BIOMETRICS 55, 754–759
September 1999

Estimation in Independent Observer Line Transect Surveys for Clustered Populations

Song Xi Chen

Department of Statistical Science, La Trobe University, Victoria 3083, Australia
email: song.chen@latrobe.edu.au

SUMMARY. This paper introduces a framework for animal abundance estimation in independent observer line transect surveys of clustered populations. The framework generalizes an approach given in Chen (1999, *Environmental and Ecological Statistics* **6**, in press) to accommodate heterogeneity in detection caused by cluster size and other covariates. Both parametric and nonparametric estimators for the local effective search widths, given the covariates, can be derived from the framework. A nonparametric estimator based on conditional kernel density estimation is proposed and studied owing to its flexibility in modeling the detection functions. A real data set on harbor porpoise in the North Sea is analyzed.

KEY WORDS: Animal abundance; Conditional likelihood; Kernel smoothing; School size.

1. Introduction

When conventional line transect surveys (Seber, 1982; Buckland et al., 1993) are used to estimate the abundance of clustered animal populations, a single observer or observation team is employed to traverse a distance L along randomly allocated transect lines within the survey area. Each cluster detected from the transect lines is counted. Its perpendicular distance from the line, the cluster size, and some other covariates that may affect the detection of the cluster are recorded.

Conventional line transect surveys rely on a fundamental assumption of 100% detection of animal clusters on the transect lines regardless of the cluster size and the other covariates. However, it is not valid in some surveys, as shown in whale surveys conducted by the International Whaling Commission and studied by Butterworth and Borchers (1988), Schweder (1990), and others, and a harbor porpoise survey reported and analyzed by Borchers et al. (1998, in press).

Independent observer line transect surveys have been proposed to get rid of this fundamental assumption. Two observers or observation teams are employed to detect animal clusters independently. A third person acts as a coordinator to identify the clusters that are the sightings common to both observers and those that were detected by only one of them. The independent observer surveys are closely connected to the mark-recapture experiments, as the objects sighted by both can be viewed as recaptures. Based on this connection, Alpizar-Jara and Pollock (1996) used variations of the Petersen estimators by grouping the sighting distances into categories. Borchers and Zucchini (1998, in press) proposed mark-recapture models for line transect data, and Borchers et al. (1998, in press) proposed Horvitz–Thompson type estimators. Recently, Chen (1999, in press) proposed an approach with the perpendicular distance as the only variable that affects detection.

This paper introduces a framework that generalizes the approach in Chen (1999, in press) to accommodate heterogeneity in detection caused by school size and other covariates. The proposed framework combines a conditional likelihood for the effective search widths of the two observers and conditional probability density estimation of the detection distances given the covariates. The estimation of the effective search widths can be carried out either parametrically or nonparametrically, depending on the way the conditional probability density estimation is formulated. The existing works in modeling heterogeneity in detection are based on parametric estimation of the detection function and require a flexible set of parametric forms for the detection function. The proposed framework has an advantage in that it allows both parametric and nonparametric estimations to be implemented.

This paper is structured as follows: Section 2 introduces the notations used and outlines the problem. A conditional estimator for the effective search width at each fixed level of the other covariates is given in Section 3. Section 4 is devoted to parametric and nonparametric estimations of a key quantity that measures the amount of local heterogeneity in detection and in the estimation of the effective search width. The theoretical properties of nonparametric abundance estimators are studied in Section 5. Section 6 deals with the choice of smoothing bandwidths. Section 7 analyzes the harbor porpoise data. General discussions are given in Section 8. This paper is a condensed version of Chen (1998) that contains more details of derivations and some simulation results, which can be downloaded from the author's homepage (http://www.latrobe.edu.au/www/statistic/sc.htm).

2. Assumptions and Outlines

There are two abundance measures for a clustered biological population. One, denoted as $D = N/A$, is the number of clus-

ters per unit area, where N is the total number of clusters and A is the survey area. The other is the number of individual animals per unit area and is denoted as $D_s = N\mu_s/A$, where μ_s is the mean cluster size.

Let x be the perpendicular distance of a cluster to the transect line and $\tilde{z} = (z_1, \ldots, z_p)$ be the other covariates, with z_1 reserved for the cluster size. If the cluster is detected, then $(x, \tilde{z})$ becomes one of the sample points $(X_i, \widetilde{Z}_i)$, where X_i is called the ith detection distance. As detection is made on both sides of the transect lines, signed perpendicular distances are used.

Let f_u be the underlying probability density function of $(x, \tilde{z})$, which is not directly observable, as the line transect sampling is biased. However, as the transect lines are allocated randomly in the survey region, it is reasonable to assume that

$$f_u(x, \tilde{z}) = (2w)^{-1} f_{u\tilde{z}}(\tilde{z}) I(-w \le x \le w), \tag{2.1}$$

where $f_{u\tilde{z}}$ is the marginal density of the covariates $\tilde{z}$, and w is the maximum perpendicular distance. This means that x is independent of $\tilde{z}$ and is uniformly distributed in $[-w, w]$.

For $i = 1$ and 2, let $g_i(x, \tilde{z})$ be the probability of detecting a cluster at a perpendicular distance x to the transect lines and having the covariates $\tilde{z}$, and let $\{(X_{ij}, \widetilde{Z}_{ij})\}_{j=1}^{n_i}$ be the sample detected by the ith observers, where n_i is the total number of sightings. Let f_1 and f_2 be the densities of the samples and $P_i = \int\int_{-w}^{w} g_i(x, \tilde{z}) f_u(x, \tilde{z})\, dx\, d\tilde{z}$ be the probability of detecting a cluster by the ith observer, then from the theory of weighted distribution (Patil and Rao, 1977), for $i = 1$ or 2,

$$f_i(x, \tilde{z}) = \frac{g_i(x, \tilde{z}) f_u(x, \tilde{z})}{P_i} = \frac{g_i(x, \tilde{z}) f_{u\tilde{z}}(\tilde{z})}{2P_i w}.$$

Define $D(\tilde{z}) = N(\tilde{z})/(2Lw)$ as the local abundance density at $\tilde{z}$, where $N(\tilde{z})$ is the total number of clusters having the covariates $\tilde{z}$. When $\tilde{z}_i$ is discrete and takes values within a discrete set $\mathcal{S}$, the total abundance measures are

$$D = \sum_{\tilde{z} \in \mathcal{S}} D(\tilde{z}) \quad \text{and} \quad D_s = \sum_{\tilde{z} \in \mathcal{S}} z_1 D(\tilde{z}).$$

When some covariates are continuous, the summation regarding those covariates should be replaced by integrals. We assume from now on that the covariates are all discrete as continuous covariates can be made discrete by binning.

The idea of this paper is to first estimate the local abundance $D(\tilde{z})$. Let $f_i(x \mid \tilde{z})$ be the conditional density of the sighting distance given $\tilde{z}$ and $\mu_i(\tilde{z}) = \int_{-w}^{w} g_i(x, \tilde{z})\, dx$ be the effective search width of the ith observer at $\tilde{z}$. Then, for $i = 1$ and 2,

$$f_i(x \mid \tilde{z}) = \frac{g_i(x, \tilde{z})}{\mu_i(\tilde{z})}. \tag{2.2}$$

The probability that at least one observer detects a cluster of $(x, \tilde{z})$ is $g(x, \tilde{z}) = g_1(x, \tilde{z}) + g_2(x, \tilde{z}) - g_1(x, \tilde{z}) g_2(x, \tilde{z})$, assuming conditional independence of detection between the two observers. The joint effective search width at $\tilde{z}$ is

$$\mu(\tilde{z}) = \int_{-w}^{w} g(x, \tilde{z})\, dx = \mu_1(\tilde{z}) + \mu_2(\tilde{z}) - \alpha(\tilde{z}) \mu_1(\tilde{z}) \mu_2(\tilde{z}),$$

where $\alpha(\tilde{z}) = \int_{-w}^{w} f_1(x \mid \tilde{z}) f_2(x \mid \tilde{z})\, dx$ and is also a key quantity of this paper.

Let the number of distinct sightings made by the two observers having the covariates $\tilde{z}$ be $n(\tilde{z})$. If the transect lines are allocated randomly such that $E\{n(\tilde{z})\} = N(\tilde{z}) P_0(\tilde{z})$, where $P_0(\tilde{z}) = \mu(\tilde{z})/(2W)$ is the probability of detecting a cluster of $\tilde{z}$, then from Burnham and Anderson (1976), an estimator for $D(\tilde{z})$ that does not require $g(0, \tilde{z}) = 1$ is

$$\widehat{D}_0(\tilde{z}) = \frac{n(\tilde{z})}{L\hat{\mu}(\tilde{z})}, \tag{2.3}$$

where $\hat{\mu}(\tilde{z})$ is an estimator for $\mu(\tilde{z})$. Then, the general estimators for D and D_s are

$$\widehat{D}_0 = L^{-1} \sum_{\tilde{z} \in \mathcal{S}} \frac{n(\tilde{z})}{\hat{\mu}(\tilde{z})} \quad \text{and} \quad \widehat{D}_1 = L^{-1} \sum_{\tilde{z} \in \mathcal{S}} \frac{z_1 n(\tilde{z})}{\hat{\mu}(\tilde{z})}. \tag{2.4}$$

3. A Conditional Likelihood Estimator for $\mu(\tilde{z})$

Let $\{X_{1i}(\tilde{z})\}_{i=1}^{n_1(\tilde{z})}$ and $\{X_{2i}(\tilde{z})\}_{i=1}^{n_2(\tilde{z})}$ be subsamples of the detection distances with covariates $\tilde{z}$, where $n_1(\tilde{z})$ and $n_2(\tilde{z})$ are the number of detections by the first and the second observers, respectively. Let $n_{11}(\tilde{z})$ be the number of common detections. Then, $n_{10}(\tilde{z}) = n_1(\tilde{z}) - n_{11}(\tilde{z})$ and $n_{01}(\tilde{z}) = n_2(\tilde{z}) - n_{11}(\tilde{z})$, respectively, are the number of detections by the first and the second observer only. Assume that detections are independent between the two observers and that $\{n_{10}(\tilde{z}), n_{01}(\tilde{z}), n_{11}(\tilde{z})\}$ given $n(\tilde{z})$ at fixed $\tilde{z}$ has a multinomial distribution $\text{mult}\{n(\tilde{z}), P_{10}(\tilde{z}), P_{01}(\tilde{z}), P_{11}(\tilde{z})\}$, where

$$\begin{aligned} P_{10}(\tilde{z}) &= \{\mu_1(\tilde{z}) - \alpha(\tilde{z}) \mu_1(\tilde{z}) \mu_2(\tilde{z})\}/\mu(\tilde{z}), \\ P_{01}(\tilde{z}) &= \{\mu_2(\tilde{z}) - \alpha(\tilde{z}) \mu_1(\tilde{z}) \mu_2(\tilde{z})\}/\mu(\tilde{z}), \\ P_{11}(\tilde{z}) &= \alpha(\tilde{z}) \mu_1(\tilde{z}) \mu_2(\tilde{z})/\mu(\tilde{z}). \end{aligned}$$

The conditional likelihood for $\mu_1(\tilde{z}), \mu_2(\tilde{z})$, and $\alpha(\tilde{z})$ is

$$\begin{aligned} &L\{\mu_1(\tilde{z}), \mu_2(\tilde{z}), \alpha(\tilde{z}) \mid n(\tilde{z})\} \\ &\quad = C P_{10}(\tilde{z})^{n_{10}(\tilde{z})} P_{01}(\tilde{z})^{n_{01}(\tilde{z})} P_{11}(\tilde{z})^{n_{11}(\tilde{z})}, \end{aligned}$$

where C is a constant free of $\mu_1(\tilde{z}), \mu_2(\tilde{z})$, and $\alpha(\tilde{z})$. Given $\alpha(\tilde{z})$, the values of $\mu_1(\tilde{z})$ and $\mu_2(\tilde{z})$ that maximize the foregoing conditional likelihood are, respectively,

$$\begin{aligned} \mu_1(\tilde{z}) &= n_{11}(\tilde{z})/\{\alpha(\tilde{z})\, n_2(\tilde{z})\}, \\ \mu_2(\tilde{z}) &= n_{11}(\tilde{z})/\{\alpha(\tilde{z}) n_1(\tilde{z})\}. \end{aligned}$$

Thus, the joint effective search width

$$\mu(\tilde{z}) = \{n(\tilde{z}) n_{11}(\tilde{z})\}/\{\alpha(\tilde{z}) n_1(\tilde{z}) n_2(\tilde{z})\}.$$

From (2.3), the estimator for $D(\tilde{z})$ is

$$\widehat{D}_0(\tilde{z}) = \frac{\hat{\alpha}(\tilde{z}) n_1(\tilde{z}) n_2(\tilde{z})}{L n_{11}(\tilde{z})},$$

where $\hat{\alpha}(\tilde{z})$ is an estimator of $\alpha(\tilde{z})$. Then, the overall abundance estimators in (2.4) become

$$\begin{aligned} \widehat{D}_0 &= L^{-1} \sum_{\tilde{z} \in \mathcal{S}} \frac{\hat{\alpha}(\tilde{z}) n_1(\tilde{z}) n_2(\tilde{z})}{n_{11}(\tilde{z})}, \\ \widehat{D}_1 &= L^{-1} \sum_{\tilde{z} \in \mathcal{S}} \frac{z_1 \hat{\alpha}(\tilde{z}) n_1(\tilde{z}) n_2(\tilde{z})}{n_{11}(\tilde{z})}. \end{aligned} \tag{3.1}$$

Note that $n_1(\tilde{z}) n_2(\tilde{z})/n_{11}(\tilde{z})$ is the Petersen estimator for $N(\tilde{z})$ by assuming constant detection. In a line transect survey, heterogeneity exists in detection with respect to x and $\tilde{z}$. At a fixed $\tilde{z}$, this heterogeneity is measured by $\alpha(\tilde{z})$. If

$n_{11}(\tilde{z})$ has low values at some $\tilde{z}$, the Chapman estimator $\{n_1(\tilde{z})+1\}\{n_2(\tilde{z})+1\}/\{n_{11}(\tilde{z})+1\}-1$ can be used to replace the Petersen estimator.

4. Estimation of $\alpha(\tilde{z})$

Note that $\alpha(\tilde{z}) = \int_{-w}^{w} f_1(x \mid \tilde{z}) f_2(x \mid \tilde{z})\, dx$ has alternative expressions

$$\alpha(\tilde{z}) = E[f_1\{X_2(\tilde{z}) \mid \tilde{z}\}] = E[f_2\{X_1(\tilde{z}) \mid \tilde{z}\}], \qquad (4.1)$$

where the expectations are with respect to $X_1(\tilde{z})$ or $X_2(\tilde{z})$. In this paper, we concentrate on deriving a nonparametric estimator for $\alpha(\tilde{z})$ based on the kernel density estimator. Kernel estimators have been used in conventional line transect surveys by Chen (1996a,b) and Mack and Quang (1998).

A kernel estimator for $f_i(x \mid \tilde{z})$ that is obtained using the sighting distances $\{X_{ij}(\tilde{z})\}_{j=1}^{n_i(\tilde{z})}$ is

$$\hat{f}_i(x \mid \tilde{z}) = \{n_i(\tilde{z})h(\tilde{z})\}^{-1} \sum_{j=1}^{n_i(\tilde{z})} K\left(\frac{x - X_{ij}(\tilde{z})}{h(\tilde{z})}\right), \qquad (4.2)$$

where K is a kernel function, and $h(\tilde{z})$ is the smoothing bandwidth. A list of commonly used kernels is available in Scott (1992) and Silverman (1996). Section 6 discusses the way in which $h(\tilde{z})$ is chosen.

Now the following method of moments estimator for $\alpha(\tilde{z})$ can be proposed from (4.1):

$$\hat{\alpha}_1(\tilde{z}) = \{n_2(\tilde{z})\}^{-1} \sum_{j=1}^{n_2(\tilde{z})} \hat{f}_1\{X_{2j}(\tilde{z}) \mid \tilde{z}\},$$

or

$$\hat{\alpha}_2(\tilde{z}) = \{n_1(\tilde{z})\}^{-1} \sum_{j=1}^{n_1(\tilde{z})} \hat{f}_2\{X_{1j}(\tilde{z}) \mid \tilde{z}\},$$

where $\hat{f}_1(\cdot \mid \tilde{z})$ and $\hat{f}_2(\cdot \mid \tilde{z})$ are given in (4.2). As K is symmetric, it can be shown from (4.2) that $\hat{\alpha}_1(\tilde{z}) = \hat{\alpha}_2(\tilde{z}) = \hat{\alpha}(\tilde{z})$, and that

$$\hat{\alpha}(\tilde{z}) = \{n_1(\tilde{z})n_2(\tilde{z})h(\tilde{z})\}^{-1} \times \sum_{i=1}^{n_1(\tilde{z})} \sum_{j=1}^{n_2(\tilde{z})} K\left(\frac{X_{2j}(\tilde{z}) - X_{1i}(\tilde{z})}{h(\tilde{z})}\right). \qquad (4.3)$$

In contrast to the preceding local kernel estimator, a global multivariate product kernel estimator for $f_i(x \mid \tilde{z})$ is

$$\hat{f}_i(x \mid \tilde{z}) = (n_1 h_i)^{-1} \sum_{j=1}^{n_i} K\left(\frac{x - X_{ij}}{h_i}\right) W_{ij}(\tilde{z}),$$

where $W_{ij}(\tilde{z})$ is the weight contributed by $(X_{ij}, \widetilde{Z}_{ij})$ and is defined as

$$W_{ij}(\tilde{z}) = \frac{K\left(\frac{z_1 - Z_{ij1}}{b_{i1}}\right) \cdots K\left(\frac{z_p - Z_{ijp}}{b_{ip}}\right)}{\sum_{j=1}^{n_i} K\left(\frac{z_1 - Z_{ij1}}{b_{i1}}\right) \cdots K\left(\frac{z_p - Z_{ijp}}{b_{ip}}\right)}.$$

In the foregoing expression, $\widetilde{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijp})$, and the same kernel K is used for all the variables. However, each variable has a different smoothing bandwidth b_{il}, $l = 1, \ldots, p$. Scott (1992) gives the available details on multivariate density estimation. Hence, a global estimator for $\alpha(\tilde{z})$ is

$$\begin{aligned}\hat{\alpha}(\tilde{z}) &= \int_{-w}^{w} \hat{f}_1(x \mid \tilde{z}) \hat{f}_2(x \mid \tilde{z})\, dx \\ &= (h_1 h_2)^{-1} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} W_{1j}(\tilde{z}) W_{2k}(\tilde{z}) \\ &\quad \times \int_{-w}^{w} K\left(\frac{x - X_{1j}}{h_1}\right) K\left(\frac{x - X_{2j}}{h_2}\right) dx. \qquad (4.4)\end{aligned}$$

Compared with the global estimator, the local estimator has two attractive features: One is its easy computation and the other is that the kernel smoothing is univariate and to some extent reduces the influence of the curse of dimensionality that is experienced in the multivariate density estimation. In multivariate density estimation, the sample size that is required to achieve a fixed level of accuracy for the kernel estimator increases exponentially with increasing dimension as discussed in Silverman (1986, p. 94).

5. Properties $\widehat{D}_0$ and $\widehat{D}_1$

The performance of $\widehat{D}_0$ and $\widehat{D}_1$ obtained by using the local kernel estimator for $\alpha(\tilde{z})$ given in (4.3) is evaluated in this section. We assume that:

(1) K is a symmetric compact kernel with a support $[-1, 1]$.
(2) $h(\tilde{z}) \to 0$ and $n(\tilde{z}) \to \infty$ as $L \to \infty$ at any $\tilde{z}$.
(3) For any $\tilde{z}$ and $i = 1, 2$, $\int_{-w}^{w} |f_i(x \mid \tilde{z})|\, dx < \infty$ and $f_i(x \mid \tilde{z})$ has a bounded third derivative with respect to x.

We see from (4.3) that $\hat{\alpha}(\tilde{z})$ is a two-sample U-statistic (Serfling, 1980). Based on a result reported in Lehmann (1951), $\hat{\alpha}(\tilde{z}) \to N[E\{\hat{\alpha}(\tilde{z})\}, \mathrm{var}\{\hat{\alpha}(\tilde{z})\}]$ in distribution as $L \to \infty$. Derivations given in Chen (1998) show that

$$\widehat{D}_0(\tilde{z}) \to N\big[D(\tilde{z}), \mathrm{var}\{\widehat{D}_0(\tilde{z})\}\big] \qquad (5.1)$$

in distribution as $L \to \infty$. It is obvious that $\{\widehat{D}_0(\tilde{z})\}_{\tilde{z} \in \mathcal{S}}$ are mutually independent as the cell subsamples are mutually independent. Thus, $\widehat{D}_0 = L^{-1} \Sigma_{\tilde{z} \in \mathcal{S}} \widehat{D}_0(\tilde{z})$ and $\widehat{D}_s = L^{-1} \Sigma_{\tilde{z} \in \mathcal{S}} z_1 \widehat{D}_0(\tilde{z})$ are the sums of independent random variables and therefore are asymptotically normal distributed from (5.1), and

$$\mathrm{var}\{\widehat{D}_0\} = L^{-2} \sum_{\tilde{z} \in \mathcal{S}} \mathrm{var}\big\{\widehat{D}(\tilde{z})\big\},$$

$$\mathrm{var}\{\widehat{D}_{l1}\} = L^{-2} \sum_{\tilde{z} \in \mathcal{S}} z_1^2 \mathrm{var}\big\{\widehat{D}(\tilde{z})\big\}.$$

It is shown in Chen (1998) that an estimator for $\mathrm{var}\{\widehat{D}(\tilde{z})\}$ is

$$\begin{aligned}\widehat{\mathrm{var}}\big\{\widehat{D}(\tilde{z})\big\} &= \widehat{D}^2(\tilde{z})\big[R(K)/\{\hat{\alpha}(\tilde{z})hn_1(\tilde{z})n_2(\tilde{z})\} \\ &\quad + n^{-1}(\tilde{z})\{1 - \hat{\mu}(\tilde{z})/(2w)\} + \{n_1(\tilde{z})n_2(\tilde{z})\}^{-1} \\ &\quad \times \{\gamma(\tilde{z}) - n(\tilde{z}) - n_{01}(\tilde{z})n_{10}(\tilde{z})/n(\tilde{z})\}\big], \qquad (5.2)\end{aligned}$$

where $R(K) = \int K^2(t)\, dt$. If the number of sightings on duplicate transect lines are available, then the binomial assumption

for $n(\tilde{z})$ can be dropped and the term $n^{-1}(\tilde{z})\{1-\hat{\mu}(\tilde{z})/(2w)\}$ in (5.2) can be replaced by $\widehat{cv}^2\{n(\tilde{z})\}$.

Approximate $100(1-2\alpha)\%$ confidence intervals for D and D_s are

$$\hat{D} \pm z_\alpha \sqrt{\widehat{\text{var}}(\hat{D})} \quad \text{and} \quad \hat{D}_s \pm z_\alpha \sqrt{\widehat{\text{var}}(\hat{D}_s)}, \tag{5.3}$$

where z_α is the upper α-percentile of the standard normal distribution. The bootstrap can be used to produce the percentile points to replace z_α used in (5.3) as well as in the method of log transformation outlined in Buckland et al. (1993).

6. Choosing the Smoothing Bandwidths

As only $\hat{\alpha}(\tilde{z})$ is involved with the smoothing, we choose the bandwidth to minimize the mean square error (MSE) of $\hat{\alpha}(\tilde{z})$. From the results of Section 5,

$$\text{MSE}\{\hat{\alpha}(\tilde{z})\} = \frac{1}{4}h^4(\tilde{z})\sigma_k^4\left\{\int_{-w}^{w} f_1''(x \mid \tilde{z}) f_2(x \mid \tilde{z})\, dx\right\}^2 + \alpha(\tilde{z})R(K)h^{-1}(\tilde{z})E\{n_1(\tilde{z})n_2(\tilde{z})\}^{-1},$$

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

The optimal bandwidth that minimizes the preceding MSE is

$$h^*(\tilde{z}) = \left[\alpha(\tilde{z})R(K)\sigma_k^{-4}\left\{\int_{-w}^{w} f_1''(x \mid \tilde{z}) f_2(x \mid \tilde{z})\, dx\right\}^{-2}\right]^{-1/5} \times \left[E\{n_1(\tilde{z})n_2(\tilde{z})\}^{-1}\right]^{1/5}. \tag{6.1}$$

The optimal MSE is $O\{N^{-8/5}(\tilde{z})\}$, which is much smaller than $O\{N^{-4/5}(\tilde{z})\}$, the optimal MSE achieved by the kernel density estimators. Thus, estimation of $\alpha(\tilde{z})$ is much easier and is less sensitive to the smoothing bandwidth.

To obtain a practical bandwidth, the method of reference to a standard distribution method proposed by Silverman (1986) is used to produce values of the unknown quantities α and η. Assuming $f_i(x \mid \tilde{z}) \sim N\{0, \sigma_i^2(\tilde{z})\}$, a practical bandwidth is

$$h(\tilde{z}) = \left\{R(K)\sigma_k^{-4}\sqrt{2\pi}\right\}^{1/5} \times \sqrt{A_1^2(\tilde{z}) + A_2^2(\tilde{z})}\{n_1(\tilde{z})n_2(\tilde{z})\}^{-1/5}, \tag{6.2}$$

where $A_i(\tilde{z}) = \min\{s_i(\tilde{z}), q_i(\tilde{z})/1.349\}$ are robust estimates of $\sigma_i^2(\tilde{z})$. Here $s_i(\tilde{z})$ and $q_i(\tilde{z})$ are the sample standard deviation and inter quantile range, respectively.

7. An Example

In this section, the harbor porpoise data from the "Small Cetacean Abundance in the North Sea" survey (Hammond et al., 1995) are analyzed by applying the method developed in this paper. The survey was conducted via vessels, with observers who made independent detections on board, in the summer of 1994. The reason for conducting the independent observers survey was because the detection near the transect line was suspected to be much less than 1. This was indeed confirmed by the data. Figure 1 presents the ratios of the number of common detections made by both observers to the number of detections made by the second observer over some categories of the detection distance for $z_1 = 1, 2, 3$, and 4. The ratios are rough estimates for $g_1(x, z_1)$ at each level of z_1, as the detections made by the second observer are used to validate the detection of the first. Figure 1 shows that the detection probability near $x = 0$ was much lower than 1, at all sizes presented, confirming that the fundamental assumption of the conventional line transect survey was violated. The fact that this also shows that a larger cluster size tended to have a larger probability of detection and an increased probability of detection at a distance of 0.4–0.8 km as the size increased from 1 to 3, was quite remarkable. These indicate the presence of size bias. The drop in detection probability at size 4 was because of the fact that there were only five sightings.

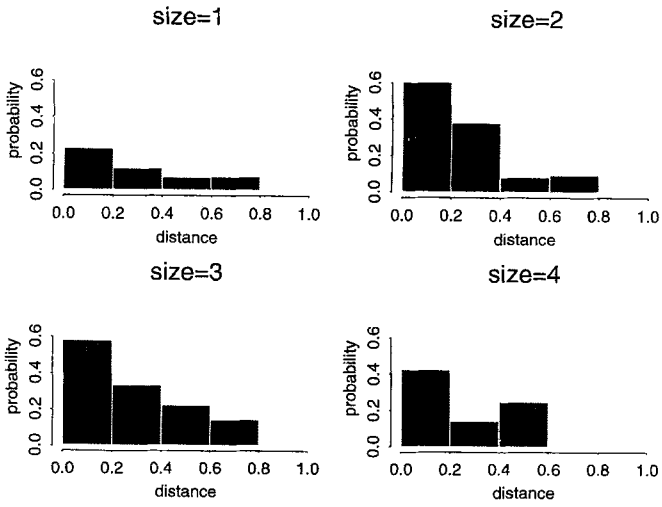

Figure 1. Rough estimates of $g_1(x, z_1)$ at a group size $z_1 = 1, 2, 3$, and 4.

There were 1157 clusters of porpoises detected in the survey. The survey effort L was 12,155 km, and the total survey region was 889,486 km^2. To remove the effect of animal movement in response to the survey vessels, the second observer scanned far ahead of the vessel using binoculars, whereas the first observer made detections via the naked eye. Here, the detection distances from the second observer are used. A comprehensive analysis of the data using a logistic regression–based Horvitz–Thompson estimator is given in Borchers et al. (1998).

Only the cluster size z_1 was considered as the covariates that ranged from one to seven. The local kernel estimator for $\alpha(z_1)$ given in (4.3) is used in conjunction with the biweight kernel $K(t) = (15/16)(1-t^2)^2 I(-1 \le t \le 1)$. Although there are enough observations for the nonparametric approach at sizes 1 to 5, there were only two sightings of size 7, and 6 sightings of size 6. The samples of sizes 6 and 7 were combined, and 6.25, which is the average of 6 and 7 weighted by the sample sizes, was used as the size for the combined sample. Then, (6.2) produced the bandwidths $h(1) = 0.067$, $h(2) = 0.0993$, $h(3) = 0.160$, $h(4) = 0.196$, $h(5) = 0.321$, and $h(6.25) = 0.3795$. The bandwidth increased as the size increased, indicating that the detection distances became more sparse as the size increased. Figure 2 shows the histograms of the detection distances of the second observer at various cluster sizes and the kernel estimates for the conditional density $f_2(x \mid z_1)$ using the previously mentioned bandwidths. As the bandwidths were designed to estimate $\alpha(z_1)$, not the

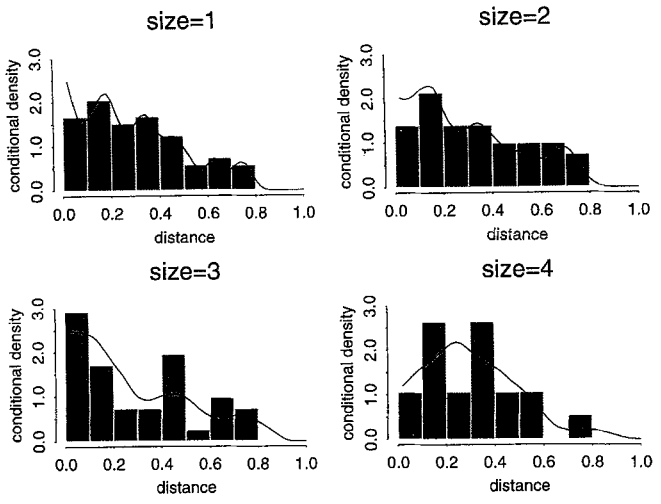

Figure 2. Kernel estimates of the conditional densities $f_2(x \mid z_1)$ at a group size $z_1 = 1, 2, 3$, and 4.

conditional densities, they were smaller than those appropriate for the conditional densities. This is the reason why the conditional density estimates were not smoothed.

The results of the analysis are given in Table 1. For each cluster size, we give the number of sightings and $\gamma(z_1)$, which is the Petersen estimator for $N(z_1)$. Estimates for the heterogeneity factors $\alpha(z_1)$ are given in Table 1. When heterogeneity is taken into account, we have the corrected estimates of $N(z_1)$ given by $\gamma(z_1)\alpha(z_1)w$ with $w = 0.85$ km. Table 1 also presents the estimates of $D_0(z_1)$, $D_s(z_1)$, and the estimates of D, D_s, and N together with their standard errors.

The Petersen estimate for the total number of clusters N was 3441. This estimate was too small as it ignored heterogeneity in detection. The amount of heterogeneity in detection owing to the sighting distance as measured by $\alpha(z_1)$ remained in the range between 1.37 and 1.9 and decreased as the cluster size was increased, indicating that the heterogeneity is larger at a smaller size. After correcting for the heterogeneity, the estimated total number of animal clusters was 6227 with a standard error of 849. All the estimates for D and D_s had coefficients of variation that were less than 14%. From the estimates of D and D_s, we derive an estimate for the mean cluster size of 1.312 with a standard error of 0.23. Multiplying the estimates $\widehat{D}_0$ and $\widehat{D}_1$ by the ratio of the entire survey area to the covered area, the estimates of the total number of animal clusters and animals were, respectively, 268,062 with a standard error of 36,554, and 351,575 with a standard error of 39,179. These estimates can be compared with those reported in Borchers et al. (1998, in press), which were 210,731 with a standard error of 31,609 and 313,375 with a standard error of 47,006, respectively. The difference between the two sets of estimates may be the result of (1) different estimation methods used and (2) the current estimates being produced by pooling observations of all vessels. However, the two sets of figures are not significantly different as the estimates from one set are contained in the 95% confidence intervals of the other.

8. Discussion

The proposed framework is based on two key assumptions: (1) independence in detection between the two observers and (2) the local multinomial distribution of $\{n_{10}(\tilde{z}), n_{01}(\tilde{z}), n_{11}(\tilde{z})\}$ given $n(\tilde{z})$. The local multinomial distribution assumption is weaker than the global multinomial distribution of $\{n_{10}, n_{01}, n_{11}\}$ given n, as the latter requires a certain homogeneity across the covariate space. The proposed framework can accommodate correlation in detection between the two observers. This point may be appreciated by noting that

$$\alpha(\tilde{z}) = (2w)\{\mu_1(\tilde{z})\mu_1(\tilde{z})\}^{-1} \operatorname{cov}\{g_1(X, \tilde{z})g_2(X, \tilde{z})\}.$$

It is shown in Section 7 that the optimal MSE of the kernel estimators for $\alpha(\tilde{z})$ is $O\{N^{-8/5}(\tilde{z})\}$, which is much smaller than $O\{N^{-4/5}(\tilde{z})\}$ obtained by the density estimators. This means that the sample size required to estimate $\alpha(\tilde{z})$ should be much smaller than that required for estimating density functions. This is further enhanced by the fact that the local kernel estimator is only univariate and is applied only on the

Table 1
Analysis of the harbor porpoise data. Figures inside the parentheses are standard errors of the estimates given before.

z_1	$n_1(z_1)$	$n_2(z_1)$	$n_{11}(z_1)$	$\gamma(z_1)$	$\alpha(z_1)$	$\widehat{N}(z_1)$	$\widehat{D}_0(z_1)$	$z_1\widehat{D}_0(z_1)$
1	419	362	46	3297.3	1.752	4911.1 (0.0313)	0.2377	0.2377
2	154	164	43	587.3	1.742	869.7 (0.0050)	0.0421	0.0842
3	55	56	16	192.5	1.891	309.4 (0.0029)	0.0150	0.0449
4	16	23	5	73.6	1.746	109.2 (0.0019)	0.0053	0.0211
5	7	7	4	12.3	1.682	17.51 (0.0003)	0.0008	0.0042
6.25	5	7	4	8.75	1.377	10.23 (0.0002)	0.0005	0.0031
Total	656	619	118	3441.2	(849.2)	6227.3 (0.041)	0.3013 (0.044)	0.3953

sighting distances. However, this will give rise to problems if $n(\tilde{z})$ is small at some $\tilde{z}$. One remedy is to combine those $\tilde{z}$ bins to increase the sample size. Another is to try parametric estimation of $\alpha(\tilde{z})$.

Parametric estimators for $\alpha(\tilde{z})$ can be developed if the parametric form for the detection functions g_1 and g_2 and the underlying distribution f_u are available. These derive the likelihood function of some parameter, for instance θ, given the sighting samples, similar to the full likelihood approach of Borchers and Zucchini (1998, in press). Suppose that $\hat{f}_i(x \mid \tilde{z}, \hat{\theta})$ for $i = 1$ and 2 are parametric conditional density estimates, for then a parametric estimate for $\alpha(\tilde{z})$ is $\hat{\alpha}(\tilde{z}) = \int_{-w}^{w} \hat{f}_1(x \mid \tilde{z}, \hat{\theta}) \hat{f}_2(x \mid \tilde{z}, \hat{\theta})\, dx$. This parametric approach shares the same strength and limitations with the existing parametric approaches (Borchers and Zucchini, 1998, in press; Borchers et al., 1998, in press). Its strength is that it fully uses the information and produces more efficient estimators. The limitations are as follows: (1) a flexible class of parametric models is not available as yet, and (2) maximizing the likelihood may be quite nontrivial.

The proposed nonparametric estimator can be easily computed according to (4.3) after choosing the smoothing bandwidth $h(\tilde{z})$ and a kernel function. Computer software written in C++ is available from the author.

Acknowledgements

The author is grateful to Dr David Borchers for allowing him to use the harbor porpoise data and thanks the two referees and the associate editor for constructive comments and suggestions.

Résumé

Cet article présente un cadre permettant l'estimation de l'abondance d'animaux à partir d'enquêtes réalisées par des observateurs indépendants avec la méthode des transects-lignes pour populations grégaires. Le cadre proposé généralise une approche développée par Chen (1999) pour tenir compte de l'hétérogénéité de détection liée à la taille des groupes et à d'autres covariables. Des estimateurs paramétriques et non paramétriques des largeurs locales efficaces de recherche peuvent être dérivés à partir de ce cadre. Un estimateur non paramétrique basé sur une estimation par noyau de la densité conditionnelle est proposé pour sa flexibilité dans la modélisation des fonctions de détection et étudié. Un jeu de données réelles sur des marsouins de la Mer du Nord est analysé.

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,** 311–327.

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

Borchers, D. L., Buckland, S. T., Goedhard, P. W., Clark, E. D., and Hedley, S. L. (1998). Horvitz–Thompson estimators for double-platform line transect surveys. *Biometrics* **54,** 1221-1237.

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

Burnham, K. P. and Anderson, D. R. (1976). Mathematical models for nonparametric inferences from line transect data. *Biometrics* **32,** 325–336.

Butterworth, D. S. and Borchers, D. L. (1988). Estimation of $g(0)$ for Minke clusters 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,** 301–313.

Chen, S. X. (1996a). Kernel estimates for density of a biological population using line transect sampling. *Journal of Royal Statistical Society, Series C* **45,** 135–150.

Chen, S. X. (1996b). Studying cluster size effects in line transect sampling using the kernel method. *Biometrics* **52,** 1283–1294.

Chen, S. X. (1998). *Estimation in Independent Observer Line Transect Surveys for Clustered Populations.* Research Report 8, Department of Statistics, La Trobe University.

Chen, S. X. (1999). Animal abundance estimation in independent observer line transect surveys. *Environmental and Ecological Statistics* **6,** in press.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Annals of Mathematics Statistics* **22,** 165–179.

Mack, Y. P. and Quang, P. X. (1998). Kernel methods in line and point transects samplings. *Biometrics* **54,** 606–619.

Patil, G. P. and Rao, C. R. (1977). The weighted distribution: A survey of their applications. In *Application of Statistics,* P. R. Krishnaiah (ed). Amsterdam: North-Holland.

Schweder, T. (1990). Independent observer experiments to estimate the detection function in line transect surveys of whales. *Report of the International Whaling Commission* **40,** 349–355.

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

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

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

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

Received April 1998. *Revised February* 1999.
Accepted February 1999.