4

Line transects

4.1 Introduction

The purpose of this chapter is to illustrate the application of the theory of Chapter 3 to line transect data, and to present the strategies for analysis outlined in Section 2.5. In general, the principal parameter of a line transect analysis does not have a closed form estimator. Instead, numerical methods are required; it is generally not possible to substitute statistics computed from the data into formulae to estimate object density. Using pen and paper and a pocket calculator, a fairly simple analysis might take months. Instead, we rely on specialized computer software to analyse distance sampling data.

This chapter uses a series of examples in which complexities are progressively introduced. The examples come from simulated data for which the parameters are known; this makes comparisons between estimates and true parameters possible. However, in every other respect, the data are treated as any real data set undergoing analysis, where the parameters of interest are unknown. A simple data set, where each object represents an individual animal, plant, nest, etc., is first introduced. Truncation of the distance data, modelling the spatial variation of objects to estimate var(n), grouping of data, and model selection philosophy and methods are then addressed. Once an adequate model has been selected, we focus on statistical inference given that model, to illustrate estimation of density and measures of precision. Finally, the objects are allowed to be clusters (coveys, schools, flocks). Cluster size is first assumed to be independent of distance and then allowed to depend on distance.

The example data are chosen to be 'realistic' from a biological standpoint. The data (sample size, distances and cluster sizes) are generated stochastically and simulate the case where the assumptions of line transect sampling are true. Thus, no objects went undetected on the line ($g(0) = 1$), no movement occurred prior to detection, and data were free

of measurement error (e.g. heaping at zero distance). In addition, the sample size n was adequate. The assumptions and survey design to ensure they are met are discussed in Chapter 7. Examples illustrating analysis of real data where some of these assumptions fail are provided in Chapter 8. The example data of this chapter are analysed using various options of program DISTANCE. In the penultimate section, some comparative analyses using program SIZETRAN (Drummer 1991) are carried out.

4.2 Example data

The example comprises an area of size A, whose boundary is well defined, and sampled by 12 parallel line transects ($l_1, l_2, \ldots, l_{12}$). The area is irregularly shaped, so that the lines running from boundary to boundary are of unequal length. We assume that no stratification is required and that the population was sampled once by a single observer to exacting standards; hence the key assumptions have been met. The distance data, recorded in metres, were taken without a fixed transect width (i.e. $w = \infty$), and ungrouped, to allow analysis of either grouped or ungrouped data. The detection function $g(x)$ was a simple half-normal with $\sigma = 10$ m, giving $f(0) = \sqrt{\{2/(\pi\sigma^2)\}} = 0.079788\,\text{m}^{-1}$. The n_i were drawn from a negative binomial distribution such that the spatial distribution of objects was somewhat clumped (i.e. $\text{var}(n) > n$). Specifically, $\text{var}(n_i|l_i) = 2E(n_i|l_i)$.

The total length ($L = \Sigma l_i$) of the 12 transects was 48 000 m and $n = 105$ objects were detected. Their distances from the transect lines were measured carefully in metres. $E(n) = 96$, thus somewhat more were observed than expected (105 vs. 96). The true density is

$$D = \frac{N}{A} = \frac{E(n) \cdot f(0)}{2L} \text{ objects/m}^2$$

where all measurements are in metres. To convert density from numbers per m^2 into numbers per km^2, multiply by 1 000 000. The true density is known in this simulated example to be approximately 80 objects per km^2; the actual value is $79.788/\text{km}^2$.

The first step is to examine the distance data by plotting histograms using various distance categories. It is often informative to plot a histogram with many fine intervals (Fig. 4.1). Here one can see the presence of a broad shoulder, no evidence of heaping, and no indication of evasive movement prior to detection; the data appear to be 'good', which we happen to know to be true here.

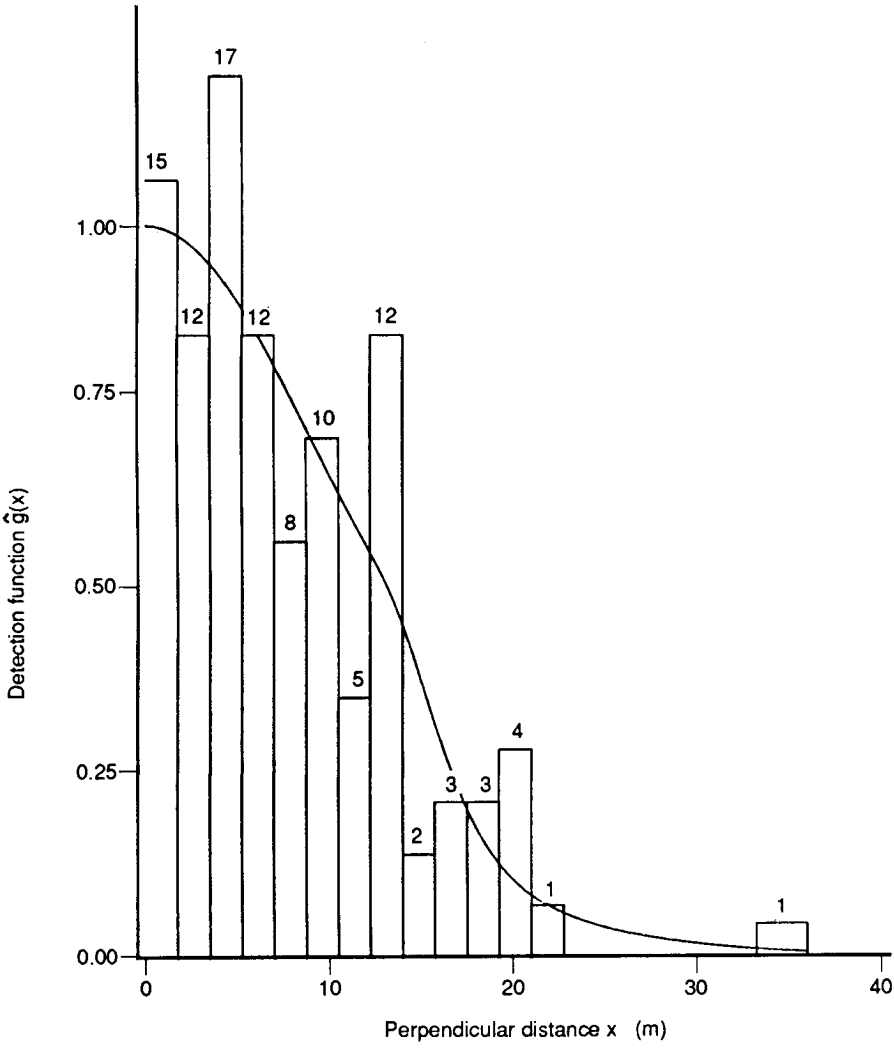

Fig. 4.1. Histogram of the example data using 20 distance categories. A fitted hazard-rate key with one cosine adjustment term is shown as a candidate model for the detection function, $g(x)$.

4.3 Truncation

Inspection of the histogram in Fig. 4.1 shows the existence of an extreme observation or 'outlier' at 35.82 m. A useful rule of thumb is to truncate at least 5% of the data; here the six most extreme distances are 19.27,

19.42, 19.44, 19.46, 21.21 and 35.82 m. Thus, w could initially be set for purposes of analysis at 19 m. An alternative is to fit a reasonable preliminary model to the data, compute $\hat{g}(x)$ to find the value of x such that $\hat{g}(x) = 0.15$, and use this value of x as the truncation point for further analysis.

As an illustration, the half-normal key function was fitted to the ungrouped, untruncated data and found to fit well. This approach suggested a truncation point of 19 m for the half-normal model, based on the criterion that $\hat{g}(x) \doteq 0.15$ (actually $\hat{g}(19) = 0.13$). The deletion of outliers is useful because these extreme observations provide little information for estimating $f(0)$, the density function at $x = 0$, but can be difficult to model. The series expansions require additional adjustment terms to fit the few data in the tail of the distance distribution, which may unnecessarily increase the sampling variance of the density estimate. In this example, both truncation rules suggest $w \doteq 19$, leaving $n = 99$ observations. For the rest of this chapter we will emphasize the results with $w = 19$ m, but estimates corresponding to no truncation will also be given and compared. The choice of truncation point is not a critical decision for these example data where all the assumptions are met and the true detection function is simple.

For the true model, the quantity $E(n) \cdot f(0)$ remains unchanged as the truncation point is varied. Consequently, for good data (i.e. data satisfying the assumptions) and a reasonable model for $g(x)$, the product $n \cdot \hat{f}(0)$ is quite stable over a range of truncation points. With increased truncation, n decreases, but $\hat{f}(0)$ increases to compensate. The estimate $n \cdot \hat{f}(0)$ under the half-normal model is 8.477 if data are truncated at 19 m, and 8.417 without truncation.

Truncation of the data at $w = 19$ m removed only six detections. If a series expansion model is used, up to three fewer parameters are required to model the truncated data than the untruncated data (Table 4.1). (Note that the truncation distance w supplied to DISTANCE must be finite; by 'untruncated data', we mean that w was at least as large as the largest recorded distance.) Outliers in the right tail of the distance distribution required additional adjustment parameters and the inclusion of such terms increased the sampling variance of $\hat{f}(0)$ and hence $\hat{D}$ when a robust but incorrect model was used (Table 4.2). If the correct model could somehow be known and used, then truncation is unimportant if the measurements are exact and no evasive movement prior to detection is present.

Truncation of the distance data for analysis deletes outliers and facilitates modelling of the data. However, as some data are discarded, one might ask if the uncertainty in $\hat{D}$ increases. First, this issue is examined when the true model is known and used (i.e. the half-normal

in this case). The coefficient of variation increased about 1% when the data were truncated at 19 m relative to untruncated (Table 4.2). Thus, little precision was lost due to truncation if the data were analysed under the true model. Of course, one never knows the true detection function except for computer simulation examples.

Table 4.1 Summary of AIC values at two truncation values w for the example data analysed as ungrouped and at three different groupings (five groups of equal width, 20 groups of equal width, and five unequal groups such that the number detected was equal in each group). The models with minimum AIC values are indicated by an asterisk

Data type	Model (key + adjustment)	w = 19 m			w = largest observation		
		No. of parameters			No. of parameters		
		Key	Adjust.	AIC	Key	Adjust.	AIC
Ungrouped	Uniform + cosine	0	1	562.98	0	2	636.48*
	Uniform + polynomial	0	1	563.28	0	4	638.18
	Half-normal + Hermite	1	0	562.60*	1	0	636.98
	Hazard-rate + cosine	2	0	565.22	2	0	639.16
Grouped (5 equal)	Uniform + cosine	0	1	300.91	0	3	224.77
	Uniform + polynomial	0	1	301.09	0	4	226.75
	Half-normal + Hermite	1	0	300.63*	1	0	222.13*
	Hazard-rate + cosine	2	0	303.18	2	0	224.21
Grouped (20 equal)	Uniform + cosine	0	1	563.58	0	2	520.88
	Uniform + polynomial	0	1	563.40	0	3	524.78
	Half-normal + Hermite	1	0	563.03*	1	0	520.31*
	Hazard-rate + cosine	2	0	565.80	2	1	523.53
Grouped (5 unequal)	Uniform + cosine	0	1	323.05*	0	2	345.13
	Uniform + polynomial	0	1	324.45	0	4	348.56
	Half-normal + Hermite	1	0	323.35	1	0	342.54*
	Hazard-rate + cosine	2	0	324.32	2	0	344.95

When series expansion models are used for the analysis of the example data, the uniform key function with either cosine or polynomial adjustments gives a smaller coefficient of variation when the data are truncated (Table 4.2). This small increase in precision is because only one parameter was required for a good model fit when $w = 19$ m, whereas two to four parameters were required to fit the untruncated data (Table 4.2). Precision was better for the untruncated data for the hazard-rate model.

The effect of truncation on the point estimates was relatively small, and estimates were not consistently smaller or larger than when data were untruncated (Table 4.2). The various density estimates ranged from 72.75 to 94.09, and their coefficients of variation ranged from 14.8% to 20.3%. The true parameter value was $D = 80$ objects/km^2.

In general, some truncation is recommended, especially for obvious outliers. Although some precision might be lost due to truncation, it is

usually slight. Often, precision is increased because fewer parameters are required to model the detection function. Most importantly, truncation will often reduce bias in $\hat{D}$ or improve precision, or both, by making the data easier to model. Extreme observations in the right tail of the distribution may arise from a different detection process (e.g. a deer seen at some distance from the observer along a forest trail, or a whale breaching near the horizon), and are generally not informative, in addition to being difficult to model. Truncation is an important tool in the analysis of distance sampling data.

Table 4.2 Summary of estimated density $\hat{D}$ and coefficient of variation cv for two truncation values w for the example data. Estimates are derived for four robust models of the detection function. The data analysis was based on ungrouped data and three different groupings (five groups of equal width, 20 groups of equal width, and five unequal groups such that the number detected was nearly equal in each group)

Data type	Model (key + adjustment)	Truncation			
		w = 19 m		w = largest obsn	
		$\hat{D}$	cv(%)	$\hat{D}$	cv(%)
Ungrouped	Uniform + cosine	90.38	15.9	80.52	16.8
	Uniform + polynomial	78.95	14.8	84.53	20.0
	Half-normal + Hermite	88.31	16.7	87.68	15.3
	Hazard-rate + cosine	84.23	18.4	72.75	15.6
Grouped (5 equal)	Uniform + cosine	88.69	15.9	94.09	16.7
	Uniform + polynomial	79.37	15.2	88.39	19.1
	Half-normal + Hermite	86.94	16.8	92.16	15.8
	Hazard-rate + cosine	84.49	19.6	80.80	16.7
Grouped (20 equal)	Uniform + cosine	89.95	15.8	80.06	15.1
	Uniform + polynomial	79.10	14.9	74.43	15.3
	Half-normal + Hermite	87.98	16.6	86.87	15.7
	Hazard-rate + cosine	85.81	19.2	84.06	18.1
Grouped (5 unequal)	Uniform + cosine	86.14	16.3	81.40	19.0
	Uniform + polynomial	78.60	15.8	86.91	17.8
	Half-normal + Hermite	85.12	17.0	88.84	16.3
	Hazard-rate + cosine	86.83	20.3	82.54	17.7

4.4 Estimating the variance in sample size

Before the precision of an estimate of density can be assessed, attention must be given to the spatial distribution of the objects of interest. If the n detected objects came from a sample of objects that were randomly (i.e. Poisson) distributed in space, then $\text{var}(n) = E(n)$ and $\widehat{\text{var}}(n) = n$.

Because most biological populations exhibit some degree of clumping, one expects $\text{var}(n) > E(n)$. Thus, empirical estimation of the sampling variance of n is recommended. This makes it nearly imperative to sample using several lines, l_i, such as the 12 used in the example. Variation in the number of detections found on each of the lines, n_i, provides a valid estimate of $\text{var}(n)$ without having to resort to the Poisson assumption and risk what may be a substantial underestimate of the sampling variance of the estimator of density.

After truncating at 19 m, the line lengths in km and numbers of detections (l_i, n_i) for the $k = 12$ lines were: (5, 14), (2, 2), (6, 8), (4, 8), (3, 3), (1, 4), (4, 10), (4, 8), (5, 17), (7, 20), (3, 0), and (4, 5). The estimator for the empirical variance of n is (from Section 3.7.2)

$$\widehat{\text{var}}(n) = L \sum_{i=1}^{k} l_i \left(\frac{n_i}{l_i} - \frac{n}{L} \right)^2 / (k-1)$$

$$= 195.8$$

This estimate is based on $k - 1 = 11$ degrees of freedom. The ratio of the empirical variance to the estimated Poisson variance is $195.8/99 = 1.98$, indicating some spatial aggregation of objects. Equivalently, one can estimate the sampling variance of the encounter rate (n/L),

$$\widehat{\text{var}}(n/L) = \frac{\sum_{i=1}^{k} \frac{l_i}{L} \left(\frac{n_i}{l_i} - \frac{n}{L} \right)^2}{k-1}$$

$$= 0.0850$$

$$\widehat{\text{se}}(n/L) = \sqrt{\widehat{\text{var}}(n/L)}$$

$$= 0.292$$

Then $\widehat{\text{se}}(n) = L \cdot \widehat{\text{se}}(n/L)$ and $\widehat{\text{var}}(n) = [\widehat{\text{se}}(n)]^2$. In most subsequent analyses of these data, we use the empirical estimate, $\widehat{\text{var}}(n) = 195.8$.

4.5 Analysis of grouped or ungrouped data

Analysis of the ungrouped data is recommended for the example because it is known that the assumptions of line transect sampling hold. General statistical theory and our experience indicate that little efficiency is lost by grouping data, even with as few as five or six well-chosen

intervals. Grouping of the data can be used to improve robustness in the estimator of density in cases of heaping and movement prior to detection (Chapter 7).

For the example, changes in the estimates of density under a given model were in most cases slight (much smaller than the standard error) whether the analysis was based on the ungrouped data or one of the three sets of grouped data. This is a general result if the assumptions of distance sampling are met. If heaping, errors in measurement, or evasive movement prior to detection are present, then appropriate grouping will often lead to improved estimates of density and better model fit. Grouping the data is a tool for the analysis of real data to gain estimator robustness. When heaping occurs, cutpoints should be selected to avoid favoured rounding distances as far as possible. Thus, if values tend to be recorded to the nearest 10 m, cutpoints might be defined at 5 m, 15 m, 25 m, ... The first cutpoint is the most critical. If assumptions hold, the first interval should be relatively narrow, so that the first cutpoint is on the shoulder of the detection function. However, it is not unusual for 10% or more of detections to be recorded as on the centreline, especially when perpendicular distances are calculated from sighting distances and angles. In this circumstance, the width of the first interval should be chosen so that few detections are erroneously allocated to the first interval through measurement error, and in particular, through rounding a small sighting angle to zero.

4.6 Model selection

4.6.1 The models

Results for fitting the detection function are illustrated using the uniform, half-normal and hazard-rate models as key functions. Cosine and simple polynomial expansions are used with the uniform key, Hermite polynomials are used with the half-normal key, and a cosine expansion is used with the hazard-rate key. Thus, four models for $g(x)$ are considered for the analysis of these data. Modelling in this example can be expected to be relatively easy as the data are well behaved, exhibit a shoulder, and the sample size is relatively large ($n = 105$ before truncation). With such ideal data, the choice of model is unlikely to affect the abundance estimate much, whereas if survey design or data collection is poor, different models might yield substantially different estimates.

From an inspection of the data in Fig. 4.1, it is clear that the uniform key function will require at least one cosine or polynomial adjustment term. The data here were generated under a half-normal detection

function so we might expect the half-normal key to be sufficient without any adjustment terms. However, the data were stochastically generated; the addition of a Hermite polynomial term is quite possible, although it would just fit 'noise'. The hazard-rate key has two parameters and seldom requires adjustment terms when data are good. In general, a histogram of the untruncated data using 15–20 intervals will reveal the characteristics of the data. Such a histogram will help identify outliers, heaping, measurement errors, and evasive movement prior to detection.

4.6.2 Likelihood ratio tests

The addition of adjustment terms to a given key function can be judged using likelihood ratio tests (LRTs). This procedure is illustrated using the example data, with $w = 19$, and ungrouped data. Assume the key function is the 1-parameter half-normal. This model is fitted to the distance data to provide the MLE of the parameter σ. Does an adjustment term significantly improve the fit of the model of $g(x)$ to the data? Let $\mathcal{L}_0$ be the value of the likelihood for the 1-parameter half-normal model and $\mathcal{L}_1$ be its value for the 2-parameter model (half-normal model plus one Hermite polynomial adjustment term). Then, the test statistic for this likelihood ratio test is

$$\chi^2 = -2 \log_e (\mathcal{L}_0/\mathcal{L}_1)$$

and is distributed asymptotically as χ^2 with 1 df if the 1-parameter model ($\mathcal{L}_0$) is the true model. In general, the degrees of freedom are calculated as the difference in the number of parameters between the two models being tested. This is a test of the null hypothesis that the 1-parameter model is the true model against the alternative hypothesis that the 2-parameter model is the true model. If the additional term makes a significant improvement in the fit, then the test statistic will be 'large'. For the example, $\log_e(\mathcal{L}_0) = -280.3000$ and $\log_e(\mathcal{L}_1) = -280.2999$. These are values of the log-likelihood function computed at the MLE values of the parameters. Then, the test statistic is

$$\begin{aligned}\chi^2 &= -2 \log_e(\mathcal{L}_0/\mathcal{L}_1) \\ &= -2[\log_e(\mathcal{L}_0) - \log_e(\mathcal{L}_1)] \\ &= -2[-280.3000 - (-280.2999)] \\ &= 0.0002\end{aligned}$$

This test statistic has 1 df, so that $p = 0.988$, and there is no reason to add a Hermite polynomial adjustment term. This does not necessarily

mean that the 1-parameter half-normal model is an adequate fit to the example data; it only informs us that the 2-parameter model is not a significant improvement over the 1-parameter model. If goodness of fit for example indicates that both models are poor, it is worth investigating whether a 3-parameter model is significantly better than the 1-parameter model. This may be done by setting the LOOKAHEAD option in DISTANCE to 2. Another solution is to try a different model.

A second example is that shown in Fig. 4.1, the untruncated example data modelled by a hazard-rate key and cosine adjustment terms. Let $\mathcal{L}_0$ be the likelihood under the 2-parameter hazard-rate model and $\mathcal{L}_1$ be the likelihood under this same model with one cosine adjustment term. MLEs of the parameters are found under both models with the resulting log-likelihood values: $\log_e(\mathcal{L}_0) = -259.898$ and $\log_e(\mathcal{L}_1) = -258.763$. Which is the better model of the data? Should the cosine term be retained? These questions are answered by the LRT statistic,

$$\begin{aligned}\chi^2 &= -2\,[\log_e(\mathcal{L}_0) - \log_e(\mathcal{L}_1)] \\ &= -2\,[-259.898 - (-258.763)] \\ &= 2.27\end{aligned}$$

Because $\mathcal{L}_0$ has two parameters and $\mathcal{L}_1$ has three parameters, the LRT has 1 df. Here, $\chi^2 = 2.27$, 1 df, $p = 0.132$. As noted in Section 3.5.2, use of $\alpha = 0.15$ instead of the conventional $\alpha = 0.05$ might be found useful as a rejection criterion. Thus, the 2-parameter model is rejected in favour of a 3-parameter model, with a single cosine adjustment to the hazard-rate key. The procedure is repeated to examine the adequacy of the 3-parameter model against a 4-parameter model with two cosine terms. Note that this illustration used the untruncated data; additional terms are frequently needed to model the right tail of the distance data if proper truncation has not been done.

If the LRT indicates that a further term is not required but goodness of fit (below) indicates that the fit is poor, the addition of two terms (using DISTANCE option LOOKAHEAD = 2) may provide a significantly better fit. If it is important to obtain the best possible fit, options SELECT = **forward** and SELECT = **all** of DISTANCE may prove useful.

4.6.3 Akaike's Information Criterion

The use of the Akaike's Information Criterion (AIC) provides an objective, quantitative method for model selection (see Burnham and Anderson (1992) for application of AIC and Akaike (1985) for theoretical

synthesis). It is similar in character to a likelihood ratio test for hierarchical models, but is equally applicable to selection between non-hierarchical models. The criterion is

$$\text{AIC} = -2 \cdot [\log_e(\mathcal{L}) - q]$$

where $\log_e(\mathcal{L})$ is the value of the log-likelihood function evaluated at the maximum likelihood estimates of the model parameters and q is the number of parameters in the model (Section 3.5.3). AIC is computed for each candidate model, and that with the lowest AIC is selected for analysis and inference. Having selected a model, one should check that it fits as judged by the usual χ^2 goodness of fit statistics. Visual inspection of the estimated detection function plotted on the histogram is also informative because one can better judge the model fit near the line, and perhaps discount some lack of fit in the right tail of the data.

AIC was computed for the four models noted above for both grouped and ungrouped data with and without truncation (Table 4.1). For computational reasons, w was set equal to the largest observation in the case of no truncation. Three sets of cutpoints were considered for each model for illustration. Set 1 had five groups of equal width, set 2 had 20 groups of equal width, and set 3 had five groups whose width increased with distance, such that the number detected in each distance category was nearly equal. Note that AIC cannot be used to select between models if the truncation distances w differ, or, in the case of an analysis of grouped data, if the cutpoints differ.

AIC values in Table 4.1 indicate that the half-normal model is the best of the four models considered. Here, we happen to know that this is the true model. All four models have generally similar AIC values within any set of analyses of Table 4.1. Still, AIC selects the half-normal model in three of the four instances, both with truncation and without. Thus, the main analysis will focus on the ungrouped data, truncated at $w = 19$ m, under the assumption that $g(x)$ is well modelled by the half-normal key function with no adjustment parameters. The fit of this model is shown in Fig. 4.2. One might suspect that all four models would provide valid inference because of the similarity of the AIC values. Often, the AIC will identify a subset of models that are clearly inferior and these should be discarded from further consideration.

4.6.4 Goodness of fit

Goodness of fit is described briefly in Section 3.5.4, and is the last of the model selection criteria we consider here. Although it is the fit of

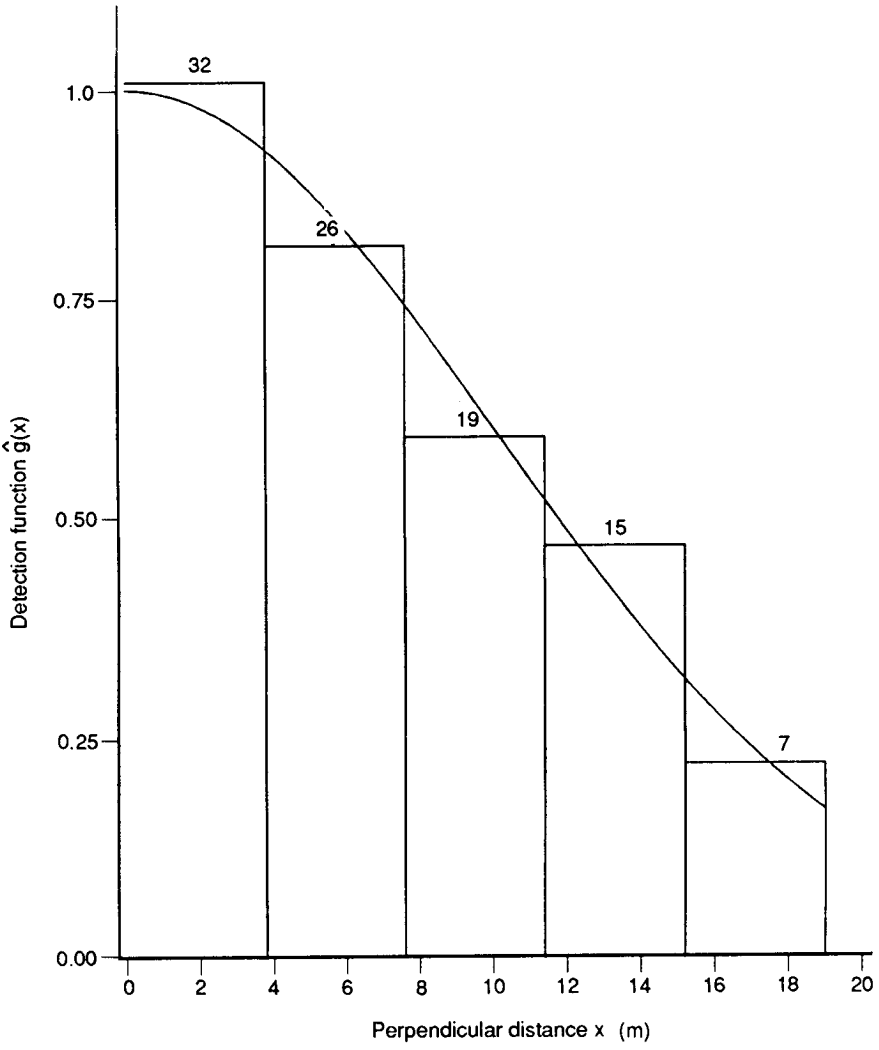

Fig. 4.2. Histogram of the example data using five distance categories. A half-normal detection function, fitted to the ungrouped data with $w = 19$ m, is shown and was used as a basis for final inference from these data.

the model near zero distance that is most critical, none of the model selection criteria of goodness of fit statistics, AIC and likelihood ratio tests give special emphasis to this region.

Some goodness of fit statistics for the example with $w = 19$ m and 20 groups are given in Table 4.3. These data were taken when all the

assumptions were met; all four models fit the data well and yield similar estimates of density.

Table 4.3 Goodness of fit statistics for models fitted to the example data with $w = 19$ m and 20 groups

Model	χ^2	*df*	*p*
Uniform + cos	14.58	17	0.62
Uniform + poly	13.11	17	0.73
Half-normal + Hermite	13.63	17	0.69
Hazard-rate + cos	14.91	16	0.53

Real data are often heaped, so that no parsimonious model seems to fit the data well, as judged by the χ^2 test. Grouping can be carried out to smooth the distance data and, thus, obtain an improved fit. While grouping usually results in little change in $\hat{D}$, it provides a more acceptable assessment of the fit of the model to the data. If possible, groups should be selected so that there is one favoured rounding distance per interval, and it should occur at the midpoint of the interval. The grouped nature of the (rounded) data is then correctly recognized in the analysis. If cutpoints are badly chosen, heaping will generate spurious significant χ^2 values.

4.7 Estimation of density and measures of precision

4.7.1 The standard analysis

Preliminary analysis leads us to conclude that the half-normal model is an adequate model of the detection function, with truncation of the distance data at $w = 19$ m, fitted to ungrouped data, and using the empirical variance of n.

Replacing the parameters of Equation 3.4 by their estimators and simplifying under the assumptions that objects on the line are detected with certainty, detected objects are recorded irrespective of which side of the line they occur, and objects do not occur in clusters, estimated density becomes

$$\hat{D} = \frac{n \cdot \hat{f}(0)}{2L}$$

where n is the number of objects detected, L is the total length of transect line, and $\hat{f}(0)$ is the estimated probability density evaluated at

zero perpendicular distance. For the example data, and adopting the preferred analysis, program DISTANCE yields $\hat{f}(0) = 0.08563$ with $\widehat{\text{se}}\{\hat{f}(0)\} = 0.007601$. The units of measure for $f(0)$ are 1/metres. Often, estimates of the effective strip width, $\mu = 1/f(0)$, are given in preference to $\hat{f}(0)$, since it has an intuitive interpretation. It is the perpendicular distance from the line for which the number of objects closer to the line that are missed is equal to the number of objects farther from the line (but within the truncation distance w) that are detected.

For the half-normal model, if data are neither grouped nor truncated, a closed form expression for $\hat{f}(0)$ exists (Chapter 3):

$$\hat{f}(0) = \sqrt{\{2/(\pi\hat{\sigma}^2)\}}$$

where $\hat{\sigma}^2 = \sum x_i^2/n$

Similarly, closed form expressions exist for the Fourier series estimator (uniform key + cosine adjustment terms) of $f(0)$ (Burnham *et al.* 1980: 56–61). However, generally the MLE of $f(0)$ must be computed numerically because no closed form equation exists.

The estimate of density for the example data truncated at 19 m, and using $\hat{f}(0)$ from DISTANCE, is

$$\begin{aligned}\hat{D} &= n \cdot \hat{f}(0)/2L \\ &= (99 \times 0.08563)/(2 \times 48) \\ &= 0.0883\end{aligned}$$

Since the units of $\hat{f}(0)$ are m and those for L are km, multiplying by 1000 gives

$$\hat{D} = 88.3 \text{ objects/km}^2$$

The estimator of the sampling variance of this estimate is

$$\widehat{\text{var}}(\hat{D}) = \hat{D}^2 \cdot \{[\text{cv}(n)]^2 + [\text{cv}\{\hat{f}(0)\}]^2\}$$

where

$$[\text{cv}(n)]^2 = \widehat{\text{var}}(n)/n^2 = 195.8/99^2 = 0.01998$$

and

$$[\mathrm{cv}\{\hat{f}(0)\}]^2 = \widehat{\mathrm{var}}\{\hat{f}(0)\}/\{\hat{f}(0)\}^2 = 0.007601^2/0.08563^2 = 0.007879$$

where $\widehat{\mathrm{var}}\{\hat{f}(0)\}$ is based on approximately $n = 99$ degrees of freedom. Then

$$\widehat{\mathrm{var}}(\hat{D}) = (88.3)^2\,[0.01998 + 0.007879]$$
$$= 217.2$$

and

$$\widehat{\mathrm{se}}(\hat{D}) = \sqrt{\widehat{\mathrm{var}}(\hat{D})}$$
$$= 14.74$$

The coefficient of variation of estimated density is $\mathrm{cv}(\hat{D}) = \widehat{\mathrm{se}}(\hat{D})/\hat{D} = 0.167$, or 16.7%, which might be adequate for many purposes. An approximate 95% confidence interval could be set in the usual way as $\hat{D} \pm 1.96 \cdot \widehat{\mathrm{se}}(\hat{D})$, resulting in the interval [59.4, 117.2]. Log-based confidence intervals (Burnham *et al.* 1987: 211–3) offer improved coverage by allowing for the asymmetric shape of the sampling distribution of $\hat{D}$ for small n. The procedure allows lower and upper 95% bounds to be computed as

$$\hat{D}_L = \hat{D}/C$$

and

$$\hat{D}_U = \hat{D} \cdot C$$

where

$$C = \exp\{1.96 \cdot \sqrt{\log_e(1 + [\mathrm{cv}(\hat{D})]^2)}\}$$

This method gives the interval [63.8, 122.2], which is wider than the symmetric interval, but is a better measure of the precision of the estimate $\hat{D} = 88.3$. The imprecision in $\hat{D}$ is primarily due to the variance component associated with n; approximately 72% (i.e. $0.01998/(0.01998 + 0.007879)$) of $\widehat{\mathrm{var}}(\hat{D})$ is due here to $\widehat{\mathrm{var}}(n)$.

The use of 1.96 in constructing the above confidence intervals is only justified if the degrees of freedom of all variance components in

$\widehat{\text{var}}(\hat{D})$ are large, say greater than 30. In this example, the degrees of freedom for the component $\widehat{\text{var}}(\hat{n})$ are only 11. If there were only one variance component, it would be standard procedure to use the t-distribution on the relevant degrees of freedom, rather than the standard normal distribution, as the basis for a confidence interval. When the relevant variance estimator is a linear combination of variances, there is a procedure using an approximating t-distribution as the basis for the confidence interval. This more complicated procedure is explained in Section 4.7.4 below, and is used automatically by program DISTANCE.

The effective strip width is estimated by $\hat{\mu} = 1/\hat{f}(0) = 11.68$ m. The unconditional probability of detecting an object in the surveyed area, $a = 2wL$, is $P_a = \hat{\mu}/w = 0.61$, which is simply the ratio of the effective strip width to the truncation distance, $w = 19$ m. These estimates are MLE as they are one-to-one transformations of the MLE of $f(0)$.

In summary, we obtain $\hat{D} = 88.3$, $\widehat{\text{se}}(\hat{D}) = 14.7$, cv = 16.7%, with a 95% confidence interval of [63.8, 122.2]. Recalling that the true parameter $D = 80$, this particular estimate is a little high, largely because the sample size ($n = 105$, untruncated) happened to be above that expected ($E(n) = 96$). This is not unusual, given the large variability in n due to spatial aggregation of the objects, and the confidence interval easily covers the parameter. Some alternative analyses and issues and their consequences will now be explored.

4.7.2 Ignoring information from replicate lines

If the Poisson assumption ($\widehat{\text{var}}(n) = n$) had been used with $w = 19$ m and $L = 48$ km, then the estimate of density would not change, but $\widehat{\text{se}}(\hat{D})$ would be underestimated at 11.84, with 95% confidence interval of [67.98, 114.70]. While this interval happens to cover D, the method underestimates the uncertainty of the estimator $\hat{D}$; if many data sets were generated, the true coverage of the interval would be well below 95%. This procedure cannot be recommended; one should estimate the variance associated with sample size empirically from the counts on the individual replicate lines, including those lines with zero counts. For example, line 11 had no observations ($n_{11} = 0$), which must be included in the analysis as a zero count.

4.7.3 Bootstrap variances and confidence intervals

The selected model for $g(x)$ for the example data was the half-normal, with $w = 19$ m, fitted to ungrouped distance data. The MLE of $f(0)$ was 0.08563 with $\widehat{\text{se}} = 0.007601$. The bootstrap procedure (Section 3.7.4) can

be used to obtain a more robust estimate of this standard error. The required number of series expansion terms can be estimated in each resample, and variance due to this estimation, ignored in the analytical method, is then a component of the bootstrap variance. As an illustration, 1000 bootstrap replications were performed, yielding an average $\hat{f}(0) = 0.08587$ with $\widehat{\text{se}} = 0.00748$. In this simple example where the true model is the fitted model without any adjustment terms, the two procedures yield nearly identical results.

A superior use of the bootstrap in line transect sampling is to sample with replacement from the replicate lines, until either the number of lines in the resample equals the number in the original data set, or the total effort in the resample approximately equals the total effort in the real data set. If the model selection procedure is automated, it can be applied to each resample, so that model misspecification bias can be incorporated in the variance estimate. Further, the density D may be estimated for each resample, so that robust standard errors and confidence intervals may be set that automatically incorporate variance in sample size (or equivalently, encounter rate) and cluster size if relevant, as well as in the estimate of $f(0)$. The method is described in Section 3.7.4, and an example of its application to point transect data is given in Section 5.7.2.

A possible analysis strategy is to carry out model selection and choice of truncation distance first, and then to evaluate bootstrap standard errors only after a particular model has been identified. Although model misspecification bias is then ignored, the bootstrap is computationally intensive, and its use at every step in the analysis will be prohibitive.

4.7.4 Satterthwaite degrees of freedom for confidence intervals

For the log-based confidence interval approach, there is a method to allow for the finite degrees of freedom of each estimated variance component in $\widehat{\text{var}}(\hat{D})$. This procedure dates from Satterthwaite (1946); a more accessible reference is Milliken and Johnson (1984). Assuming the log-based approach, $[\log_e(\hat{D}) - \log_e(D)]/\text{cv}(\hat{D})$ is well approximated by a t-distribution with degrees of freedom computed in the case of two variance components by the formula

$$df = \frac{[\text{cv}(\hat{D})]^4}{\dfrac{[\text{cv}(n)]^4}{k-1} + \dfrac{[\text{cv}\{\hat{f}(0)\}]^4}{n}}$$

where

$$[\mathrm{cv}(\hat{D})]^2 = \frac{\widehat{\mathrm{var}}(\hat{D})}{\hat{D}^2} \qquad [\mathrm{cv}(n)]^2 = \frac{\widehat{\mathrm{var}}(n)}{n^2} \qquad [\mathrm{cv}\{\hat{f}(0)\}]^2 = \frac{\widehat{\mathrm{var}}\{\hat{f}(0)\}}{\{\hat{f}(0)\}^2}$$

Given the computed degrees of freedom, one finds the two-sided $100(1 - 2\alpha)$ percentile of the t-distribution with these degrees of freedom; df is in general non-integer, but may be rounded to the nearest integer. Usually $\alpha = 0.025$, giving a 95% confidence interval. Then one uses the value of $t_{df}(0.025)$ in place of 1.96 in the confidence interval calculations, so that

$$\hat{D}_L = \hat{D}/C$$

and

$$\hat{D}_U = \hat{D} \cdot C$$

where

$$C = \exp\{t_{df}(0.025) \cdot \sqrt{\log_e(1 + [\mathrm{cv}(\hat{D})]^2)}\}$$

This lengthens the confidence interval noticeably when the number of replicate lines is small.

We illustrate this procedure with the current example for which $[\mathrm{cv}(n)]^2 = 0.01998$ on 11 degrees of freedom, and $[\mathrm{cv}\{\hat{f}(0)\}]^2 = 0.007879$ on 99 degrees of freedom. Thus $[\mathrm{cv}(\hat{D})]^2 = 0.027859$. The above formula for df gives

$$df = \frac{0.0007761}{\frac{0.0003992}{11} + \frac{0.00006208}{99}} = 21.02$$

which we round to 21 for looking up $t_{21}(0.025) = 2.08$ in tables. Using 2.08 rather than 1.96, we find that $C = 1.4117$, and the improved 95% confidence interval is [62.6, 124.7], compared with [63.8, 122.2] using $z = 1.96$. The Satterthwaite procedure is implemented in DISTANCE, so that it produces the improved interval.

This procedure for computing the degrees of freedom for an approximating t-distribution generalizes to the case of more than two components, for example when there are three parameter estimates, n, $\hat{f}(0)$ and $\hat{E}(s)$. Section 3.7.1 gives the general formula.

4.8 Estimation when the objects are in clusters

Often the objects are detected in clusters (flocks, coveys, schools) and further considerations are necessary in this case. The density of clusters (D_s), the density of individual objects (D), and average cluster size $E(s)$ are the biological parameters of interest in surveys of clustered populations, and several intermediate parameters are of statistical interest (e.g. $f(0)$). Here, we will assume that the clusters are reasonably well defined; populations that form loose aggregations of objects are more problematic.

It is assumed that the distance measurement is taken from the line to the geometric centre of the cluster. If a truncation distance w is adopted in the field (as distinct from in the analysis), then a cluster is recorded if its centre lies within distance w and a count made of **all** individuals within the cluster, including those individuals that are at distances greater than w. If the geometric centre of the cluster lies at a distance greater than w, then no measurement should be recorded, even if some individuals in the cluster are within distance w of the line. The sample size of detected objects n is the number of clusters, not the total number of individuals detected.

4.8.1 Observed cluster size independent of distance

If the size of detected clusters is independent of distance from the line (i.e. $g(x)$ does not depend on s), then estimation of D_s, D and $E(s)$ is relatively simple. The sample mean $\bar{s}$ is taken as an unbiased estimator of the average cluster size. Then $\hat{E}(s) = \bar{s} = \Sigma\, s_i/n$, where s_i is the size of the ith cluster. In general the density of clusters D_s and measures of precision are estimated exactly as given in Sections 4.3–4.7. Then, $\hat{D} = \hat{D}_s \cdot \bar{s}$; the estimator of the density of individuals is merely the product of the density of clusters times the average cluster size. Alternatively, the expression can be written

$$\hat{D} = \frac{n \cdot \hat{f}(0) \cdot \bar{s}}{2L}$$

The example data set used throughout this chapter is now reconsidered in view of the (now revealed) clustered nature of the population. The distribution of true cluster size in the population was simulated from a Poisson distribution and the size of detected clusters was independent

of distance from the line. The value of s was simulated as 1 + a Poisson variable with a mean of two or, equivalently, $(s - 1) \sim$ Poisson (2), so that $E(s) = 3$. Theoretically, $\text{var}(s) = \text{var}(s - 1) = E(s - 1) = E(s) - 1 = 2$. Under the independence assumption, the sample mean $\bar{s}$ is an unbiased estimate of $E(s)$. The true density of individuals was 240. Estimates of D_s (called D in previous sections), $f(0)$, effective strip width and the various measures of precision are exactly those derived in Section 4.7.

The estimated average cluster size, $\bar{s}$, for the example data with $w = 19$ m is 2.859 ($\bar{s} = 283/99$) and the empirical sampling variance on $n - 1 = 98$ degrees of freedom is

$$\widehat{\text{var}}(\bar{s}) = \frac{\sum_{i=1}^{n} (s_i - \bar{s})^2}{n(n-1)}$$

$$= 0.02062$$

so that

$$\widehat{\text{se}}(\bar{s}) = \sqrt{0.02062}$$

$$= 0.1436$$

These empirical estimates compare quite well with the true parameters; $\text{var}(\bar{s}) = 2/n = 2/99 = 0.0202$, $\widehat{\text{se}}(\bar{s}) = 0.142$. If one uses the knowledge that cluster sizes were based on a Poisson process, one could estimate this true standard error as $\sqrt{\{(\bar{s} - 1)/n\}} = \sqrt{(1.859/99)} = 0.137$, which is also close to the true value. The point here is that the empirical estimate is quite good and can be computed when the Poisson assumption is false.

A plot of cluster size s_i vs. distance x_i (Fig. 4.3) provides only weak evidence of dependence ($r = 0.16, p = 0.10$). In this case, we take $\hat{E}(s) = \bar{s}$, the sample mean. Thus, the density of individuals is estimated as

$$\hat{D} = \hat{D}_s \cdot \bar{s}$$

$$= 88.3 \times 2.859$$

$$= 252.4 \text{ individuals/km}^2$$

Then, for large samples,

$$\widehat{\text{var}}(\hat{D}) = \hat{D}^2 \cdot \{[\text{cv}(n)]^2 + [\text{cv}\{\hat{f}(0)\}]^2 + [\text{cv}(\bar{s})]^2\}$$
$$= 252.4^2 \cdot [0.01998 + 0.007879 + 0.002523]$$
$$= 1936$$

so that

$$\widehat{\text{se}}(\hat{D}) = \sqrt{1936} = 44.0$$

This gives $\hat{D} = 252.4$ individuals/km^2 with cv = 17.4%. The log-based 95% confidence interval, using the convenient multiplier $z = 1.96$, is [179.8, 354.3]. This interval is somewhat wide, due primarily to the spatial variation in n; $\widehat{\text{var}}(n)$ makes up 66% of $\widehat{\text{var}}(\hat{D})$, while $\widehat{\text{var}}(\hat{f}(0))$ contributes 26% and $\widehat{\text{var}}(\bar{s})$ contributes only 8% (e.g. 66% = $\{0.01998/(0.01998 + 0.007879 + 0.002523)\} \times 100$).

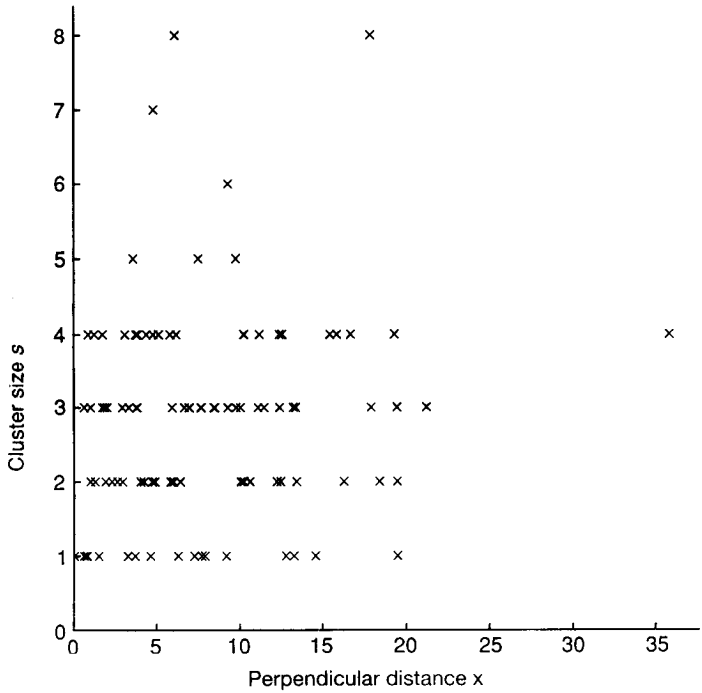

Fig. 4.3. Scatterplot of the relationship between the size of a detected cluster and the distance from the line to the geometric centre of the cluster for the example in which size and detection distance are independent. The correlation coefficient is 0.16.

A theoretically better confidence interval is one based not on the standard normal distribution (i.e. on $z = 1.96$) but rather on a t-distribution with degrees of freedom computed here as

$$df = \frac{[\text{cv}(\hat{D})]^4}{\frac{[\text{cv}(n)]^4}{k-1} + \frac{[\text{cv}\{\hat{f}(0)\}]^4}{n} + \frac{[\text{cv}(\bar{s})]^4}{n-1}}$$

$$= \frac{0.0009235}{\frac{0.0003992}{11} + \frac{0.00006208}{99} + \frac{0.000006366}{98}} = 24.97$$

which we round to 25. Given this computed value for the degrees of freedom, one finds the two-sided $100(1 - 2\alpha)$ percentile $t_{df}(\alpha)$ of the t-distribution. In this example, we find $t_{25}(0.025) = 2.06$. Using 2.06 rather than 1.96 in the log-based confidence interval procedure, $C = 1.4282$, giving an improved 95% confidence interval of [176.8, 360.5]. It is this latter interval which DISTANCE computes, applying the procedure of Satterthwaite (1946) to $\log_e(\hat{D})$.

Plots and correlations should always be examined prior to proceeding as if cluster size and detection distance were independent. In particular, some truncation of the data will often have the added benefit of weakening the dependence between s_i and x_i. If truncation is appropriate, then $\hat{E}(s)$ should be based on only those clusters within $(0, w)$. Our experience suggests that data from surveys of many clustered populations can be treated under the assumption that s_i and x_i are independent. For small clusters (e.g. coveys of quail of 5–12 or family groups of antelope of 2–4), the independence assumption is likely to be reasonable. This allows the analysis of $f(0)$, D_s, and measures of their precision to be separated from the analysis of the data on cluster size and its variability. Then, estimation of the density of individuals is fairly simple.

4.8.2 Observed cluster size dependent on distance

The analysis of survey data where the cluster size is dependent on the detection distance is more complicated because of difficulties in obtaining an unbiased estimate of $E(s)$ (Drummer and McDonald 1987; Drummer *et al.* 1990; Otto and Pollock 1990). The dependence arises because large clusters might be seen at some distance from the line (near w), while small clusters might remain undetected. This phenomenon causes an overestimation of $E(s)$ because too few small clusters are detected (i.e. they are underrepresented in the sample). Thus, $\hat{D} = \hat{D}_s \cdot \bar{s}$ is also an overestimate. Another complication is that large

clusters might be more easily detected near w than small clusters, but their size might be underestimated due to reduced detectability of individuals at long distances. This phenomenon has a counter-balancing effect on the estimates of $E(s)$ and D. The dependence of x_i on s_i is a case of size-biased sampling (Cox 1969; Patil and Ord 1976; Patil and Rao 1978; Rao and Portier 1985).

The analysis of sample data from clustered populations where a dependence exists between the distances x_i and the cluster sizes s_i can take several avenues. Some of these require that the detection function $g(x)$ is fitted unconditional on cluster size, using the robust models and model selection tools already discussed. The simplest method exploits the fact that size bias in detected clusters does not occur at distances from the line for which detection is certain. Hence, $E(s)$ may be estimated by the mean size of clusters detected within distance v, where $g(v)$ is reasonably close to one, say 0.6 or 0.8. In the second method, a cluster of size s_i at distance x_i from the line is replaced by s_i objects, each at distance x_i. Thus, the sampling unit is assumed to be the individual object rather than the cluster, and the issue of estimating true mean cluster size is side-stepped. For the third method, data are stratified by cluster size (Quinn 1979, 1985). The selected model is then fitted independently to the data in each stratum. If size bias is large or cluster size very variable, smaller truncation distances are likely to be required for strata corresponding to small clusters. The fourth method estimates cluster density D_s conventionally, as does the first. Then, given the x_i, $E(s)$ is estimated by some form of regression of s_i on the x_i (i.e. an appropriate model is identified for $E(s|x)$). This sequential procedure seems to have a great deal of flexibility. In the final approach considered here, a bivariate model of $g(x, s)$ is fitted to the data to obtain the estimates of D, D_s and $E(s)$ simultaneously. The first four approaches are illustrated in this section using program DISTANCE, and the fifth using program SIZETRAN (Drummer 1991).

The data used in this section are sampled from the same population as in earlier sections of this chapter (i.e. $L = 48$, $D_s = 80$ $f(0) = 0.079$, $E(s) = 3$ and $E(n) = 96$, so that $D = 240 = 3 \times 80$). The half-normal detection function was used, as before, but σ was allowed to be a function of cluster size:

$$\sigma(s) = \sigma_0 \left(1 + b \cdot \frac{s - E(s)}{E(s)}\right)$$

where $b = 1$ and $E(s) = 3$ in the population. Selecting $b = 1$ represents a strong size bias and corresponds to Drummer and McDonald's (1987) form with $\alpha = 1$. Cluster size in the entire population (detected or not)

was distributed as $s \sim (1 + \text{Poisson})$. Given a cluster size s, the detection distance was generated from the half-normal detection function $g(x_i|s_i)$. Because of the dependence between cluster size and detection distance, the distance data differ from those in the earlier parts of this chapter. In particular, some large clusters were detected at greater distances (e.g. one detection of 5 objects at 50.9 m). Because of the dependence on cluster size, the bivariate detection function $g(x, s)$ is not half-normal. This detection function is monotone non-increasing in x and monotone non-decreasing in s. In addition, the detected cluster sizes do not represent a random sample from the population of cluster sizes, as small clusters tend to remain undetected except at short distances. Thus, the size of **detected** clusters is not any simple function of a Poisson variate. Generation of these data is a technical matter and is treated in Section 6.7.2.

A histogram of the distance data indicates little heaping and a somewhat long right tail (Fig. 4.4). Truncation at 20 m seemed reasonable and eliminated only 16 observations, leaving $n = 89$ (15% truncation). Truncation makes modelling of the detection function easier and always reduces, at least theoretically, the correlation between detection distance and cluster size. Three robust models were chosen as candidates for modelling $g(x)$: uniform + cosine, half-normal + Hermite, and hazard-rate + cosine. All three models fit the truncated data well. AIC suggested the use of the uniform + cosine model by a small margin (506.26 vs. 506.96 for the half-normal), and both models gave very similar estimates of density. The hazard-rate model (AIC = 509.209) provided rather high estimates of density with less precision, although confidence intervals easily covered the true parameter. The uniform + cosine model and the half-normal model both required only a single parameter to be estimated from the data, while the hazard-rate has two parameters. This may account for some of the increased standard error of the hazard-rate model, but the main reason for the high estimate and standard error is that the hazard-rate model attempts to fit the spike in the histogram of Fig. 4.4 in the first distance category. Because we know the true model in this case, we know the spike is spurious, and arises because for this data set, more simulated values occurred within 2.5 m than would be expected. Generally, if such a spike is real, the hazard-rate model yields lower bias (but also higher variance) than most series expansion models, whereas its performance is poor if the spike is spurious. Since AIC selected the uniform + cosine model, we use it below to illustrate methods of analysis of the example data.

The uniform + cosine model for the untruncated data required five cosine terms to fit the right tail of the data adequately (Fig. 4.4). Failure to truncate the data here would have resulted in lower precision, the model would have required five cosine terms instead of just one

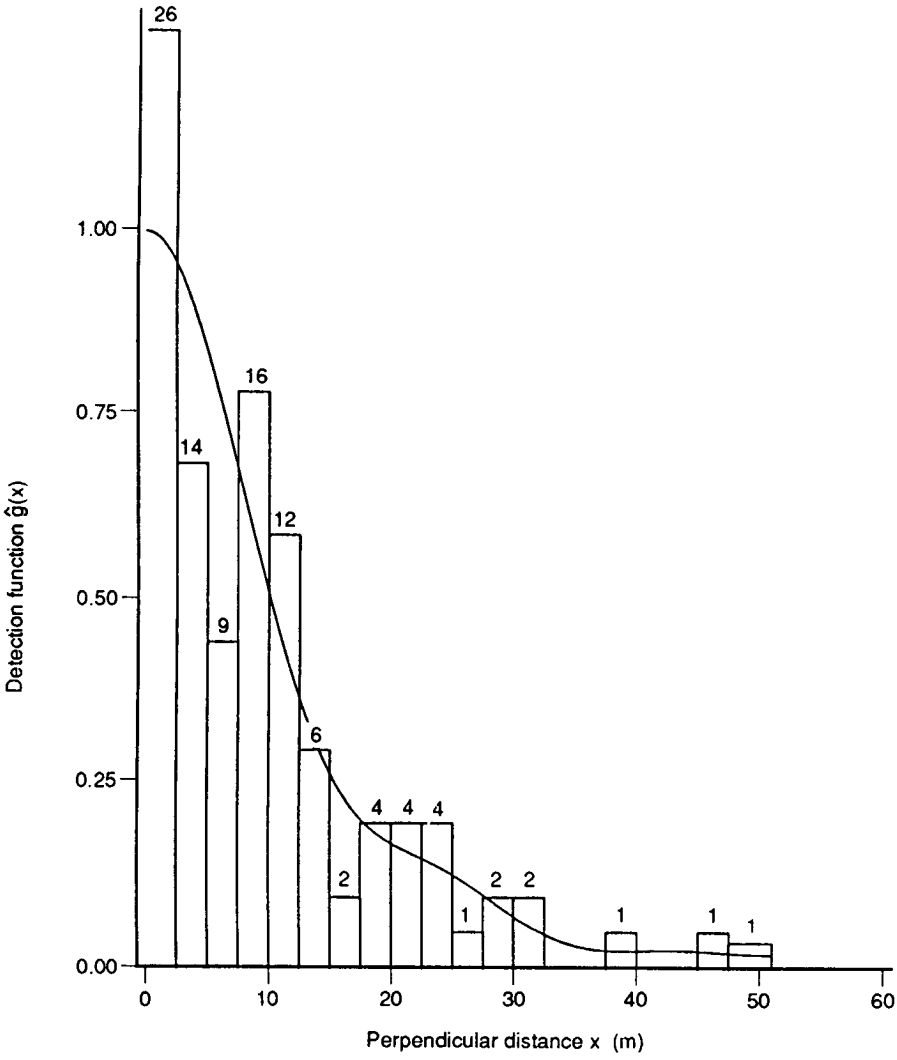

Fig. 4.4. Histogram of the example data using 20 distance categories for the case where cluster size and detection distance are dependent. The fit of a uniform + 5-term cosine model is shown.

(Fig. 4.5), and the mean cluster size would have been less reliably estimated (below).

The uniform key function and a 1-term cosine model fit the truncated data well (Fig. 4.5, $0.38 \leq p \leq 0.71$, depending on the grouping used for the χ^2 test). The estimated density of clusters was 81.56 ($\widehat{se} = 12.43$),

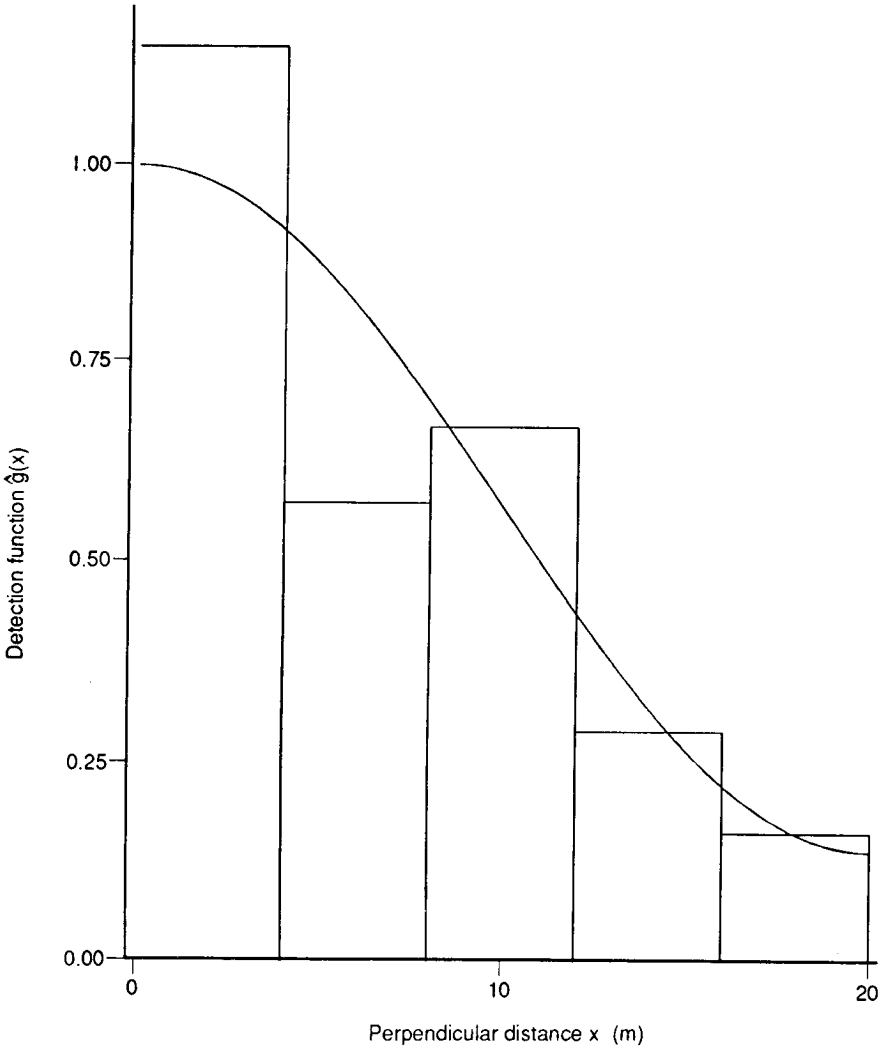

Fig. 4.5. Histogram of the example data using five distance categories and truncation at $w = 20$ m for the case where cluster size and detection distance are dependent. The fit of a uniform + 1-term cosine model is shown.

quite close to the true value (80). The mean cluster size from the sample data was 3.258 ($\widehat{se} = 0.134$) which is surely too high in view of the size-biased sampling caused by the correlation between cluster size and detection distance. However, truncation at $w = 20$ reduced this correlation from 0.485 to 0.224 so this uncorrected estimate of $E(s)$ may not

be heavily biased. Multiplying the density of clusters by the uncorrected estimate of mean cluster size (3.258), the density of individuals is estimated as 265.7 ($\widehat{\text{se}} = 41.9$; 95% CI = [195.4, 361.4]), which is a little high, but still a quite acceptable estimate for these data, for which actual density was 240 individuals/km^2.

(a) *Truncation* Observed mean cluster sizes and standard errors for various truncation distances are shown in Table 4.4. The detection function $g(x)$ was estimated using truncation distance w and mean cluster size was estimated after truncating data at distance v, $v \leq w$. The gain in precision by reducing the truncation distance from 51 m to 20 m arises because predominantly large clusters are truncated, and variability in the size of remaining clusters is reduced. It is clear from Table 4.4 that 20 m is too large a truncation distance for unbiased estimation of mean cluster size, since mean cluster size continues to fall when the truncation distance is reduced further. The choice of truncation distance is a compromise between reducing bias and retaining adequate precision. Here, mean cluster size appears to stabilize at a truncation distance of 10.5 m, for which $\hat{g}(10.5) = 0.5$. Thus, mean cluster size is estimated to be 3.116 with $\widehat{\text{se}} = 0.152$. Replacing the estimates $\bar{s} = 3.258$ and $\widehat{\text{se}} = 0.134$ by these values, density is estimated as 254.1 individuals/km^2, with $\widehat{\text{se}} = 40.7$ and approximate 95% confidence interval [186.1, 347.0] (based on $z = 1.96$ rather than Satterthwaite's correction). This estimate is closer to the true value of 240 individuals/km^2, and precision is almost unaffected, because the contribution to the overall variance due to variation in cluster size is slight.

Table 4.4 Observed mean cluster sizes and standard errors for various truncation distances v. Probability of detection $\hat{g}(v)$ at the truncation distance v for cluster size estimation was estimated from a uniform + 1-term cosine model with $w = 20$ m (Fig. 4.5) for $v \leq 20$ m, and from a uniform + 5-term cosine model with $w = 51$ m for $v = 51$ m

Truncation distance v(m)	n	$\bar{s}$	$\widehat{\text{se}}(\bar{s})$	$\hat{g}(v)$
51.0	105	3.581	0.150	0.02
20.0	89	3.258	0.134	0.14
13.6	80	3.188	0.144	0.30
10.5	69	3.116	0.152	0.50
9.1	61	3.098	0.166	0.60
7.6	49	3.061	0.192	0.70
6.0	44	3.114	0.206	0.80
4.1	37	3.081	0.214	0.90

Hence if sample size is large, one may select a truncation point for the estimation of $E(s)$ that is smaller than the truncation point for the estimation of $f(0)$ to reduce the size bias in $\hat{E}(s)$. For example, $E(s)$ might be estimated by the mean size of clusters detected within a distance x of the line, where $g(x) = 0.6$. Often bias reduction is more important than precision in estimating mean cluster size because its relative contribution to $\text{var}(\hat{D})$ may be small, as in this example.

(b) *Replacement of clusters by individuals* If a cluster of size s_i is replaced by s_i objects at the same distance, the assumption that detections are independent is violated. This compromises analytic variance estimates and model selection procedures. The first difficulty may be overcome by using robust methods for variance estimation, but model selection is more problematic. If likelihood ratio tests are used to determine the number of terms, too many terms are fitted on average, since heaping in the data at distances where large clusters were recorded yield significant departures from a smooth detection function when observations are assumed to be independent. The effect may be reduced by imposing a monotonicity constraint (Section 3.4.5). Another option is to select a model taking clusters as the sampling unit, then refit the model (with the same series terms, if any) to the data with object as the sampling unit. Neither of these is entirely satisfactory. If both strategies are adopted in the same analysis, so that a uniform + 1-term cosine model is fitted to the distance data truncated at 20 m, the following estimates are obtained. Number of objects detected, $n = 290$. Estimated density, $\hat{D} = 255.6$ objects/km^2, with analytic $\widehat{\text{se}} = 38.3$ and 95% confidence interval [184.7, 353.8]. These estimates are very close to those obtained assuming cluster size is independent of distance, although the point estimate is rather closer to the true density of 240 objects/km^2. Average cluster size can be estimated by the ratio of estimated object density (255.6) to estimated cluster density (81.56), giving 3.134. The precision of this estimate could be quantified using the bootstrap. In each bootstrap resample, both densities, and hence their ratio, would be estimated, and a variance and confidence interval obtained as described in Section 4.7.3.

This procedure cannot generally be recommended. However, it may be useful if the population being sampled occurs in loose aggregations, rather than tight, easily defined clusters. The distance to each individual object should ideally be measured in this case, although it may be sufficient to record positions and sizes of smaller groups within a cluster. The method will often perform poorly unless sample size is fairly large.

(c) *Stratification* Choice of number of strata is determined largely by sample size. The more strata, the greater the reduction in size bias, but an adequate sample size for estimating $f(0)$ is required in each stratum (perhaps at least 20–30 per stratum). Defining two strata, corresponding to cluster sizes 1–3 and ≥ 4, sample sizes before truncation are 52 and 53 respectively. If four strata are defined, for cluster sizes 1–2, 3, 4 and ≥ 5, sample sizes before truncation are 29, 23, 26 and 27. The data were analysed for both choices of stratification.

Table 4.5 Summary of results for different stratification options. Model was uniform with cosine adjustments; distance data were truncated at $w = 20$ m, except for the stratum comprising clusters of size 5–9, for which $w = 35$ m

Cluster sizes	Sample size after truncation	Effective strip width (m)	$\hat{D}$	$\widehat{\text{se}}(\hat{D})$	95% CI for D
All	89	11.4	265.7	41.9	(195.4, 361.4)
1–3	51	11.0	96.9	17.6	
4–9	38	12.1	147.4	35.0	
All			244.3	39.2	(178.8, 333.8)
1–2	29	10.0	51.2	16.0	
3	22	13.7	50.1	16.3	
4	22	11.7	78.2	21.0	
5–9	24	22.6	53.3	18.0	
All			232.8	35.8	(172.5, 314.2)

Results are summarized in Table 4.5. In this case, no precision is lost by stratification, despite the small samples from which $f(0)$ was estimated, and the estimated densities were closer to the true value of 240 objects//km^2 than for the case without stratification. In our experience, loss of precision arising from stratification by cluster size is seldom large, provided sample size in each stratum does not fall below 20, and the method is a simple way of reducing the effects of size-biased sampling. Mean cluster size may be estimated by a weighted average of the mean size per stratum, with weights equal to the estimated density of clusters by stratum. Alternatively, $E(s)$ may be estimated as overall $\hat{D}$ from the stratified analysis divided by $\hat{D}_s$ from the unstratified analysis. For two strata, this yields $\hat{E}(s) = 244.3/81.56 = 2.995$, and for four strata, $\hat{E}(s) = 232.8/81.56 = 2.854$. Both estimates are close to the true mean cluster size of 3.0. The reader is referred to Drummer (1985) and Quinn (1985) for further information on stratification.

(d) *Regression estimator* The procedure we recommend in most cases is a regression of s_i or $\log_e(s_i)$ on $\hat{g}(x_i)$ (Section 3.6.3). This allows an

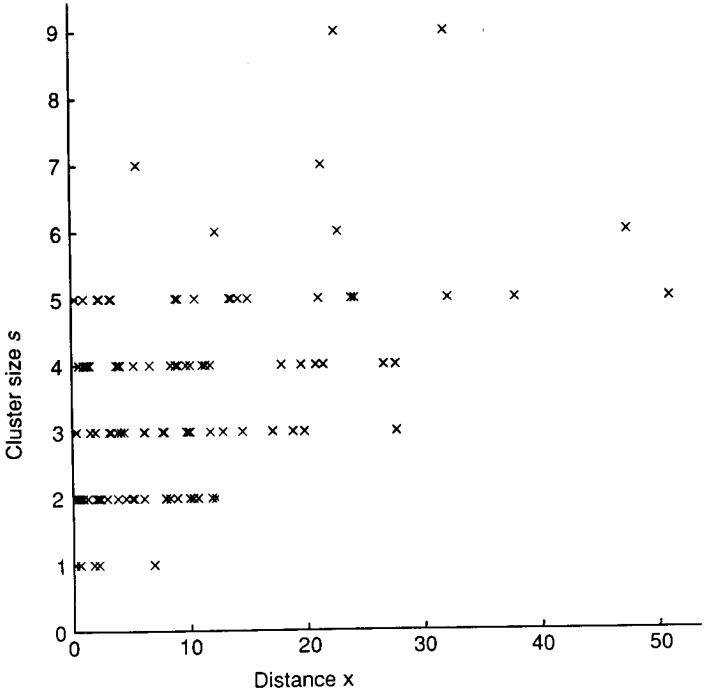

Fig. 4.6. Scatterplot of the relationship between the size of a detected cluster and the distance from the line to the geometric centre of the cluster for the example in which probability of detection is a function of cluster size. The correlation coefficient is 0.485 ($w = \infty$).

estimate of $E(s)$ at the point where $\hat{g}(x_i) = 1$; that is, the point at which detectability is certain, where size bias should not occur. Proper truncation of the distance data should be considered prior to the regression analysis (e.g. $g(x) \doteq 0.15$). Applying this method to the example, with dependent variable $\log_e(s_i)$, yields $\hat{E}(s) = 2.930$ and $\widehat{\text{se}}\{\hat{E}(s)\} = \sqrt{[\widehat{\text{var}}\{\hat{E}(s)\}]} = 0.139$, which is close to the true mean cluster size of 3.0 with good precision. The corresponding density estimate is 239.0 individuals/km^2, with $\widehat{\text{se}} = 38.1$, cv = 16.0% and 95% confidence interval [171.3, 333.5]. The regression approach allows $f(0)$ to be estimated using all the robust theory available and then treats the estimation of mean cluster size as a separate problem. The analyst has good control over the estimation under this procedure. A scatter plot of cluster size against distance or estimated detection probability can be used to investigate the form of the relationship, although the scatter can be wide (Fig. 4.6).

The regression estimate of $E(s)$ reduces the bias but some precision may be lost in correcting for the size-biased sampling of cluster size.

Often, $\text{var}(\bar{s})$ is a small component of $\text{var}(\hat{D})$, so that little precision is lost by applying an adjustment for size bias. Use of $\bar{s}$ as an estimate of $E(s)$ is not recommended if dependence between cluster size and detection distance is suspected.

(e) *Bivariate models for the detection function* This methodology, due to Drummer and McDonald (1987), relies on several parametric models, each incorporating a parameter (α) to reflect the size bias in the sample data. The data are transformed to x/s^{α} and the bivariate detection function is expressed as $g(x, s) = g(x/s^{\alpha})$; $\alpha = 0$ represents the special case where cluster size and detection distance are independent and no size bias exists (Otto and Pollock 1990). Program SIZETRAN was used for the analysis of the example data and four models were considered: the negative exponential, the half-normal, the reversed logistic, and the generalized exponential (Drummer and McDonald 1987, Table 1). These models incorporate the size bias by modelling the scale parameter as a simple increasing power function of observed cluster size, s^{α}. Under this approach, truncation is not an option with current software and separate consideration of $\hat{E}(s)$ is not necessary.

The choice of the best model might be guided by AIC and this leads to the half-normal model (Table 4.6). The generalized exponential model failed to converge and cannot be considered. The estimates under the half-normal model are quite good (Table 4.6), but the precision is poorer than under the sequential approach using regression (cv = 21.3% instead of 16.0%). This model fits the data well as judged by χ^2 goodness of fit tests ($p > 0.15$). The data were simulated from this model with $\alpha = 1$, so that it would be expected to fit well. Estimation under the other models seems less satisfactory; the estimated density of individuals is too high under both the negative exponential and the reversed logistic (Table 4.6). Still, the confidence intervals cover the true density, partially because the estimated standard errors are so large. In each case, there is clear evidence of a size-biased sample (i.e. α is significantly greater than zero, Table 4.6).

Table 4.6 Summary of results for three models allowing for dependence between cluster size and detection distance (standard errors). The parameter α is incorporated into these models to account for size-biased sampling (Drummer and McDonald 1987). Data are untruncated ($w = \infty$)

Model	AIC	$\hat{\alpha}$		$\hat{D}_s$		$\hat{D}$	
Negative exponential	674.4	1.113	(0.214)	140.7	(41.4)	397.6	(114.6)
Half-normal	668.9	1.092	(0.150)	87.1	(19.3)	247.5	(52.9)
Reversed logistic	671.1	1.076	(0.171)	95.2	(27.8)	271.4	(77.5)

If the size-biased nature of the data had been ignored (i.e. $\alpha = 0$), the point estimates would have been less satisfactory. In general, the point estimates of the density of individuals would have been too high (e.g. $\hat{D} = 385.0$ and 384.6 for the negative exponential and reversed logistic, respectively). In this example, AIC selected a good model and the point estimates were quite satisfactory. Again, this illustrates the importance of meeting the key assumptions and obtaining an adequate sample of quality data.

The bivariate approach is interesting and appealing from a statistical viewpoint, but precision seems poorer than for the sequential regression approach. Software development is required to address convergence problems, and to allow the user to specify a finite truncation width. Further study is needed to investigate the robustness of the approach. The method of Quang (1991), using Fourier series, may partially address this aspect.

4.9 Assumptions

Assumptions of line transect sampling are covered in detail in Section 2.1, and further discussion is given in Chapter 7. We outline a few issues here, to counter some of the more common misconceptions about what is assumed in deriving density or abundance estimates.

Most model selection procedures and some variance estimation procedures assume that objects are randomly and independently distributed throughout the study area. Provided lines are randomly located, or a systematic grid of lines is randomly positioned in the study area, the assumption is not required. If object density is highly variable, or dependence between detections is strong, then the possible effect on model selection should be borne in mind, and robust variance methods should be adopted, with care taken to ensure that the correct sampling unit is selected; whether a detection associated with one sampling unit is made should be largely independent of detections made in other sampling units. Often, all effort carried out by a single observer in a single session comprises a suitable sampling unit. Detections made within the unit might be highly dependent (e.g. if one bird calls in response to the calls of another, both might be detected by the observer, or encounter rate might be abnormally high on one leg of a marine mammal survey because of exceptional sighting conditions). Between units, dependence should be slight. If random lines are used, the appropriate sampling unit is the line, and all data associated with it. If the design comprises a systematic grid of lines, use of lines as sampling units should again prove satisfactory. These issues are discussed further in

Section 3.7. The most extreme departure from independent detections is when objects occur in clusters. Strategies for this case are outlined in Section 4.8.

Estimation of $g(0)$ for line transect surveys is considered in Chapter 6. There is currently no approach to this problem that is wholly satisfactory, so whenever possible, surveys should be designed to ensure that $g(0) = 1$. The solution of guarding the centreline can be counter-productive if this gives rise to two detection processes. If one process generates detections at large distances, but is such that $g(0)$ is appreciably less than one, and the other generates detections only at small distances, then the composite detection function will be impossible to model adequately. If such a field strategy is adopted, data from the two processes should be recorded and analysed separately, although problems are likely to remain. If it is suspected that $g(0)$ is less than one, methods that might increase it include using more observers to cover the line, travelling more slowly along the line, using only experienced observers, improving the training of observers, and upgrading optical aids. For terrestrial surveys in which animals are flushed, trained dogs can be an effective aid, allowing a wider area to be efficiently searched.

Random movement of objects before detection generates positive bias in estimates of object density. Hiby (1986) showed that bias is small provided that object movement is slow relative to that of the observer (up to around a third of the observer's speed). A strategy for line transect analysis of fast moving objects is outlined in Section 7.6. Movement in response to the observer is problematic, and is discussed in Section 2.1. From a practical viewpoint, field procedures should be developed that ensure that most detections occur at distances for which responsive movement is unlikely to have occurred. In other words, the observer should strive to detect the object before the object is able to move far from its initial position in response to the observer's presence. If this is not possible, the methods of Turnock and Quinn (1991) or Buckland and Turnock (1992), which use ancillary data to adjust for the effect of movement, might be attempted. The latter method is described in Section 6.4.

If the distance data appear to have a distinct mode away from the origin, the analysis is problematic. This might happen by chance, as a result of heaping, or through the presence of evasive movement prior to detection. Some robust models will attempt to fit the data near the origin, so that the mode of the density function is to the right of the origin. In these cases, it is often prudent to constrain the estimated detection function to be monotone non-increasing in an attempt to minimize bias. A weak constraint is to impose the condition $\hat{g}(x) \leq g(0) = 1$ for all $x > 0$. This condition is often sufficient to achieve

a monotone non-increasing function and satisfactory estimates of $f(0)$. The alternative is a constraint forcing a strict non-increasing function such that $\hat{g}(x_i) \geq \hat{g}(x_{i+1})$, where $x_i < x_{i+1}$. The default option of DISTANCE imposes the strong constraint. To reduce the computational cost of applying the constraint, the estimated density is evaluated at just ten points, and the fit is modified if the constraint is not satisfied for all successive pairs of these points. The user may instead select the weak constraint, or override both constraints. DISTANCE warns the user if a constraint has caused the model fit to be modified. In this case, the bootstrap estimate of $\text{var}(\hat{f}(0))$ is recommended, since the assumptions on which the analytic variance is based are violated.

Consistent bias in distance estimation should be avoided. If distances are overestimated by 10%, densities are underestimated by 9%; if they are underestimated by 10%, densities are overestimated by 11%. If on the other hand distance estimation is unbiased on average, measurement errors must be large to be problematic. For marine surveys, reticles or graticules (Section 7.4.2) are almost essential for accurate distance estimation. In terrestrial surveys, distances can often be measured, and if this is not practical, good range finders can be effective up to around 300 m. Distance categories can be accurately determined in aerial surveys by lining up markers on the windows with markers or streamers on the wing struts, although the height of the aircraft must be accurately measured and constant. In hilly terrain, perpendicular distances from the aircraft must be determined by other means.

4.10 Summary

Data analysis is relatively easy if the survey is well designed and the data properly collected. The analysis of small samples, especially where some assumptions have been violated, is more problematic. The analysis of 'good' data, such as here, is relatively easy using available software. Adequate analysis cannot be carried out without specialist software.

An objective strategy must be followed, such as that outlined in this chapter and Section 2.5. The data must be checked for recording or data-entry errors. Plotting the distance data as histograms will often reveal anomalies that must be further considered. Truncation of some observations in the right tail of the distance data should always receive consideration. Several candidate, robust models should be considered. The use of AIC and other criteria are helpful in selecting the best model, or a small subset of good models, for final analysis and inference. Once a model is selected, MLE is used to obtain parameter estimates and measures of their precision. With good data (adequate sample size and

validity of the key assumptions), inference using two or three good, robust models is likely to yield similar estimates. This is reassuring because the methods to select the best model are subject to uncertainty.

If objects on the centreline are missed, $E(\hat{D})$ will be too low. If 20% of objects on the centreline are missed, the density estimate can be expected to show a negative bias of around 20%. Movement prior to detection is also problematic. Measurement errors, especially near the centreline, are more difficult to treat. If measurement errors are random, then the sampling variance may be somewhat inflated, but bias may be small. Systematic measurement errors invariably generate bias and should be avoided. Valid inference depends on field design and attention to the assumptions. While analysis procedures are robust to some types of assumption failure, there is no substitute for quality data taken carefully under the assumptions. Searching should be conducted such that the distance data have a broad shoulder. The presence of a shoulder makes model selection less important and improves the quality of inference. The reader is urged to study the material in Chapter 7 prior to the conduct of a survey involving distance sampling.

These strategies for analysis carry over to more complicated surveys involving stratification, surveys repeated in time using the same lines, multiple observers, aerial or underwater platforms, or samples of very large areas. Some of these issues are illustrated in Chapter 8 (and by Burnham *et al.* 1980: 41–55), and specialized theory is extended in Chapter 6.

Surveys of clustered populations require additional care in counting the number of individuals in each cluster detected and addressing the possible size-biased aspects of such sampling. Plotting the cluster sizes s_i against the x_i distances is always recommended. Our experience suggests that size bias is often a minor issue if cluster size is not too variable; proper truncation of perpendicular distance data can often allow simple models to provide valid inference concerning the density of clusters and individuals. However, if the largest cluster is, say, more than five times the size of the smallest, correction for possible size bias should be investigated. When cluster size is highly variable (e.g. from one or two individuals to many thousands, as in some species of marine mammals), then very careful modelling and analysis of the data is required.

Populations in large, loose aggregations, scattered around the sample area, are problematic. Theory and software are readily available for the analysis of sample data from populations of individuals randomly distributed in space, and the same is true of populations distributed under some regular stochastic process that generates some degree of spatial aggregation, by computing var(n) empirically. Good theory and software

now exist for the analysis of populations that are clustered in definable clusters where the cluster size is not too variable. Difficulties arise when populations are spatially distributed in loose clusters whose boundaries, and therefore size, must be determined subjectively. This situation is in need of additional research, but bootstrap methods may play an important role in the analysis of such data. If at all possible, the location of each individual object should be recorded in this circumstance, so that the method of Section 4.8.2(b) can be applied, but the cluster to which each individual belongs should also be noted, to allow comparative analyses of clusters. Populations in large or highly variable groups require great care in estimating $E(s)$ in ways that minimize or avoid bias. Estimation of average cluster size must receive special emphasis in the design of the survey and the pilot study (e.g. temporarily leaving the planned centreline in aerial surveys of cetaceans to count individuals more accurately).

The following is intended as a crude checklist of the stages required to carry out a full analysis of line transect data. Not all steps are necessarily required in any given analysis, especially if similar data sets have been analysed previously.

1. Key in and validate the data. The data should not be aggregated in any way prior to entry. Thus if distances are ungrouped, they should not be entered as grouped data, even if they are subsequently grouped for analysis. Distances should be entered by line, so that individual lines can be defined as the sampling units. For stratified designs, these lines should be allocated to their strata.
2. Plot histograms of the perpendicular distance data, using different choices for the cutpoints, and fit a preliminary model to the data. Examine the histograms for evidence of failure of assumptions. If data are ungrouped, assess whether they should be grouped before analysis, selecting group cutpoints to reduce the effect of heaping, or to alleviate the effects of a spurious spike in the data at zero distance (Section 4.5). If data are grouped, assess whether any groups should be amalgamated.
3. Identify a truncation point w for perpendicular distances, preferably such that $\hat{g}(w) \doteq 0.15$, although truncation of roughly 5% of observations is often satisfactory (Section 4.3). Assess from the histograms whether this truncation distance is reasonable; if not, select one or more alternatives. Try fitting a few models, possibly with different grouping options or different truncation points.
4. Where relevant, select an appropriate truncation point w and an appropriate choice of grouping (if any). Fit several models that satisfy the model robustness, shape and estimator efficiency criteria. We recommend some or all combinations of a half-normal, uniform

or hazard-rate key with simple or Hermite polynomial or cosine adjustments. Select a single model, for example using Akaike's Information Criterion, and assess its adequacy using goodness of fit (Section 4.6). If the fit is poor, investigate the reasons, and evaluate possible solutions. Assess the sensitivity of estimation to the model selected; if sensitivity is high (e.g. the detection curve is excessively spiked under one or more models), examine whether the estimates from the selected model should be replaced or supplemented by those from other models that yield adequate fits.

5. If the detections are of clusters of objects, assess whether there is evidence of size bias, and if necessary, try one or more of the methods of Section 4.8 to correct for it.
6. Having identified a model for the perpendicular distance data, review the options for variance estimation, for stratifying some or all components of estimation, and for including covariates. Select options that are likely to reduce bias; of the options remaining, select those that yield the most efficient estimation. Fit the data using the preferred model(s) and options.