BIOMETRICS 58, 263–269
June 2002

Sequential Estimation in Line Transect Surveys

Song Xi Chen,[1,*] Paul S. F. Yip,[2,] and Yong Zhou[3,***]**

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore
[2]Department of Statistics, University of Hong Kong, Hong Kong, China
[3]University of Hong Kong, Hong Kong, China and Chinese Academy of Science, Beijing, China
*email: stacsx@nus.edu.sg
**email: sfpyip@hku.hk
***email: yzhou@amath4.amt.ac.cn

SUMMARY. This article considers using sequential procedures to determine the amount of survey effort required in a line transect survey in order to achieve a certain precision level in estimating the abundance of a biological population. Sequential procedures are constructed for both parametric and nonparametric animal abundance estimators. The criterion used to derive the stopping rules is the width of confidence intervals for the animal abundance. For each estimator considered, we develop stopping rules based on the asymptotic distributions and the bootstrap. A sequential analysis on an aerial survey of the southern bluefin tuna indicates substantial saving of survey effort can be made by employment of the proposed sequential procedures. This savings of survey effort is also observed in a simulation study designed to evaluate the empirical performance of the proposed sequential procedures.

KEY WORDS: Bootstrap; Confidence interval; Kernel estimator; Parametric estimator; Stopping rule.

1. Introduction

Aerial line transect surveys for the commercially valuable southern bluefin tuna have been conducted over the Great Australia Bight as part of a recruitment monitoring program for juvenile tuna within a trilateral framework among Australia, Japan, and New Zealand. The aerial surveys have been conducted in the summer months since 1991 to provide prompt and fishery-independent information on the level of tuna abundance; see Chen (1996a,b) for more information on the surveys as well as their statistical analyses. Line transect surveys are a popular sampling method for estimating the abundance of a biological population. Comprehensive reviews on these surveys are available in Seber (1982) and Buckland et al. (1993). In a line transect survey, randomly allocated transect lines of length L are placed within a survey region. Properly trained observers are employed to traverse the transect lines and make detections of the target population. All the detections made while traveling along the lines are counted and the perpendicular distances of detected objects to the transect line are measured, which carry vital information on the detection pattern and the distribution of the population relative to the transect line. Seber (1982) and Buckland et al. (1993) provide details on how to formulate an abundance estimate based on these detected distances.

The focus of this article is on determination of L, the total length of transect lines. In a typical line transect study, L has been fixed and the focus of the study has been on estimation of the animal abundance $D = N/A$, where N is the unknown size of a population in a region with area A. Obviously, L should be long enough to allow a sufficient amount of information to be gathered to make an accurate estimation of D. At the same time, we do not want to waste the survey effort because some surveys are quite expensive to conduct. For the tuna aerial survey, the leasing and running costs of the airplane together with maintaining a survey crew on call every day (waiting for good weather) during the survey period is very high. There are discussions in Seber (1982) and Buckland et al. (1993) about how to choose L. A pilot survey is suggested to create a pilot estimator for D. The length L is chosen to make the coefficient of variation smaller than a predetermined level.

This article considers sequential procedures to determine the amount of survey effort L without conducting a pilot survey. The sequential procedures are formulated for both parametric and nonparametric animal abundance estimators and are based on the length of confidence intervals for D. For each type of estimator, large-sample stopping rules based on an asymptotic distribution and finite-sample stopping rules based on the bootstrap are considered. It is revealed that, for the parametric estimator, the minimum survey effort depends on the shape and the scale parameters of the detection function. For the nonparametric estimator, the dependence of the stopping rule on the underlying detection mechanism is largely reduced, implying the robustness of the nonparametric rule. A simulation presented in Section 7 shows that the parametric sequential rule needs less survey effort than the nonparametric rule, provided corrected parametric models are used, and the finite-sample sequential rules based on the bootstrap reduce the survey effort of the asymptotic rules.

The article is structured as follows. In Section 2, we review parametric and nonparametric estimators for D in line transect surveys. A general sequential procedure based on the length of confidence intervals is introduced in Section 3. Large-sample stopping rules based on the asymptotic distributions of the estimators are developed in Section 4, whereas the bootstrap-based finite-sample stopping rules are considered in Section 5. A sequential analysis of a data set from the tuna aerial survey is given in Section 6. Results from a simulation study are reported in Section 7.

2. Line Transect Estimation

Suppose there are n_L detections made after traveling the transect lines of length L and $X_1, X_2, \cdots, X_{n_L}$ are the perpendicular sighting distances with f as their common probability density function. Here we write n_L to emphasis that the number of sightings depends on L. Let $g(x)$, commonly called the detection function, be the conditional probability function of detecting an object given that the object is at a perpendicular distance x from the transect line and let w be the maximum perpendicular sighting distance. Only a proportion, say P_0, of the objects is detectable. Assume that f is uniform in $[-w, w]$. Standard line transect survey theory shows that $P_0 = w^{-1} \int_0^w g(x)dx$ and $D = \mathrm{E}(n)/\{2L \int_0^w g(x)dx\}$. Then the usual assumption $g(0) = 1$ means that $\int_0^w g(x)dx = \{f(0)\}^{-1}$ and $D = \mathrm{E}(n)f(0)/\{2L\}$, which leads to a general estimator for D,

$$\hat{D} = \frac{n\hat{f}(0)}{2L}, \qquad (1)$$

where $\hat{f}(0)$ is an estimator for $f(0)$.

Parametric and nonparametric estimators for $f(0)$ have been developed in line transect surveys. The parametric estimation is made within the general exponential power series model established by Pollock (1978), which assumes

$$f(x) = f(0)\exp\{-(x/\lambda)^p\}, \qquad \lambda > 0 \; p > 0, \qquad (2)$$

where $f(0) = \{\lambda\Gamma(1 + 1/p)\}^{-1}$ and λ and p are the scale and the shape parameters, respectively. Maximum likelihood estimators for the parameters and hence an estimator for $f(0)$ by substitution can be found in Pollock (1978).

There are basically two nonparametric estimators. One is based on the Fourier series expansion with a key function as proposed in Buckland et al. (1993). The other is the kernel smoothing estimator proposed by Chen (1996a) and Mack and Quang (1998). In this article, we consider the kernel estimator when developing nonparametric sequential rules.

3. A Framework for Sequential Analysis

In this section, we propose sequential procedures to determine L without conducting a pilot survey. The sequential procedures reflect a predetermined level of precision required in estimating D and produces stopping rules. The stopping rules are evaluated constantly during the survey, and once satisfied, the survey is stopped. The criterion used for developing the procedures is the length of a confidence interval for D with certain confidence level, say 95% . This criterion has been used in Carroll (1976) for probability density estimation.

For a line transect estimator $\hat{D}$ given in (1), either parametric or nonparametric, we assume

(a) $\mathrm{bias}^2(\hat{D}) = o\{\mathrm{var}(\hat{D})\}$;
(b) the estimator $\hat{D}$ has an asymptotic normal distribution as $L \to \infty$;
(c) $f'(0) = 0$ and $f''(x)$ exists for any x;
(d) $g(0) = 1$;
(e) $\mathrm{E}(n) = NP_0$ and $\mathrm{var}(n) = \gamma\mathrm{E}(n)$, where γ is a dispersion parameter.

Condition (a) means that the squared bias is of a smaller order than the variance and limits the effect of the bias on the coverage of confidence intervals. The above condition is satisfied by the parametric estimator under the general exponential power series model. For the nonparametric kernel estimators, the condition may be satisfied by a bias correction given in Chen (1996a). Conditions (b)–(e) are standard conditions in line transect surveys. When the distribution of the biological population is random, n should be Poisson distributed and $\gamma = 1$. There is a tendency of clustering in biological populations, which makes $\gamma > 1$. In the latter case, γ can be estimated by independent replication of transect lines and the number of sightings made on each replicate (cf., Buckland et al., 1993).

We write $\hat{D}_L$, $\hat{f}_L(0)$, and $\mathrm{var}(\hat{D}_L)$ to emphasize the role of L on these quantities. A equal-tailed confidence interval for D derived from $\hat{D}_L$ with confidence level $1 - \alpha$ is

$$\left(\hat{D}_L + \beta_{\alpha/2}\widehat{\mathrm{SD}}(\hat{D}_L), \hat{D}_L + \beta_{1-\alpha/2}\widehat{\mathrm{SD}}(\hat{D}_L)\right),$$

where $\widehat{\mathrm{SD}}(\hat{D}_L)$ is an estimator for the standard deviation of $\hat{D}$ and $\beta_{1-\alpha/2}$ and $\beta_{\alpha/2}$ are the $(1 - \alpha/2)$th and the $\alpha/2$th percentiles of the Studentized estimator

$$t_L = \{\hat{D}_L - D\}/\sqrt{\widehat{\mathrm{var}}(\hat{D}_L)},$$

respectively. The standard deviation $\widehat{\mathrm{SD}}(\hat{D}_L)$ can be constructed based on the following general formula for the variance of $\hat{D}_L$, which is valid regardless of which estimator is used for $f(0)$:

$$\begin{aligned}\mathrm{var}\{\hat{D}_L\} &= (2L)^{-2}\left[\mathrm{E}\left[n_L^2\mathrm{var}\{\hat{f}_L(0)|n_L\}\right] + \mathrm{var}\left[n_L\mathrm{E}\{\hat{f}_L(0)|n_L\}\right]\right] \\ &= (2L)^{-2}\left[\mathrm{E}\left[n_L^2\mathrm{var}\{\hat{f}_L(0)|n_L\}\right] + \gamma f^2(0)\mathrm{E}(n_L)\right] \\ &\quad \times \{1 + o(1)\}. \end{aligned} \qquad (3)$$

The sequential procedure considered in this article is to find the smallest L_1 such that the length of the above confidence interval is

$$(\beta_{1-\alpha/2} - \beta_{\alpha/2})\widehat{\mathrm{SD}}(\hat{D}_{L_1}) \le d_0, \qquad (4)$$

where d_0 is a desired length of the confidence interval representing a level of precision in the estimation of D. The sequential rule is to stop the survey at $L = L_1$. Although the above rule gives an impression of instant stoppage, the termination of a survey can be facilitated after one transect line or even after one survey day to allow time to test whether (4) has been met or not. The rule can be modified to stop if $L \ge \max\{L_0, L_1\}$, where L_0 represent a predetermined minimum survey effort, in order to gain minimum sample coverage of the study area. This modification limits the chance of the survey being terminated prematurely, although our simulation reported in Section 7 indicates (4) performs generally well.

4. Asymptotic Stopping Rules

The asymptotic stopping rules for the parametric and the kernel estimators are based on the asymptotic normality of $\hat{D}_L$ as $L \to \infty$. Here the standard normal percentile Z_r is used to approximate β_r, the exact percentile of $\hat{D}_L$ used in (4).

4.1 *Stopping Rule for the Parametric Estimator*

The parametric estimator assumes f has the parametric form given in (2). From Pollock (1978), $(\hat{\lambda}_L, \hat{p}_L)$, the maximum likelihood estimators of λ and p, are the roots of the following equations:

$$\lambda = \left\{\left(p\sum X_i^p/n\right)\right\}^{1/p}$$

$$n\phi(1+1/p)/p^2 = \sum (X_i/\lambda)^p/\log(X_i/\lambda),$$

where $\phi(\cdot)$ is the digamma function. The parametric estimator for D is

$$\hat{D}_{pL} = \frac{n_L \hat{f}_{pL}(0)}{2L},$$

where $\hat{f}_{pL}(0) = \{\hat{\lambda}_L \Gamma(1+1/\hat{p}_L)\}^{-1}$. Denote $\text{var}\{\hat{f}_{pL}(0)\} = \sigma^2(\lambda, p)n^{-1}$, where

$$n^{-1}\sigma^2(\lambda,p) = \{f(0)/\lambda\}^2\text{var}(\hat{\lambda}) + \{f(0)\phi(1+1/p)\}^2\text{var}(\hat{p}) + 2\{f^2(0)\phi(1+1/p)/\lambda\}\text{cov}(\hat{\lambda},\hat{p});$$

see Pollock (1978) for the asymptotic covariance of $(\hat{\lambda}, \hat{p})$. From (3),

$$\text{var}(\hat{D}_{pL}) = (2L)^{-1}D\{\sigma^2(\lambda,p)/f(0) + \gamma f(0)\}. \quad (5)$$

Then

$$\widehat{\text{SD}}(\hat{D}_{pL}) = \sqrt{(2L)^{-1}\hat{D}_{pL}\{\sigma^2(\hat{\lambda},\hat{p})/\hat{f}_{pL}(0) + \hat{\gamma}\hat{f}_{pL}(0)\}}, \quad (6)$$

where $\hat{\gamma}$ is an estimate for the dispersion parameter, the estimation of which has been discussed in Section 3.

The asymptotic stopping rule for the parametric estimator is to stop the survey if

$$2Z_{1-\alpha/2}\widehat{\text{SD}}(\hat{D}_{pL}) \le d_0.$$

The relationship between L and the detection parameters can be revealed from (5); i.e., for a fixed abundance density D and L, $\text{var}(\hat{D}_{pL})$ depends on

$$\eta(\lambda,p,\gamma) = \sigma^2(\lambda,p)/f(0) + \gamma f(0) = \sigma^2(\lambda,p)\lambda\Gamma(1+1/p) + \gamma/\{\lambda\Gamma(1+1/p)\}.$$

Clearly, the larger the $\eta(\lambda,p,\gamma)$, the larger the variance, the longer distance we need to travel.

We plot the η function in Figure 1, with panels (a) and (b) for the surface and the contour plots of $\eta(\lambda,p,1)$, respectively, and panels (c) and (d) for those of $\eta(\lambda,p,5)$. The ranges for p and λ are $[0,4]$ and $[1,4]$, respectively. It shows that the increase of γ only slightly increases the height of the surface without affecting its shape much. The η function is unimodal with respect to the shape parameter p, with a mode at around $p = 0.75$, and is monotonic decreasing with respect to λ.

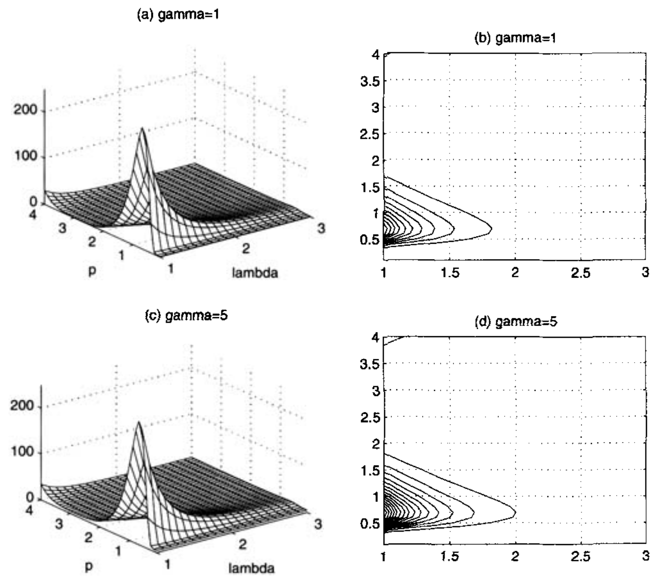

Figure 1. Surface and contour plots of $\eta(\lambda,p,\gamma)$ function.

These imply that the larger the shape parameter p ($p \ge 1$) and the larger the scale parameter, the smaller the variance, and thus a smaller L is required. Indeed, an observer with an exponential detection pattern ($p = 1$) will need more sampling effort than an observer with a half normal detection pattern ($p = 2$), provided both have the same scale of observation.

4.2 *Stopping Rule for Kernel Estimator*

A kernel density estimator for the probability density function of the sighting distance $f(x)$ is

$$\hat{f}_{kL}(x) = (n_L h_{n_L})^{-1}\sum_{i=1}^{n_L}\left\{K\left(\frac{x-X_i}{h_L}\right) + K\left(\frac{x+X_i}{h_L}\right)\right\}, \quad (7)$$

where K is a kernel function and h_L is the smoothing bandwidth, which controls the amount of smoothness. Comprehensive reviews and practical implementations of the kernel density estimation are available in Silverman (1986).

The kernel estimator is a result of applying the standard kernel estimator on an extended sighting sample and contains both the original sighting distance $X_1, \ldots, X_{n_L}$ and their reflections $-X_1, \ldots, -X_{n_L}$. The modification would be unnecessary if the signs of the sighting distances are kept in the original data.

From (7), a kernel estimator for $f(0)$ is

$$\hat{f}_{kL}(0) = 2(n_L h_{n_L})^{-1}\sum_{i=1}^{n_L} K\left(\frac{X_i}{h_L}\right).$$

The corresponding estimator of D is

$$\hat{D}_k(L) = \frac{n_L \hat{f}_{kL}(0)}{2L}.$$

The properties of $\hat{D}_k$ have been studied in Chen (1996a), including various methods for choosing h_L. The optimal bandwidth that minimizes the mean squared error of $\hat{f}_{kL}(0)$ is $h_L = C_0(n_L)^{-1/5}$, where C_0 is a constant free of L. If the Gaussian kernel is used and f is not too far away from a half normal density, then h_L can be chosen as

$$h_L = 0.812A(L)(n_L)^{-1/5},$$

where $A(L) = \min\{\sigma(L), IQ(L)/1.34\}$ and $\sigma(L)$ and $IQ(L)$ are the standard deviation and the interquantile range of the reflected sighting sample, respectively.

The bias of $\hat{D}_d$, as shown in Chen (1996a), is

$$\text{bias}(\hat{D}_{kL}) = \frac{1}{4}\sigma_k^2 f^{(2)}(0)C_0^2\{2D/f(0)\}^{3/5}L^{-2/5}+o\left(L^{-2/5}\right),$$

where $\sigma_k^2 = \int t^2K(t)dt$ and $\text{var}\{\hat{D}_{kL}\} = O(L^{-4/5})$. Thus, $\text{bias}^2\{\hat{D}_{kL}\} \neq o(\text{var}\{\hat{D}_{kL}\})$, as required in condition (a). To make the kernel estimator satisfy the condition, we use the bias-corrected estimator given in Chen (1996a),

$$\hat{D}_{kc}(L) = \frac{n_L\hat{f}_{kcL}(0)}{2L},$$

where $\hat{f}_{kcL} = \hat{f}_{kL}(0) - (1/2)\hat{f}^{(2)}(0)h_L^2\sigma_k^2$ is a bias-corrected estimator for $f(0)$,

$$\hat{f}^{(2)}(0) = 2(n_Lh_{1L})^{-1}\sum_{i=1}^{n_L}K^{(2)}(X_i/h_{1L})$$

is an estimator of $f^{(2)}(0)$ with another smoothing bandwidth h_{1L}, and $K^{(2)}$ is the second derivative of the kernel function K. The optimal h_{1L} is of order $n^{-1/9}$. A rule of thumb for choosing h_{1L} is $h_{1L} = h_Ln_L^{1/5-1/9}$. It can be shown that $\text{var}\{\hat{f}_{kcL}(0) \mid n\} = 2f(0)R(K)/(n_Lh_L) + O(n_L^{-1})$, where $R(K) = \int K^2(t)dt$. From (3), the variance of $\hat{D}_{kc}(L)$ is

$$\text{var}\{\hat{D}_{kcL}\} = (2L)^{-1}D\{2R(K)\text{E}(n_L/h_L)/\text{E}(n_L) + \gamma f(0).\}$$

An estimator for $\text{SD}(\hat{D}_{kcL})$ is

$$\widehat{\text{SD}}(\hat{D}_{kcL}) = \sqrt{(2L)^{-1}\hat{D}_{kcL}\{2R(K)/h_L + \hat{\gamma}\hat{f}_{kc}(0)\}}. \quad (8)$$

Then the asymptotic stopping rule for the kernel estimator is to stop the survey if $2Z_{1-\alpha/2}\widehat{\text{SD}}(\hat{D}_{kcL}) \leq d_0$.

5. Finite Sample Procedures

The asymptotic sequential rules use the standard normal percentile Z_r to approximate β_r, the percentile of $t(\hat{D}) = (\hat{D} - D)/(\widehat{\text{var}}(\hat{D}))^{1/2}$. The accuracy of the asymptotic rules relies on the quality of the normal approximations. In this section, β_r is approximated by the bootstrap method. Before the bootstrap was introduced, the asymptotic approximation might have been the only tool available. The bootstrap (Efron and Tibshirani, 1993) provides a powerful tool to profile the exact distribution of $t(\hat{D})$ and hence β_r in finite samples.

The finite-sample stopping rule for the parametric estimator is to stop the survey if $(\beta_{1-\alpha/2} - \beta_{\alpha/2})\widehat{\text{SD}}(\hat{D}_{pL}) \leq d_0$ and the stopping rule for the kernel estimator is to stop the survey if $(\beta_{1-\alpha/2} - \beta_{\alpha/2})\widehat{\text{SD}}(\hat{D}_{kcL}) \leq d_0$, where $\widehat{\text{SD}}(\hat{D}_{pL})$ and $\widehat{\text{SD}}(\hat{D}_{kcL})$ are given in (6) and (8).

For the nonparametric kernel estimator, the following bootstrap procedure can be used for approximating β_r:

Step 1. Generate a resample $X_1^*, \ldots, X_{n_L}^*$ of the original sighting sample $X_1, \ldots, X_{n_L}$ by random sampling with replacement from the original sample.

Step 2. Calculate the Studentized estimator for $\hat{D}_L$ and denote it as $t^*(\hat{D}_{kcL})$.

Step 3. Repeat steps 1 and 2 B times and obtain $t^{*1}(\hat{D})(L), \ldots, t^{*B}(\hat{D}_{kcL})$, which has been ranked in ascending order.

Step 4. Approximate β_r by the rth "sample" percentile of $t^{*1}(\hat{D}_{kcL}), \ldots, t^{*B}(\hat{D}_{kcL})$.

This is the so-called nonparametric bootstrap because the resamples are drawn from the empirical distribution of the original sample in step 1. For the parametric estimator $\hat{D}_p$, the parametric bootstrap may be used. It has exactly the same steps as the nonparametric bootstrap outlined above except that, in step 1, the bootstrap resamples are generated from a distribution with $f(x; \hat{\lambda}, \hat{p})$ as its density, where $\hat{\lambda}$ and $\hat{p}$ are the maximum likelihood estimates.

We have so far assumed the sighting distance is the basic sampling unit. Due to the clustering of biological populations, the basic sampling unit for a line transect survey may be transect segments or all the segments covered in 1 day. In this case, the bootstrap resampling carried out at step 1 should be modified so that the transect segments or survey days are resampled; see Buckland et al. (1993) and the next section for an implementation for the tuna survey.

6. Surveys for Southern Bluefin Tuna

We now apply the proposed sequential procedures for the southern bluefin tuna aerial survey data collected in 1993. There are 258 sightings after a total survey effort of $L =$ 14,065 (kilometers) from 19 survey days during a period of 2 1/2 months. The survey is run only when certain weather conditions (wind speed, cloud cover, etc.) are satisfied. Because the leasing and running costs of the airplane together with maintaining a survey crew is very high, it is of great interest to see if the sequential procedures produce abundance estimates that are cost effective while maintaining the desired scientific precision of the estimation. Three levels of d, 0.0004, 0.0005, and 0.0006, are used with the sequential rules. The results of the sequential analysis together with that of the full-sample analysis are presented in Table 1. Bootstrap estimates for D along with the standard errors are also provided for comparison. Because the plane flies transects that are mostly continuous, the daily transects rather than the sighting observations are the basic sampling unit, as suggested in Buckland et al. (1993). Thus, the resampling of survey days has been adopted for the bootstrap when considering the finite-sample rules.

The results contained in Table 1 can be summarized as follows: (i) Substantial reduction in survey effort L and subsequently the sample size n_L is realized due to the use of the sequential procedures, and the reduction is achieved at little cost of estimation precision, as shown by $\hat{D}$. (ii) The employment of the bootstrap sequential procedures further reduce L of the asymptotic procedures. (iii) At a given d, the nonparametric procedure requires less effort than its parametric counterpart, which is because p, as revealed by its parametric estimate, is in a region that leads to high variance of $\hat{D}$, as shown in Figure 1. (iv) The bootstrap reduces the difference between the parametric and nonparametric procedures.

Table 1
Estimation of population density for tuna data with different interval length d

	Full	$d = 0.0006$	$d = 0.0005$	$d = 0.0004$
	Parametric Estimator			
$\hat{D}_{pL}$	0.00155	0.00151	0.00154	0.00155
$\text{SE}(\hat{D}_{pL})$	0.00026	0.00029	0.00025	0.00026
$\hat{\lambda}$	6.5056	9.0471	7.3137	6.5056
$\hat{p}$	1.4391	1.8137	1.5633	1.4391
n_L (asymptotic)	258	117	223	258
n_L (finite sample)		101	117	223
L (asymptotic)	14,065.62	4813.42	11,011.65	14,065.62
L (finite sample)		3555.57	4813.42	11,011.65
Survey days (asymptotic)	19	6	14	19
Survey days (finite sample)		4	6	14
	Nonparametric Estimator			
$\hat{D}_{kcL}$	0.00145	0.00144	0.00149	0.00140
$\text{SE}(\hat{D}_{kcL})$	0.00018	0.00027	0.00024	0.00020
n_L (asymptotic)	258	117	154	223
n_L (finite sample)		101	123	171
L (asymptotic)	14,065.62	4813.41	6540.02	11,011.65
L (finite sample)		3555.57	5256.02	7508.78
Survey days (asymptotic)	19	6	9	14
Survey days (finite sample)		4	7	10

7. Simulation Results

Monte Carlo simulations are performed to investigate the performance of the proposed sequential procedures on the estimation of D. In each simulation, the positions of N animals are generated according to a Poisson distribution with intensity D in a strip with length $L = 100$ and width $2w = 10$. The transect line runs through the middle of the strip, with w distance to each side. The exponential power series detection function $g(x) = \exp\{-(x/\lambda)^p\}$ was used to detect the simulated population. An animal at a perpendicular distance X_i to the transect is detected with probability $g(X_i)$. We choose the population size $N = 500$ and the detection parameters $\lambda = 2$ and $p = 1.5, 2, 2.5$. The sequential procedures are based on the length of the 95% confidence intervals, which are chosen to be $d = 0.15$ and 0.2.

Tables 2 and 3 contain the average abundance estimates

Table 2
Sequential parametric estimation for $N = 500$ with different interval length d and shape parameter p. The numbers inside the parentheses are the bootstrap (finite sample) counterparts of the quantities beside them.

Interval length	Full	$d = 0.2$	$d = 0.15$
	$\lambda = 2$, $p = 1.5$		
$\hat{D}_{pL}$	0.483	0.480	0.480
$\text{SE}(\hat{D}_{pL})$	0.083 (0.085)	0.093 (0.116)	0.084 (0.104)
n_L	178.81 (178.16)	116.98 (93.74)	143.31 (105.92)
L	99.43 (99.44)	65.14 (52.42)	79.74 (59.34)
Coverage	94.4% (91.8%)	99.8% (99.6%)	97.4% (99.6%)
	$\lambda = 2$, $p = 2$		
$\hat{D}_{pL}$	0.504	0.496	0.491
$\text{SE}(\hat{D}_{pL})$	0.067 (0.068)	0.082 (0.095)	0.073 (0.085)
n_L	177.09 (177.72)	101.75 (91.79)	119.86 (100.49)
L	99.41 (99.45)	57.23 (51.29)	67.52 (56.40)
Coverage	96.4% (94.4%)	100% (99.2%)	98.0% (97.2%)
	$\lambda = 2$, $p = 2.5$		
$\hat{D}_{pL}$	0.508	0.509	0.504
$\text{SE}(\hat{D}_{pL})$	0.053 (0.054)	0.073 (0.077)	0.067 (0.072)
n_L	177.68 (177.73)	94.65 (90.20)	102.50 (95.34)
L	99.41 (99.41)	53.03 (50.73)	57.46 (53.34)
Coverage	97.4% (93.8%)	99.8% (99.2%)	98.0% (99%)

Table 3

Sequential nonparametric estimation for $N = 500$ with different interval length d and shape parameter p. The numbers inside parentheses are the bootstrap (finite-sample) counterparts of the quantities beside them.

Interval length	Full	$d = 0.2$	$d = 0.15$
	$\lambda = 2,\ p = 1.5$		
$\hat{D}_{kcL}$	0.486	0.482	0.473
$\mathrm{SE}(\hat{D}_{kcL})$	0.070 (0.070)	0.084 (0.095)	0.076 (0.084)
n_L	178.81 (178.87)	120.30 (88.99)	146.79 (112.51)
L	99.43 (99.49)	66.70 (49.45)	81.93 (62.89)
Coverage	96.4% (92.8%)	98.8% (97.6%)	96.6% (93.2%)
	$\lambda = 2,\ p = 2$		
$\hat{D}_{kcL}$	0.502	0.497	0.493
$\mathrm{SE}(\hat{D}_{kcL})$	0.077 (0.077)	0.091 (0.101)	0.079 (0.083)
n_L	177.09 (176.53)	120.71 (90.73)	163.06 (138.13)
L	99.41 (99.42)	67.78 (51.35)	91.74 (78.32)
$\hat{N}(d)$	502 (502)	497 (497)	493 (493)
Coverage	96.8% (94.8%)	99.4% (97.6%)	95.4% (92.2%)
	$\lambda = 2,\ p = 2.5$		
$\hat{D}_{kcL}$	0.506	0.499	0.499
$SE(\hat{D}_{kcL})$	0.081 (0.080)	0.094 (0.103)	0.082 (0.083)
n_L	177.68 (177.24)	123.92 (96.44)	169.20 (157.98)
L	99.41 (99.43)	69.47 (54.44)	94.85 (89.08)
Coverage	97.2% (95.8%)	99.2% (96.8%)	93.8% (92.4%)

$\hat{D}$ and their standard errors, the average sample size n, the survey effort L, and the coverage probabilities. The parametric results given in Table 2 also contain average estimates for the parameters p and λ. Because the simulated points are detected independently, the basic sampling unit is each individual point, which means that, when applying the finite-sample rule, the detected points are resampled in the bootstrap.

Both Tables 2 and 3 confirm that, by paying a small price in the precision of estimating D, i.e., by increasing the length of the confidence intervals, there will be substantial reduction in the survey effort by applying both the parametric and nonparametric sequential procedures. In comparison with their full-sample counterparts, the sequential estimates of D are little changed, which is especially the case for the parametric method. The sequential estimates have larger standard error, which is expected because the length of the confidence interval has been increased. The survey effort L is further reduced by applying the bootstrap (finite-sample) sequential rules. When $d = 0.2$, the reduction in L ranges from 35 to 55% for the asymptotic parametric rule and ranges from 35 to 40% for the asymptotic nonparametric rule. The reduction is increased to around 50% by using the bootstrap. In general, we observe that the parametric/bootstrap rule requires less effort than the nonparametric/asymptotic rule. These are expected because the parametric estimation within a correctly specified parametric family is more efficient than the nonparametric estimation.

Table 2 also shows that, when p increases from 1.5 to 2.5, the standard error of the parametric estimator $\mathrm{SE}(\hat{D}_{pL})$ decreases, which confirms our early observation in Section 4.1. As a result, the required survey effort L decreases as p increases. The above phenomena are reversed in Table 3 for the nonparametric estimator, i.e., both the $\mathrm{SE}(\hat{D}_{kcL})$ and the required survey effort L increase as p increases. The dependence of the nonparametric rules on the detection parameters is weaker than the parametric rules, which indicates that the nonparametric rules are less sensitive to the parameters.

Both Tables 2 and 3 show that the use of the bootstrap increases the standard errors and reduces L. That this happens at the same time seems contradictory, because an increased standard error generally means an increased survey effort in order to reduce the variability. To reconcile these, we report in Table 4 the average $\beta_{0.975} - \beta_{0.025}$, which is the difference between the two bootstrap quantiles of Studentized statistics $t(\hat{D}_L)$. We find the average differences of the two bootstrap quantiles are much smaller than 2×1.96, which is the corresponding quantile difference of $\mathrm{N}(0, 1)$. This means that the distribution of $t(\hat{D}_L)$ concentrates more around zero than $\mathrm{N}(0, 1)$. Therefore, the increase variation is well compensated by the reduction in $\beta_{0.975} - \beta_{0.025}$.

8. Discussion

The sequential procedure proposed has been found to be useful and relevant to density estimation of wildlife abundance. The parametric method performs better than the nonparametric method at the risk of making wrong assumptions regarding the detection function. It is shown that reduction in the survey effort can be achieved at little cost of estimation precision. Furthermore, the choice of d in most wildlife experiments does not have to be very small in order to be useful.

In this article, we have reported on our study of univariate line transect surveys. After understanding the sequential pro-

Table 4
Bootstrap quantile estimation of $\beta_{0.025}$, $\beta_{0.975}$, *and* $\beta_{0.975} - \beta_{0.025}$

	Parametric bootstrap (nonparametric bootstrap)		
	$p = 1.5$	$p = 2$	$p = 2.5$
$\beta_{0.025}$	−1.12 (−1.57)	−1.31 (−1.69)	−1.33 (−1.74)
$\beta_{0.975}$	1.53 (1.89)	1.76 (1.91)	1.88 (1.94)
$\beta_{0.975} - \beta_{0.025}$	2.65 (3.46)	3.06 (3.60)	3.20 (3.68)

cedures in univariate line transect surveys, we can extend the sequential analysis to line transect surveys for clustered populations, which has another covariate, namely, the school size, in additional to the sighting distance. In this extension, the work of Drummer and McDonald (1987) for parametric estimation and the works of Chen (1996b) and Mack and Quang (1998) can be incorporated into the sequential procedures considered in this article. When the detection of animals depends on a general set of covariates, similar sequential procedures can be developed based on the estimation method considered in Quang and Becker (1996).

Acknowledgements

The authors thank a referee for suggesting using the survey day as the resampling unit in the bootstrap for the analysis of the tuna aerial survey. We also thank another referee, the associate editor, and the coeditor, Marie Davidian, for comments that improved the presentation of the article. Part of the research was supported by a grant from Hong Kong University.

Résumé

Cet article étudie l'intérêt de l'utilisation de procédures d'échantillonnage séquentiel pour la détermination de l'effort d'échantillonnage requis dans une étude par transects linéaires pour atteindre une certaine précision dans l'estimation de l'abondance de populations biologiques. Des procédures séquentielles sont construites pour des estimateurs paramétriques et non paramétriques de l'abondance des animaux. Le critère utilisé pour décider de l'arrêt est la largeur de l'intervalle de confiance de l'abondance. Pour chaque estimateur étudié, nous développons des règles d'arrêt à partir de distributions asymptotiques et de distributions estimées par la méthode du bootstrap. Une analyse séquentielle sur des données d'une étude aéroportée d'une espèce de thon indique qu'une économie substantielle d'effort d'échantillonnage peut être faite en utilisant notre procédure. Cette économie est aussi observée dans une étude par simulation destinée à estimer la performance empirique de notre procédure.

References

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

Carroll, R. J. (1976). On sequential density estimation. *Probability Theory and Related Fields* **36,** 137–151.

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

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

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

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* London: Chapman and Hall.

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

Pollock, K. H. (1978). A family of density estimators for line-transect sampling. *Biometrics* **34,** 475–478.

Quang, P. X. and Becker, E. F. (1996). Line transect sampling under varying conditions, with applications to aerial surveys. *Ecology* **77,** 1297–1302.

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

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

Received January 2001. *Revised January* 2002.
Accepted January 2002.