}^{n_2} R(K) f_1(X_{2j})\{1 + o(1)\}.
\end{aligned}$$

Thus, the first term on the right side of (5.5) is

$$E\{\text{Var}(\hat{\alpha}|X^2, n_2, n_1)\} = \alpha R(K) h^{-1} E\{(n_1 n_2)^{-1}\}. \tag{5.6}$$

From the derivation for (5.4),

$$\begin{aligned}\mathrm{Var}\{E(\hat{\alpha}|\widetilde{X}_2, n_2, n_1)\} &= \mathrm{Var}\left[n_2^{-1}\sum_{j=1}^{n_2}\{f_1(X_{2j}) + \frac{1}{2}h^2\sigma_k^2 f_1''(X_{2j}) + O(h^3)\}\right] \\ &= \mathrm{Var}\{f_1(X_{2j})\}E(n_2^{-1})\{1+o(1)\} \\ &= \left\{\int_{-w}^{w} f_1^2(x)f_2(x)dx - \alpha^2\right\}E(n_2^{-1})\left\{1+o(1)\right\}. \end{aligned} \quad (5.7)$$

Combining (5.6) and (5.7) and ignoring the remainder terms we have

$$\mathrm{Var}(\hat{\alpha}) \approx \left[\alpha R(K)h^{-1}E\{(n_1n_2)^{-1}\} + \left\{\int_{-w}^{w} f_1^2(x)f_2(x)dx - \alpha^2\right\}E(n_2^{-1})\right]. \quad (5.8)$$

As $\hat{\alpha}$ is a two-sample U-statistic,

$$\hat{\alpha} \to N\{E(\hat{\alpha}), \mathrm{Var}(\hat{\alpha})\} \quad \text{in distribution} \quad \text{as} \quad L \to \infty \quad (5.9)$$

based on a result in Lehmann (1951).

Let $\gamma = n_1n_2/n_{11}$ be the Petersen estimator. It may be shown that

$$\gamma/L \to D/\alpha \quad \text{in probability} \quad \text{as} \quad L \to \infty. \quad (5.10)$$

Combining (5.9) and (5.10) we have

$$\hat{D} \to N\{D, \mathrm{Var}(\hat{D})\} \quad \text{in distribution} \quad \text{as} \quad L \to \infty \quad (5.11)$$

where $\mathrm{Var}(\hat{D}) = D^2\{\mathrm{Var}(\hat{\alpha})/E^2(\hat{\alpha}) + \mathrm{Var}(\gamma)/E^2(\gamma)\} = D^2\{cv^2(\hat{\alpha}) + cv^2(\gamma)\}$ as $E(\gamma\hat{\alpha}) = E(\gamma)E(\hat{\alpha})$.

6. Confidence intervals

The asymptotic result in (5.11) can be used to construct confidence intervals for D. However, $\mathrm{Var}(\hat{D})$ has to be estimated which, in turn, requires estimation of $\mathrm{Var}(\hat{\alpha})$ and $\mathrm{Var}(\gamma)$.

An estimator for $\mathrm{Var}(\hat{\alpha})$ can be derived based on the asymptotic variance given in (5.8). Let $\beta = \int_{-w}^{w} f_1^2(x)f_2(x)dx = E\{f_1^2(X_{2j})\}$, which can be estimated by

$$\hat{\beta} = n_2^{-1}\sum_{j=1}^{n_2}\hat{f}_1^2(X_{2j}).$$

An estimator for $\mathrm{Var}(\hat{\alpha})$ is

$$\widehat{\mathrm{Var}}(\hat{\alpha}) = \hat{\alpha}R(K)h^{-1}(n_1n_2)^{-1} + \{\hat{\beta} - \hat{\alpha}^2\}n_2^{-1}. \quad (6.1)$$

Let $P_1 = P_{10} + P_{11}$ and $P_2 = P_{01} + P_{11}$ where P_{01}, P_{10} and P_{11} are the conditional probabilities used in Section 3 to derive the conditional likelihood. Derivation (deferred until the Appendix) shows that, under the multinomial distribution of (n_{10}, n_{01}, n_{11}) given $n = n_{10} + n_{01} + n_{11}$,

$$\mathrm{Var}(\gamma) = vE(n) + \mathrm{Var}(n)P_1^2P_2^2/P_{11}^2 \tag{6.2}$$

where $v = P_1P_2(P_1P_2/P_{11} - 1 - P_{01}P_{10})/P_{11}^2$. If n is Binomial (N, P) distributed where $P = \mu/w = (\mu_1 + \mu_2 - \alpha\mu_1\mu_2)/w$ then

$$\mathrm{Var}(\gamma) = vE(n) + E(n)(1 - \mu/w)P_1^2P_2^2/P_{11}^2.$$

Let $\hat{P}_i = n_i/n$ for $i = 1$ and 2 and $\hat{P}_{jk} = n_{jk}/n$ for $j = 0$ or 1 and $k = 0$ or 1 respectively; $\hat{v}$ be an estimator for v by substituting the above estimated probabilities into the formula for v given above, and $\hat{\mu} = n/(\gamma\hat{\alpha})$. Then, an estimator for $\mathrm{Var}(\gamma)$ is

$$\widehat{\mathrm{Var}}(\gamma) = \hat{v}n + (1 - \hat{\mu}/w)\gamma^2/n. \tag{6.3}$$

Combining (6.1) and (6.3) we have

$$\widehat{\mathrm{Var}}(\hat{D}) = \hat{D}^2(\widehat{\mathrm{Var}}(\hat{\alpha})/\hat{\alpha}^2 + \widehat{\mathrm{Var}}(\gamma)/\gamma^2). \tag{6.4}$$

Now a confidence interval for D with an asymptotic confidence level ξ is

$$\left(\hat{D} - z_{(1+\xi)/2}\sqrt{\widehat{\mathrm{Var}}(\hat{D})}, \hat{D} + z_{(1+\xi)/2}\sqrt{\widehat{\mathrm{Var}}(\hat{D})}\right) \tag{6.5}$$

where $z_{(1+\xi)/2}$ is the $(1 + \xi)/2$-th quantile of the standard normal distribution.

If the numbers of sightings on duplicate transect lines are available, the Binomial assumption for n can be dropped and

$$\widehat{\mathrm{Var}}(\gamma) = \hat{v}n + c\hat{v}^2(n)\gamma^2$$

should be used to replace (6.3).

The above method of constructing confidence intervals is based on the delta method. The bootstrap can be used as well for confidence intervals. However, we will not discuss it further here as the bootstrap has been fairly well covered in the literature.

7. Choosing the smoothing bandwidth

The smoothing bandwidth h can be chosen to minimize either the mean squared error of $\hat{\alpha}$ or that of $\hat{D}$. It can be shown that the bandwidths produced by minimizing the two mean square errors are of the same order of magnitude. Since $\hat{D} = \hat{\alpha}\gamma$ and only $\hat{\alpha}$ is involved with the kernel smoothing, the bandwidth is chosen to minimize the mean square errors of $\hat{\alpha}$.

From (5.4) and (5.8), the mean square error of $\hat{\alpha}$ is approximately

$$MSE\{\hat{\alpha}\} = \frac{1}{4}h^4\sigma_k^4\left\{\int_{-w}^{w} f_1''(x)f_2(x)dx\right\}^2 + \alpha R(K)h^{-1}E\{n_1n_2\}^{-1}.$$

The optimal bandwidth which minimizes the above mean square error is

$$h^* = \{\alpha R(K)\sigma_k^{-4}\}^{-1/5}\left\{\int_{-w}^{w} f_1''(x)f_2(x)dx\right\}^{2/5}[E\{n_1n_2\}^{-1}]^{1/5}, \tag{7.1}$$

which implies the mean square error of $\hat{\alpha}$ is of the order of $N^{-8/5}$ which is much smaller than the standard order of $N^{-4/5}$ in kernel density estimation. Thus, the estimation for α, which is a functional of the densities, is much more accurate than the probability density estimation and is less demanding on the bandwidth. Note that the mean square error of $\hat{D}$ is larger than $O(N^{-8/5})$ due to the involvement of γ.

In practice a useful bandwidth can be derived by estimating the unknown quantities α, $\int_{-w}^{w} f_1''(x) f_2(x) dx$ and $E\{n_1 n_2\}$ in (7.1). The last one can be estimated simply by $\{n_1 n_2\}$. Values for α and $\int_{-w}^{w} f_1''(x) f_2(x) dx$ can be obtained by assuming $f_i|(x|\tilde{z}) \sim N\{0, \sigma_i^2\}$, which give

$$\alpha = 1/\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}$$

$$\eta(f_1'', f_2) = -1/\{\sqrt{2\pi}(\sigma_1^2 + \sigma_2^2)^{3/2}\}.$$

Let A_i be estimates for σ_i. Then, the bandwidth is

$$h = \{R(K)\sigma_k^{-4}\sqrt{2\pi}\}^{1/5}\sqrt{A_1^2 + A_2^2}(n_1 n_2)^{-1/5}. \quad (7.2)$$

The above method is similar to the method of referring to a standard distribution proposed by Silverman (1986) for kernel density estimation. In practice, robust estimates for σ_i can be used, for instance

$$A_i = \min\{s_i, q_i/1.349\}$$

where s_i and q_i are the sample standard deviation and the sample inter-quantile range respectively.

8. Analyzing the stake data

In this section we analyze Laake's (1978) stake data using the proposed method. One hundred and fifty ($N = 150$) wooden stakes were randomly placed in a rectangular area. Eleven observers were asked to traverse a transect line 1000 m long with $w = 20$. The true value of D is known to be 0.00375. The stake data is well known in the study of line transect surveys. Various estimates have been developed. There are negative biases associated with these estimates, which some authors believe were due to $g(0) < 1$ for some of the observers.

Here we analyze the data from the same two observers used in Alpizar-Jara and Pollock (1996). There were 72 (n_1) and 48 (n_2) sightings by the two observers respectively, and 39 (n_{11}) sighted by both.

The bandwidth used was $h = 2.849$ according to (7.2), and $\hat{\alpha} = 0.0722$. The resulting estimate for D is 0.0032 with a standard error of 0.0003. The corresponding estimate for N is 128.05 with a standard error of 12.02. The true values for D and N were all contained in the 95% confidence intervals. The estimates for N given above were more accurate than that given in Alpizar-Jara and Pollock (1996) where $\hat{N} = 105$ was reported. This was because they used grouped data in their estimation rather than the exact distance used here. The Petersen estimator give an estimate of 88.62 for N. It severely underestimates as the heterogeneity in detection is ignored. In Fig. 1, the stake data is presented as histograms together with kernel estimates for density functions of the sighting distances.

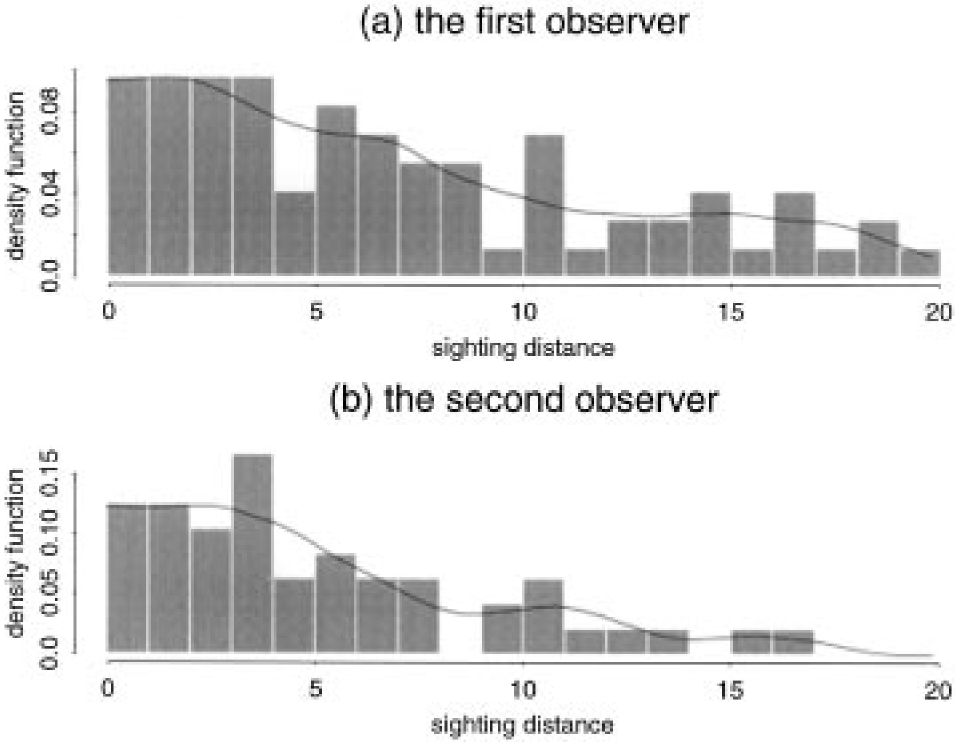

Figure 1. The Stake data and probability density estimates.

To enhance our understanding of the data, we calculated estimates for $g_1(0)$ and $g_2(0)$ as given by

$$\hat{g}_i(0) = \hat{f}_i(0)\hat{\mu}_i \quad \text{for} \quad i = 1, 2.$$

We found that $\hat{g}_1(0) = 1.066$ and $\hat{g}_2(0) = 0.916$. That $\hat{g}_1(0)$ is larger than one is due to the fact that the estimate is not range respecting and it should be regarded as one. Their standard errors, obtained by the bootstrap, were 0.078 and 0.095 respectively. There was some evidence to suggest that $g_2(0)$ is less than 1, but the evidence was not very strong. A close study of the data sets reveals that the first observer only missed one stake at 1.32 m among all the stakes less than 3.3 m, and the second observer missed four stakes less than 3.3 m, but only missed one less than 1.32 m at 0.31 m. In the kernel estimation for $g_i(0)$, only the data within a small area near zero are used rather than the entire data set used by parametric methods.

We also analyzed the data under the conventional line transect model assuming $g_i(0) = 1$ by pooling the sighting distances from the two samples. The kernel estimates for D and N were 0.00347 and 138.84 respectively. The improvement in the estimate for N suggested that $g_i(0)$ are close to one. In the case where $g_i(0)$ are actually one, using models which assume otherwise leads to loss of accuracy. However, as shown in the simulation study reported in the next section this loss of accuracy is much less than that found when $g_i(0)$ are in fact less than 1 and conventional line transect models are used.

9. Simulation results

In this section we present some simulation results designed to provide empirical outcomes to the theory developed in this paper. The aims of the simulation study were to investigate the bias and variance of the nonparametric estimator $\hat{D}$, and the coverage accuracy of the proposed confidence intervals. The results were based on 1000 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 $D = N/(Lw)$ was known. We fixed $D = 0.15$ and $w = 10$, and the population size $N = 300$ and 600.

Two scenarios of detection functions were considered to generate the sighting distances in the simulation. In the first scenario, monotonic functions

$$g_1(x) = 0.7\exp\{-(bx)^a\} \quad \text{and} \quad g_2(x) = 0.6\exp\{-(bx)^a\}\} \tag{9.1}$$

were used as the detection functions for the two observers respectively. These two detection functions were simple modifications of the usual generalized exponential power series detection functions used in conventional line transect surveys. In the second scenario, we chose

$$g_1(x) = 0.7\exp\{-(b|x-2|)^a\} \quad \text{and} \quad g_2(x) = 0.6\exp\{-(b|x-2|)^a\}. \tag{9.2}$$

These two are not monotonic and have modes at $x = 2$. They reflect to some degree a situation confronted in some aerial surveys where the detection for the strip on the flying path is relatively low. In defining the detection functions above, a and b were the shape and scale parameters respectively. We fixed $b = 0.5$ and chose the shape parameter $a = 1.0$, 1.5, 2.0 and 2.5. Clearly, in both scenarios $g_i(0) < 1$ for $i = 1$ and 2. The smoothing bandwidth h was chosen according to (7.2).

Table 1 contains the point estimates $\hat{D}$ and their standard errors, and the coverage of the 95% confidence intervals for D and their length. We observe that the proposed estimator $\hat{D}$ was quite satisfactory for the two scenarios of detection functions and the proposed bandwidth selection method worked well. There were two surprises in the simulation results. One is that the estimates had quite small bias in the second scenario of detection functions even though the bandwidths were chosen by referring to the normal distributions. The other is that quite satisfactory results were observed when $a = 1$ where f_i were not smooth at $x = 0$, which violated a condition assumed in (5.1). Both were results of α being a functional of the densities and having a much smaller mean square error of $O(N^{-8/5})$ which implies it is easier to be estimated. We observe, for both the point estimates and confidence intervals, that when the size of the population N increased from 300 to $N = 600$ (which increased the sample sizes n_1 and n_2), the biases of the point estimates, the standard errors of the estimates and the length of the confidence intervals were reduced.

For comparison purposes, we also supplied kernel estimates for D when $g(0) = 1$ is assumed. There were severe biases in those estimates. The coverage of the confidence

Table 1. Point estimates ($\hat{D}$) and their standard errors (SE), and the empirical coverage of 95% confidence intervals and their length for $D = 0.15$. The point estimates ($\hat{D}_{g(0)=1}$) for D when $g(0) = 1$ is assumed are also given.

(a) $g_1(x) = 0.7\exp\{-(0.5x)^a\}$ and $g_2(x) = 0.6\exp\{-(0.5x)^a\}$

	a = 1.0		*a = 1.5*		*a = 2*		*a = 2.5*	
N	*300*	*600*	*300*	*600*	*300*	*600*	*300*	*600*
$\hat{D}$	0.154	0.152	0.155	0.153	0.155	0.153	0.155	0.153
SE	0.036	0.022	0.030	0.020	0.029	0.020	0.028	0.019
$\hat{D}_{g(0)=1}$	0.126	0.127	0.130	0.130	0.132	0.132	0.132	0.132
SE	0.041	0.031	0.045	0.034	0.049	0.036	0.050	0.037
Coverage	0.937	0.960	0.955	0.949	0.936	0.945	0.938	0.941
Length	0.140	0.094	0.125	0.084	0.115	0.077	0.109	0.073
n	35.8	71.3	32.5	65.2	31.9	63.6	31.9	63.8
$\hat{P}_0$	0.116	0.117	0.105	0.107	0.103	0.104	0.103	0.104

(b) $g_1(x) = 0.7\exp\{-(0.5|x-2|)^a\}$ and $g_2(x) = 0.6\exp\{-(0.5|x-2|)^a\}$

	a = 1.0		*a = 1.5*		*a = 2*		*a = 2.5*	
N	*300*	*600*	*300*	*600*	*300*	*600*	*300*	*600*
$\hat{D}$	0.147	0.148	0.149	0.149	0.150	0.150	0.150	0.150
SE	0.022	0.014	0.018	0.013	0.017	0.012	0.016	0.012
$\hat{D}_{g(0)=1}$	0.071	0.068	0.071	0.069	0.074	0.071	0.075	0.072
SE	0.025	0.020	0.027	0.021	0.029	0.023	0.029	0.023
Coverage	0.908	0.932	0.910	0.918	0.917	0.926	0.910	0.929
Length	0.083	0.057	0.072	0.050	0.066	0.046	0.071	0.43
n	58.2	115.8	57.4	115.3	58.8	117.6	59.82	119.9
$\hat{P}_0$	0.198	0.196	0.193	0.193	0.196	0.196	0.198	0.20

intervals were very poor, and are not reported here. Generally speaking, the effect of wrongly assuming $g(0) = 1$ when in fact $g(0) < 1$ is very severe indeed.

10. Discussion

In this paper we have proposed a framework for estimating animal abundance from independent observer line transect surveys. The framework allows both parametric and nonparametric methods. A nonparametric estimator has been specifically proposed and studied. Its theoretical properties are investigated and good empirical performance has been observed. An attractive feature of this nonparametric estimator is that it is free of parametric assumptions on the detection process. A limitation of the nonparametric method is that it requires ''enough'' sample size in order to have a reasonable density estimate. This is not a problem in the current univariate situation. It would be more significant when other covariates are introduced. However, the curse of dimensions is not just a problem for nonparametric estimators.

If parametric forms for the detection functions are available, the maximum likelihood method can be used to estimate the unknown parameters. Suppose $g_i(x) = g_i(x; \theta^i)$ where $\theta^i = (\theta^i_1, \ldots, \theta^i_k)$ are the parameters for $i = 1$ and 2. Then $\mu_i = \int_{-w}^{w} g_i(x; \theta^i)dx$ can be expressed as functions of θ^i, say $\mu_i(\theta^i)$. Then, parametric forms of probability density functions f_i of the observed sighting distances can be established as $f_i(x; \theta^i) = g_i(x; \theta^i)/\mu_i(\theta^i)$ which produces the likelihood function for each of the sighting distance samples. After computing the maximum likelihood estimate, say $\hat{\theta}^i$, α can be estimated by either

$$n_2^{-1} \sum f_1(X_{2j}; \hat{\theta}^1) \quad \text{or} \quad n_1^{-1} \sum f_2(X_{1j}; \hat{\theta}^2).$$

Then, substituting the estimate for α, an estimate for D is obtained. Here only the partial likelihood based on the sighting distances is used. The reason for not using the full likelihood is that the partial likelihood is easier to compute. This may lead to some loss of efficiency in the estimation of α. However, the MSE of the parametric estimator for α should be at a smaller order than $N^{-8/5}$ which is the order of the nonparametric estimator $\hat{\alpha}$ as shown in Section 7. As we can achieve a MSE of such small order, there should be little loss of efficiency.

The attractions of the proposed framework are (i) its easy computation compared with the full likelihood approach, (ii) its flexibility to be carried out either parametrically or nonparametrically, (iii) its ability to measure the amount of non-uniform detection, and (iv) its linkage to the Petersen estimator.

The approach proposed in this paper is developed from univariate line transect surveys with the sighting distance as the only variable. It can be generalized to independent line transect surveys that are subject to heterogeneity in detection due to other covariates. The idea is to fix the other covariates that introduce heterogeneity first, apply the current approach on the sighting distance to obtain local abundance estimates and then derive the total abundance estimates by integrating or summing up the local estimates. The details of the extension is reported in Chen (1999).

Acknowledgments

The author is grateful to Dr Jeff Laake for allowing him to use the Stake data in this paper, and thanks the referees for constructive comments.

Appendix: Derivation of (6.2)

Let $\hat{P}_i = n_i/n$ and $\hat{P}_{jk} = n_{jk}/n$ for $i = 1$ and 2, and $j, k = 0$ and 1. Then

$$\gamma = nD(\hat{P}_1, \hat{P}_2, \hat{P}_{11})$$

where $D(x, y, z) = xy/z$. By a Taylor expansion of $D(\hat{P}_1, \hat{P}_2, \hat{P}_{11})$ near (P_1, P_2, P_{11}), we have

$$\gamma = n\left\{\frac{P_1P_2}{P_{11}} + \frac{P_2}{P_{11}}(\hat{P}_1 - P_1) + \frac{P_1}{P_{11}}(\hat{P}_2 - P_2) - \frac{P_1P_2}{P_{11}^2}(\hat{P}_{11} - P_{11}) \right.$$
$$- \frac{P_2}{P_{11}^2}(\hat{P}_1 - P_1)(\hat{P}_{11} - P_{11})$$
$$- \frac{P_1}{P_{11}^2}(\hat{P}_2 - P_2)(\hat{P}_{11} - P_{11}) + \frac{1}{P_{11}}(\hat{P}_1 - P_1)(\hat{P}_2 - P_2) + \frac{P_1P_2}{P_{11}^3}(\hat{P}_{11} - P_{11})^2$$
$$\left. + O_p(n^{-3/2})\right\}$$

the above Taylor expansion by the delta method, we have

$$\mathrm{Var}(\gamma|n) = n^2\left\{\frac{P_2^2}{P_{11}^2}\mathrm{Var}(\hat{P}_1) + \frac{P_1^2}{P_{11}^2}\mathrm{Var}(\hat{P}_2) - \frac{2P_1P_2^2}{P_{11}^3}COV(\hat{P}_1, \hat{P}_{11}) \right.$$
$$\left. - \frac{2P_1^2P_2}{P_{11}^3}COV(\hat{P}_2, \hat{P}_{11}) + \frac{2P_1P_2}{P_{11}^2}COV(\hat{P}_1, \hat{P}_2) + \frac{P_1^2P_2^2}{P_{11}^4}\mathrm{Var}(\hat{P}_{11}) + o(n^{-1})\right\}.$$

Notice that $COV(\hat{P}_1, \hat{P}_{11}) = P_{11}P_{01}n^{-1}, COV(\hat{P}_2, \hat{P}_{11}) = P_{11}P_{10}n^{-1}, COV(\hat{P}_1, \hat{P}_2) = -P_{10}P_{01}n^{-1}$, $\mathrm{Var}(\hat{P}_i) = P_i(1 - P_i)n^{-1}$ for $i = 1$ and 2, and $\mathrm{Var}(\hat{P}_{11}) = P_{11}(1 - P_{11})n^{-1}$. Substituting these into the above Equation (6.2) will be obtained after some simple algebra.

References

Alpizar-Jara, R. and Pollock, K.H. (1996) A combination line transect and capture-recapture sampling models for multiple observers in aerial surveys. *Environmental and Ecological Statistics,* **3**, 311–27.

Borchers, D.L., Buckland, S.T., Goedhart, P.W., Clark, E.D., and Cumerworth, S.L. (1998) Horvitz-Thompson estimators for double-platform line transect surveys. *Biometrics*, **54**, 1221–37.

Borchers, D.L., Zucchini, W., and Fewster, R. (1998) Mark-recapture models for line transect surveys. *Biometrics*, **54**, 1207–1220.

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

Butterworth, D.S. and Borchers (1988) Estimation of $g(0)$ for Minke schools from results of independent observer experiments on the 1985/86 and 1986/87 IWC/IDCR Antarctic assessment cruise, 1978/79. *Report of the International Whaling Commission*, **38**, 301–13.

Chen, S.X. (1996a) Kernel estimates for density of a biological population using line transect sampling. *Journal of Royal Statistical Society*, Series C, **45**, 135–50.

Chen, S.X. (1996b) Studying school size effects in line transect sampling using the kernel method. *Biometrics*, **52**, 1283–94.

Chen, S.X. (1999) Estimation in independent observer line transect surveys for clustered populations. *Biometrics*, **55**, 754–59.

Laake, J.L. (1978) *Line transect estimators robust to animal movement, MS thesis*, Utah State University, Logan, UT, USA, 55pp.

Lehmann, E.L. (1951) Consistency and unbiasedness of certain nonparametric tests. *The Annals of Mathematics Statistics,* **22**, 165–79.

Mack, Y.P. and Quang, P.X. (1998) Kernel methods in line and point transect sampling. *Biometrics,* **54**, 606–19.

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. (1993) Nonparametric estimators for variable circular plot survey. *Biometrics,* **49**, 837–52.

Seber, G.A.F. (1982) *The Estimation of Animal Abundance*, 2nd edn. Griffin, London.

Schweder, T. (1990) Independent observer experiments to estimate the detection function in line transect surveys of whales. *Report of the International Whaling Commission*, **40**, 349–55.

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

Biographical sketch

Song Xi Chen is an Associate Professor in the Department of Statistics and Applied Probability at the National University of Singapore. After receiving his doctrine from Australian National University, he worked as a statistician for CSIRO Division of Fisheries from 1992 to 1994 and was responsible for the Aerial Line Transect Survey of the Southern Bluefin Tuna in the Great Australian Bight. He was a lecturer/Senior Lecturer at La Trobe University from 1995 to 2000. His research interests are in the estimation of animal abundance and recreational fishing effort (using the bus-route survey), besides his other interests in the empirical likelihood and nonparametric curve estimation.