5

Point transects

5.1 Introduction

Songbird surveys often utilize point transects rather than line transects for several reasons. Once the observer is at the point, he or she can concentrate solely on detecting, locating and identifying birds, without the need to traverse what may be difficult terrain; he or she can take the easiest route into and away from the point, whereas good line transect practice dictates that the observer follows routes determined in advance and according to a randomized design. Further, patchy habitats can be sampled more easily by point transects. Frequently, density estimates are required for each habitat type, or estimation is stratified by habitat to improve precision. Designing a point transect survey so that each habitat type is represented in the desired proportions is easier than for line transects, and describing the vegetation structure associated with a point is also easier than for a line. If line transects are used in patchy habitats, either each line traverses several habitat types, and data must be recorded separately for each section of line within a single habitat type, or the design comprises many short lines, so that end effects (e.g. objects detected behind the observer when he or she first starts a transect, or objects detected by the observer as he or she approaches the end of a transect, which are beyond the area to be surveyed) become problematic. Other advantages of point transects are that known distances from the points may be flagged, to aid distance estimation, and only the observer-to-object distance is required, which is easier to estimate than the perpendicular distance required in line transect sampling if the observer is far from that part of the line closest to the object.

At the time of writing, point transect sampling seems to be restricted to bird surveys, although the theory also applies to the cue count and trapping web methods described in Chapter 6. The disadvantages of point transect sampling that make it unsuitable for many purposes

include the following. Objects may be disturbed or flushed by an observer approaching the point. It is difficult to determine which of these would have been detected from the point, but if they are ignored, density will be underestimated. The observer may detect many objects and waste much time while travelling between points; for line transects, a higher proportion of time in the field is spent surveying, and a higher proportion of detections is made while surveying. Thus, point transects may be inefficient for objects that occur at low densities.

This chapter illustrates point transect analysis through simulated data sets for which the parameters are known. As in Chapter 4, a simple data set 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. Having selected an appropriate model, estimation of density and measures of precision are discussed. In a final example, the objects are assumed to occur in clusters (e.g. family parties or flocks).

5.2 Example data

The example data were generated from a half-normal detection function, $g(r) = \exp(-r^2/2\sigma^2)$, $0 \leq r < \infty$ with $\sigma = 10$ m. There were $k = 30$ points, and the number of sightings per point followed a Poisson distribution with parameter $E(n_i) = 5$, $i = 1, \ldots, k$. Each sighting is of a single object. Thus $E(n) = 5 \times 30 = 150$, $h(0) = 1/\sigma^2 = 0.01$, and true density is

$$D = \frac{E(n) \cdot h(0)}{2\pi k} = 0.00796 \text{ objects/m}^2 = 79.6 \text{ objects/ha}$$

Untruncated data generated from this model, together with the fitted half-normal model, are shown in Fig. 5.1. Data truncated at 20 m and the corresponding fit of the half-normal are shown in Fig. 5.2. For comparison, the fit of the uniform + one term simple polynomial detection function is shown in Fig. 5.3.

The histograms of Figs 5.1 and 5.2 illustrate two methods of presenting point transect data. In Figs 5.1b and 5.2b, the frequency of distances is shown by distance interval, as for a conventional histogram. The curve is the fitted probability density function of recorded distances, with scale chosen to match that of the frequency data. The parameter $h(0)$ is estimated by the slope of this curve at distance $r = 0$. At small distances, the function increases because area surveyed at a given distance increases with distance from the point. For example, the area

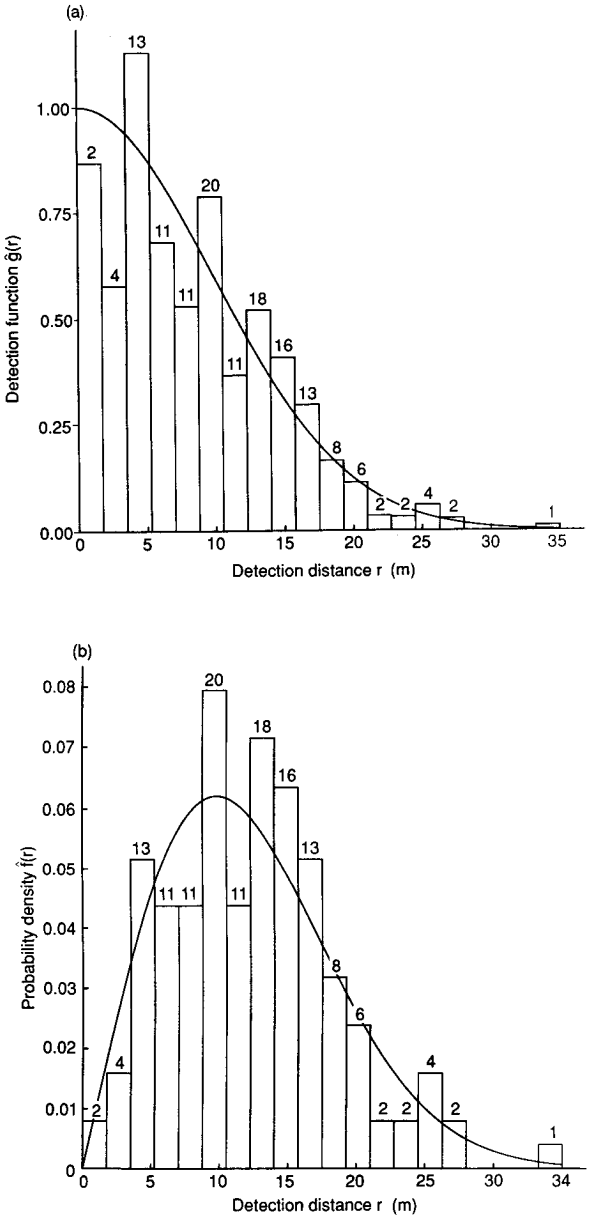

Fig. 5.1. Histograms of the example data using 20 distance categories. The fit of the half-normal detection function to untruncated data is shown in (a), in which frequencies are divided by detection distance, and the corresponding density function is shown in (b).

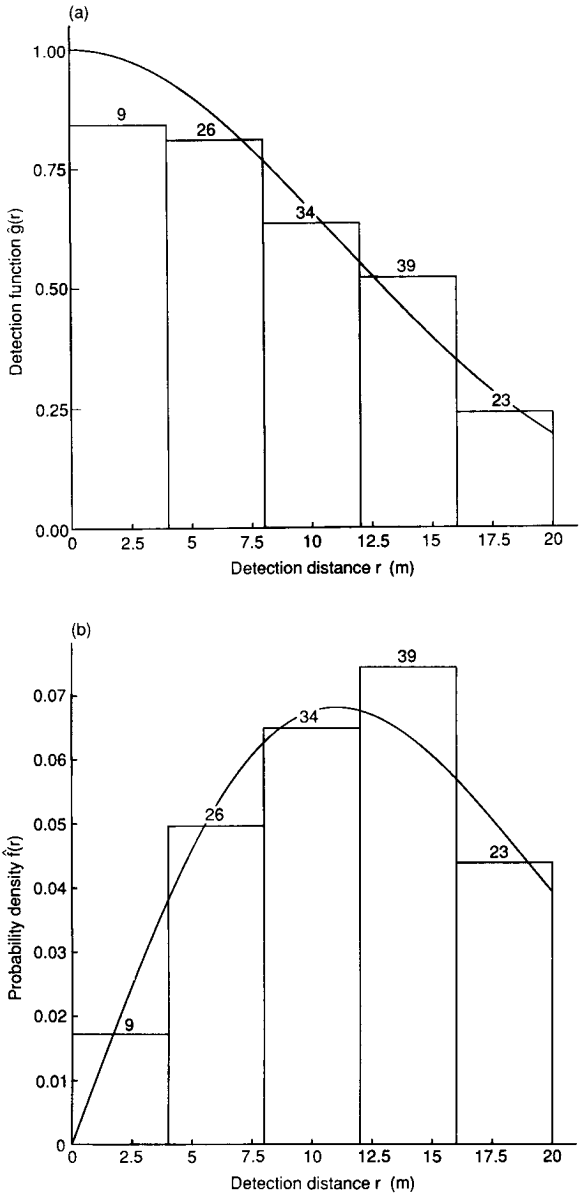

Fig. 5.2. Histograms of the example data, truncated at 20 m, using five distance categories. The half-normal model, fitted to the ungrouped data, is shown and was used for final analysis of these data. The fitted detection function is shown in (a), in which frequencies are divided by detection distance, and the corresponding density function is shown in (b).

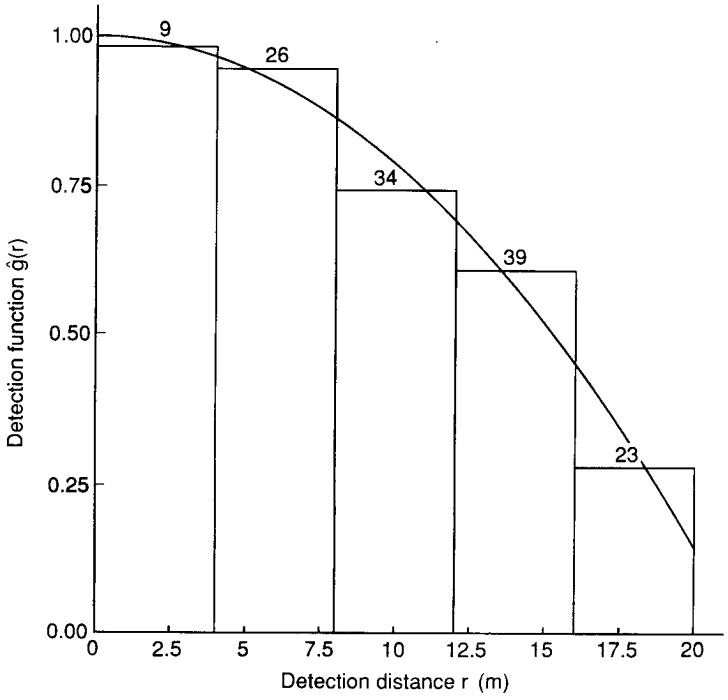

Fig. 5.3. The fit of the detection function using a uniform key with a single simple polynomial adjustment term to the example data, ungrouped and truncated at $w = 20$ m.

surveyed between r and $r + \delta$, where δ is small, is approximately $2\pi r\delta$, whereas the area between $2r$ and $2r + \delta$ is roughly $4\pi r\delta$. To correct for this increase in area, the ith distance r_i may be assigned a weight $1/r_i$. For each distance interval, these weights are summed across those observations falling within the interval. The sums are the 'corrected frequencies' of Figs 5.1a and 5.2a. To guard against infinite weights, program DISTANCE assigns weights to any zero distances equal to the weight for the smallest non-zero distance. If data are in frequency form, DISTANCE approximates the weights by the reciprocals of the mid-points of the groups. The fitted detection function, plotted so that its scale corresponds to that of the data, is also given in Figs 5.1a and 5.2a. Note that the detection function may sometimes appear to fit badly at small distances, as in Fig. 5.2; this is not a programming error, but arises because of the deceptive nature of point transect data. Relatively few distances are recorded close to the point, where area surveyed is small, so the fit of the model is not heavily influenced by distances close

to zero, whereas the height of the first histogram bar is dominated by small distances. The corresponding density is plotted with untransformed frequencies, and should appear to fit the data well, as in Fig. 5.2b, provided heaping is not severe, adequate truncation is carried out, and an appropriate model is selected.

5.3 Truncation

The largest detection distance in the example data was 34.16 m, considerably greater than the second largest of 26.87 m (Fig. 5.1). Unless the true detection function is somehow known (as it is for this simulated example), large distances can prove difficult to model, and the extra terms required increase the variance in $\hat{h}(0)$. If the uniform + simple polynomial model is fitted to the untruncated data of Fig. 5.1, four polynomial terms are required, and a less plausible shape for the detection function is obtained than for the single term fit of Fig. 5.3. In Chapter 4, we suggested as rules of thumb that either roughly 5% of observations be truncated or truncation distance w be chosen such that $\hat{g}(w) \doteq 0.15$. These rules do not carry across to point transects, for which

Table 5.1. Summary of AIC values for two truncation values (w) for the example data analysed as ungrouped 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). For each analysis, the model with the smallest AIC is indicated by an asterisk

Data type	Model (key + adjustment)	w = 20 m			w = largest obsn		
		No. of parameters			No. of parameters		
		Key	Adjust.	AIC	Key	Adjust.	AIC
Ungrouped	Uniform + cosine	0	1	765.51	0	2	918.46*
	Uniform + polynomial	0	1	764.48	0	4	922.32
	Half-normal + Hermite	1	0	764.31*	1	0	919.16
	Hazard-rate + cosine	2	0	767.22	2	1	919.79
Grouped (5 equal)	Uniform + cosine	0	1	403.36	0	2	374.06*
	Uniform + polynomial	0	1	400.83*	0	4	377.78
	Half-normal + Hermite	1	0	401.97	1	1	374.22
	Hazard-rate + cosine	2	0	403.52	2	1	375.91
Grouped (20 equal)	Uniform + cosine	0	1	768.16	0	2	764.20*
	Uniform + polynomial	0	1	766.90	0	2	830.04
	Half-normal + Hermite	1	0	766.85*	1	0	764.34
	Hazard-rate + cosine	2	1	769.94	2	1	765.25
Grouped (5 unequal)	Uniform + cosine	0	1	426.28	0	2	468.50
	Uniform + polynomial	0	1	426.69	0	4	476.91
	Half-normal + Hermite	1	0	425.93*	1	0	467.17*
	Hazard-rate + cosine	2	0	428.04	2	0	472.39

a higher proportion of detections occurs in the tail of the detection function. This can be seen in Fig. 5.1; Fig. 5.1a shows that probability of detection is as small as 0.12 at a distance of 20 m, yet 13 of 144 observations (9%) lie beyond 20 m (Fig. 5.1b). We suggest that roughly 10% of observations should be truncated for point transects, or alternatively, w should be chosen such that $g(w) \doteq 0.1$, where $g(w)$ is estimated from a preliminary fit of a plausible model to the data. In this example, a truncation distance of 20 m roughly satisfies both criteria, and is used subsequently for the example data.

Truncation of the data at $w = 20$ m removed 13 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 5.1). Outliers in the right tail of the distance distribution required additional adjustment parameters. Except for the hazard-rate model, which performed relatively poorly on these data, density estimates varied more when data were untruncated (Table 5.2). The poor performance of the uniform + polynomial model when fitted to untruncated data divided into 20

Table 5.2 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 equal in each group)

Data type	Model (key + adjustment)	Truncation			
		$w = 20$ m		w = largest obsn	
		$\hat{D}$	cv (%)	$\hat{D}$	cv (%)
Ungrouped	Uniform + cosine	75.05	14.4	74.13	10.6
	Uniform + polynomial	60.76	12.1	70.88	18.0
	Half-normal + Hermite	70.82	15.7	79.62	12.6
	Hazard-rate + cosine	62.36	18.7	71.02	18.1
Grouped (5 equal)	Uniform + cosine	73.74	14.5	73.77	11.5
	Uniform + polynomial	62.01	12.7	70.14	25.7
	Half-normal + Hermite	69.06	16.0	64.53	26.3
	Hazard-rate + cosine	52.14	14.5	79.13	17.2
Grouped (20 equal)	Uniform + cosine	75.54	14.0	74.24	10.7
	Uniform + polynomial	61.09	12.2	42.26	9.0
	Half-normal + Hermite	71.30	15.6	80.25	12.3
	Hazard-rate + cosine	82.98	26.9	70.85	18.6
Grouped (5 unequal)	Uniform + cosine	74.91	14.3	74.36	15.8
	Uniform + polynomial	61.93	12.8	57.45	37.0
	Half-normal + Hermite	71.57	15.8	80.73	13.0
	Hazard-rate + cosine	84.76	37.2	57.50	13.9

groups was because DISTANCE failed to converge when attempting to fit a better model; convergence problems are encountered more commonly when appropriate truncation is not carried out. If the correct model is known and used, then truncation is not necessary provided the measurements are exact and no evasive movement prior to detection occurs. However, the true model is never known in field surveys.

Truncation of the distance data deletes outliers and facilitates model fitting. However, as some data are discarded, the uncertainty in $\hat{D}$ may increase. For the example data, when the true model was used (i.e. the half-normal), the coefficient of variation was around 3% higher for analyses of truncated data in three of the four analyses, but 10% lower (16.0%, compared with 26.3%) in one of the analyses of grouped data. When an incorrect model was fitted, the cv increased after truncation in eight analyses and decreased in four.

The true density in this example was 79.6 objects/ha. In exactly one half of the 16 analyses of Table 5.2, the estimate was closer to the true density after truncation than before. The case for truncation is therefore not compelling for these simulated data. However, real data tend to be less well behaved, and if no truncation is imposed in the field, truncation at the analysis stage is advisable.

5.4 Estimating the variance in sample size

If objects were known to be distributed at random, the distribution of sample size n would be Poisson with $\text{var}(n) = E(n)$, so that $\widehat{\text{var}}(n) = n$. Most biological populations exhibit some degree of clumping, so that $\text{var}(n) > E(n)$. If the survey is well designed so that points are spread either systematically or randomly throughout the study area, or within each stratum if the study area is divided into strata, then point transect methods are ideally suited to estimating $\text{var}(n)$ empirically, from the variability in sample size between individual points. For the example data, there were $k = 30$ points, and sample sizes n_i within the truncation distance of $w = 20$ m were 1, 1, 5, 6, 3, 8, 7, 5, 3, 4, 1, 8, 3, 1, 2, 7, 4, 4, 6, 7, 6, 8, 4, 3, 5, 2, 4, 2, 9 and 2.

From Section 3.7.2,

$$\widehat{\text{var}}(n) = k \sum_{i=1}^{k} (n_i - \bar{n})^2/(k-1)$$

with $k - 1$ degrees of freedom. All counts here are positive; had any been zero, they would be retained when calculating the variance. For

the example, the above formula yields $\widehat{\text{var}}(n) = 172.7$, or $\widehat{\text{se}}(n) = 13.14$. Equivalently, the variance of the mean number of objects per point, $\bar{n} = n/k = 4.367$, may be estimated:

$$\widehat{\text{var}}(\bar{n}) = \sum_{i=1}^{k} (n_i - \bar{n})^2 / \{k \cdot (k-1)\}$$

so that $\widehat{\text{var}}(\bar{n}) = 0.1919 = \widehat{\text{var}}(n)/k^2$. Since $n = 131$ after truncation, $\widehat{\text{var}}(n) > n$, indicating possible clumping of objects. However, the variance-mean ratio does not differ significantly from one ($p = 0.12$):

$$(k-1) \cdot \frac{\widehat{\text{var}}(n)}{n} = 38.2$$

which is a value from $\chi^2_{k-1} = \chi^2_{29}$ if the true distribution of n is Poisson. For this simulated example, we know the true distribution is indeed Poisson.

5.5 Analysis of grouped or ungrouped data

Because the assumptions of point transect sampling are known to hold for the example, analysis of ungrouped data is preferred. Generally, little efficiency is lost by grouping data prior to analysis, even with as few as five or six well-chosen intervals. If recorded distances tend to be rounded to favoured values (heaping), or if there is evidence of movement of objects in response to the observer before detection, appropriate grouping of data can lead to more robust estimation of density (Chapter 7). Often, there are sound practical reasons for recording data by distance group, instead of measuring each individual detection distance, in which case the field methods determine the analysis option.

For the example, estimated densities tended to be rather more variable between models when analysis was based on grouped data, although coefficients of variation were not consistently higher (Table 5.2). Provided distances can be measured accurately, and movement in response to the observer before detection is not a problem, we recommend that analysis should be of ungrouped data. Otherwise, data should be grouped. If heaping occurs, group cutpoints should be selected so that favoured distances for rounding tend to occur midway between cutpoints. Choice of group interval is often more critical than for line transect sampling, since a smaller proportion of detections occurs near zero distance, yet it is the value of a function at zero distance, $h(0)$, that must be estimated.

It is this difficulty that gives rise to the relatively large variability in density estimates in Table 5.2. On the other hand, although poor practice, it is not uncommon for 10% or more of perpendicular distances to be recorded as on the line in line transect sampling. It is rare for an object to be recorded as at the point ($r = 0$) in point transect sampling, so that spurious spikes in the detection function at small distances are uncommon.

5.6 Model selection

5.6.1 The models

The same four models for the detection function are considered as in Chapter 4. Thus, the uniform, half-normal and hazard-rate models are used 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. The data 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, so that the addition of a Hermite polynomial term in one analysis of Table 5.1 is not particularly surprising. A histogram of the data using 15–20 intervals, as in Fig. 5.1, tends to reveal the characteristics of the data, such as outliers, heaping, measurement errors, and evasive movement prior to detection.

5.6.2 Likelihood ratio tests

If default settings are accepted, DISTANCE determines the number of adjustment terms required to attain an adequate fit of the data using likelihood ratio tests. Consider the example data, ungrouped and with $w = 20$ m, analysed using the uniform key with a single polynomial adjustment (Table 5.1). How was it determined that a single adjustment was required for this model? Let $\mathcal{L}_0$ be the value of the likelihood for fitting a uniform key alone, let $\mathcal{L}_1$ be the maximum value of the likelihood when a single polynomial term is added, and $\mathcal{L}_2$ be the value after fitting two polynomial terms. Program DISTANCE gives $\log_e(\mathcal{L}_0) = -394.986$, $\log_e(\mathcal{L}_1) = -381.239$ and $\log_e(\mathcal{L}_2) = -381.061$. The likelihood ratio test of the hypothesis that the uniform key provides an adequate description of the data against the alternative that a single polynomial adjustment to the key provides a better fit is carried out by calculating

$$\chi^2 = -2 \log_e (\mathcal{L}_0/\mathcal{L}_1)$$
$$= -2[\log_e (\mathcal{L}_0) - \log_e (\mathcal{L}_1)]$$
$$= -2[-394.986 + 381.239]$$
$$= 27.49$$

If the true model is the uniform key without adjustment, this statistic is distributed asymptotically as χ^2_1. In general, the *df* for this test statistic is the difference in the number of parameters between the two models being tested. A value of 27.49 is much larger than would be expected if the distribution really was χ^2_1 ($p < 0.001$), suggesting that a uniform detection function is not an adequate description of the data, a conclusion that is obvious from Fig. 5.3. Less obvious is whether an additional polynomial term should be fitted. The above test is now carried out, but with $\mathcal{L}_1$ replacing $\mathcal{L}_0$ and $\mathcal{L}_2$ replacing $\mathcal{L}_1$:

$$\chi^2 = -2[\log_e (\mathcal{L}_1) - \log_e (\mathcal{L}_2)]$$
$$= 0.36$$

Again comparing with χ^2_1, this test statistic is not significant ($p = 0.55$), so a further term does not improve the fit of the model significantly. Our experience suggests that a larger value than the conventional $\alpha = 0.05$ is often preferable for the size of the test, and we suggest $\alpha = 0.15$ (Section 3.5.2).

If the likelihood ratio test 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**) rather than just one may provide a significantly better fit. Another solution is to change the default setting of SELECT = **sequential** to SELECT = **forward** or SELECT = **all** in DISTANCE.

5.6.3 Akaike's Information Criterion

Akaike's Information Criterion (AIC) provides a quantitative method for model selection, whether models are hierarchical or not (Section 3.5.3). The adequacy of the selected model should still be assessed, for example using the usual χ^2 goodness of fit statistics and visual inspection of both the estimated detection function and the corresponding density plotted on histograms of the data, as shown in Figs 5.1 and 5.2. The plots allow the fit of the model near the point to be assessed; some lack of fit in the right tail of the data can be tolerated.

AIC was computed for the four models for both grouped and ungrouped data, with truncation distance w set first to 20 m (13 observations

truncated) and then to the largest observation, selected so that no observations were truncated (Table 5.1). Three sets of cutpoints were considered for grouped analyses under each model. Set 1 had five equal groups, set 2 had 20 equal groups, and set 3 had five groups whose width varied, such that the number detected in each distance category was nearly equal. 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, so AIC values can only be compared within each of the eight sets of results in Table 5.1.

The AIC values in Table 5.1 select the half-normal (true) model in four of the eight sets of results. The uniform key with cosine adjustments is selected three times, and the uniform key with simple polynomial adjustments once. Since the half-normal model is selected for the preferred analysis of ungrouped data, truncated at 20 m, the main analysis will be based upon it. However, the AIC value for the uniform + polynomial model is almost the same as for the half-normal + Hermite model, and might equally well be adopted on this basis. We examine the consequences of selecting this model later. The only model that might reasonably be excluded from further consideration on the basis of its AIC value is the hazard-rate + cosine model.

Table 5.3 Goodness of fit statistics for models fitted to the example data with $w = 20$ m and 20 groups

Model	χ^2	df	p
Uniform + cosine	20.34	17	0.26
Uniform + polynomial	19.84	17	0.28
Half-normal + Hermite	19.20	17	0.32
Hazard-rate + cosine	20.32	16	0.21

5.6.4 Goodness of fit

Goodness of fit is another useful tool for model selection (Section 3.5.4). Goodness of fit statistics for the example data without grouping, with $w = 20$ m, and using 20 groups of equal width to evaluate the χ^2 statistic, are given in Table 5.3. These data were taken when all the assumptions were met, and all four models fit the data well. If a model was to be selected from these results, there might be a marginal preference for the half-normal + Hermite polynomial model, which we know to be the correct choice in this case. Heaping in real data sets generally means that fewer than 20 groups should be used, with perhaps six to eight usually being reasonable. If heaping is severe, fewer groups might be required, ideally with each preferred rounding distance falling near

the middle of each group. The grouped nature of the (rounded) data is then correctly recognized in the analysis. If cutpoints are badly chosen, heaping will lead to spurious significant χ^2 values. If data are collected as grouped, the group cutpoints are determined before analysis, although consecutive groups may be merged.

5.7 Estimation of density and measures of precision

5.7.1 The standard analysis

The preferred analysis from the above considerations comprises the fit of the half-normal key without adjustments to ungrouped data, truncated at $w = 20$ m. The variance of n is estimated empirically.

Replacing the parameters of Equation 3.5 by their estimators and simplifying under the assumptions that objects at zero distance are detected with certainty, detected objects are recorded irrespective of their angle from the observer, and objects do not occur in clusters, estimated density becomes

$$\hat{D} = \frac{n \cdot \hat{h}(0)}{2\pi k}$$

where n is the number of objects detected, k is the number of point transects sampled, and $\hat{h}(0)$ is the slope of the estimated density $\hat{f}(r)$ of observed detection distances evaluated at $r = 0$; $\hat{h}(0) = 2\pi/\hat{v}$, where $\hat{v}$ is the effective area of detection.

For the example data, and adopting the preferred analysis, program DISTANCE yields $\hat{h}(0) = 0.01019$, with $\widehat{\text{se}}\{\hat{h}(0)\} = 0.001233$ (based on approximately $n = 131$ degrees of freedom). The units of $\hat{h}(0)$ are m^{-2}. Thus

$$\hat{D} = \frac{131 \times 0.01019}{2\pi \times 30} = 0.00708 \text{ objects/m}^2 \text{ or } 70.8 \text{ objects/ha}$$

The estimator of the sampling variance of this estimate is

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

where $$[\text{cv}(n)]^2 = \widehat{\text{var}}(n)/n^2 = 172.7/131^2 = 0.010065$$

and $$[\text{cv}\{\hat{h}(0)\}]^2 = \widehat{\text{var}}\{\hat{h}(0)\}/\{\hat{h}(0)\}^2 = 0.001233^2/0.01019^2 = 0.01464$$

Then $$\widehat{\text{var}}(\hat{D}) = (70.8)^2\,[0.010065 + 0.01464]$$

$$= 123.84$$

and $$\widehat{\text{se}}(\hat{D}) = \sqrt{\widehat{\text{var}}(\hat{D})}$$

$$= 11.13$$

The coefficient of variation of estimated density is $\text{cv}(\hat{D}) = \widehat{\text{se}}(\hat{D})/\hat{D} = 15.7\%$, which is likely to be adequate for some purposes. Note that even with a sample size of $n = 131$ after truncation, the coefficient of variation is over 15%. A 95% confidence interval could be calculated as $\hat{D} \pm 1.96(\widehat{\text{se}}(\hat{D}))$, giving the interval [49.0, 92.6]. Log-based confidence intervals offer improved coverage by allowing for the asymmetric shape of the sampling distribution of $\hat{D}$ for small n. Applying the procedure of Section 3.7.1, the interval [52.1, 96.2] is obtained, which is wider than the symmetric interval, but is a better measure of the uncertainty in the estimate $\hat{D} = 70.8$. In line transect sampling, the variance of $\hat{D}$ is usually primarily due to the variance in n, but this is less often the case in point transect sampling, where precision in $\hat{h}(0)$ can be poor; here, variance in n accounts for 41% of the total variance estimate.

If the uniform model with polynomial adjustments is adopted, estimated density is 60.9 objects/ha, with 95% log-based confidence interval [48.2, 77.0]. The true parameter value, $D = 79.6$ objects/ha, lies above the upper limit of this interval. We return to this example later, to show how the bootstrap may be used to estimate variances and to determine confidence limits that incorporate model misspecification uncertainty.

For some purposes it is convenient to have a measure of detectability. For example, it may be useful to assess whether the detectability for a species is a function of habitat, which may have implications for survey design. The effective radius of detection $\rho = \sqrt{(v/\pi)}$, estimated by $\hat{\rho} = \sqrt{\{2/\hat{h}(0)\}}$, may be used for this purpose. For long-tailed detection functions, ρ may be considerably larger than intuition would suggest, because large numbers of objects are detected at far distances, where the area surveyed is great, relative to close distances, where the surveyed area is small. A parameter that is unaffected either by this phenomenon or by truncation is $r_{1/2}$, the distance at which the probability of detecting an object is one-half. For any fitted detection function $\hat{g}(r)$, it may be estimated by solving $\hat{g}(\hat{r}_{1/2}) = 0.5$ for $\hat{r}_{1/2}$. For the example with $w = 20$ m and $\hat{\rho} = 14.0$ m and $\hat{r}_{1/2} = 13.0$ m.

We know that the true detection function is half-normal for the example. Using that knowledge, closed form estimators are available and the analysis is simple to carry out by hand, provided the data are

both ungrouped and untruncated. Using the results of Chapter 3, Section 3.4.4,

$$\hat{\sigma}^2 = \sum_{i=1}^{n} r_i^2/2n = 94.81 \text{ m}^2$$

It follows that

$$\hat{h}(0) = 2\pi/\hat{v} = 1/\hat{\sigma}^2 = 0.01055$$

and estimated density is

$$\hat{D} = 144 \times \hat{h}(0)/(2\pi \times 30) = 0.00806 \text{ objects/m}^2, \text{ or } 80.6 \text{ objects/ha}$$

The effective radius of detection is estimated as $\hat{\rho} = \sqrt{(2\hat{\sigma}^2)} = 13.8$ m, and the radius at which probability of detection is one-half is estimated by $\hat{r}_{1/2} = \sqrt{(2\hat{\sigma}^2 \log_e 2)} = 11.5$ m. These estimates are in excellent agreement with the true values of $D = 79.6$ objects/ha, $\rho = 14.1$ m, and $r_{1/2} = 11.8$ m.

The results of Section 3.4 also yield variance estimates for this special case:

$$\widehat{\text{var}}[\hat{h}(0)] = 4/\sum_{i=1}^{n} (r_i^2 - 2\hat{\sigma}^2)^2 = 8.850 \times 10^{-7}, \text{ or } \widehat{\text{se}}[\hat{h}(0)] = 9.407 \times 10^{-4}$$

Thus $[\text{cv}\{\hat{h}(0)\}]^2 = \widehat{\text{var}}\{\hat{h}(0)\}/\{\hat{h}(0)\}^2 = 0.0009407^2/0.01055^2 = 0.007951$

Also, $[\text{cv}(n)]^2 = 0.010065$ from above

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

$= (80.6)^2 [0.010065 + 0.007951]$

$= 117.04$

and $\widehat{\text{se}}(\hat{D}) = 10.82$

The 95% log-based confidence interval is then [62.0, 104.7] objects/ha.

5.7.2 Bootstrap variances and confidence intervals

The bootstrap is a robust method, based on resampling, for quantifying precision of estimates. One circumstance in which the bootstrap is likely to be preferred is when the user wishes to incorporate in the standard error the component of variation arising from estimating the number of

polynomial or Fourier series adjustments to be carried out. We recommend the following inplementation.

Generate a bootstrap sample by selecting points **with replacement** from the k points recorded until the bootstrap sample also comprises k points. Repeat until B bootstrap samples have been selected. Typically, B will be around 200 to 1000. Density D is estimated from each bootstrap sample, and the estimates are ordered, to give $\hat{D}_{(i)}$, $i = 1, \ldots, B$. Then

$$\hat{D}_B = \left\{ \sum_{i=1}^{B} \hat{D}_{(i)} \right\} / B$$

and

$$\widehat{\text{var}}_B(\hat{D}_B) = \left\{ \sum_{i=1}^{B} (\hat{D}_{(i)} - \hat{D}_B)^2 \right\} / (B - 1)$$

while a $100(1 - 2\alpha)\%$ confidence interval for D is given by $[\hat{D}_{(j)}, \hat{D}_{(j')}]$, with $j = (B + 1)\alpha$ and $j' = (B + 1)(1 - \alpha)$. It is convenient to select B so that j and j' are integer. Thus for $\alpha = 0.025$, one might select from the following values: 199, 239, 279, . . ., 999. The estimate $\hat{D}$ calculated from the original data set is usually used in preference to the bootstrap estimate $\hat{D}_B$, with $\text{se}(\hat{D})$ estimated by $\sqrt{\{\widehat{\text{var}}_B(\hat{D}_B)\}}$. Applying this to the example with $B = 399$ (so that $(B + 1)\alpha = 10$, an integer, for $\alpha = 0.025$), we take a sample of 30 points at random and with replacement from the 30 in the example data set. Suppose this yields the following points: 1, 1, 3, 5, 6, 6, 6, 8, 10, 10, 11, 12, 15, 15, 17, 17, 17, 18, 18, 20, 21, 22, 22, 25, 26, 26, 26, 28, 30, 30. The bootstrap sample therefore comprises each detection distance recorded at points 6, 17 and 26 three times, each distance recorded at points 1, 10, 15, 18, 22 and 30 twice, and each distance from points 3, 5, 8, 11, 12, 20, 21, 25 and 28 once. Those for remaining points are excluded. This bootstrap sample is analysed in exactly the same way as the actual sample, to yield an estimate $\hat{D}_1$. The exercise is repeated 399 times. The sample variance of these bootstrap estimates was 159.8, giving $\widehat{\text{se}}(\hat{D}) = 12.6$ objects/ha. After ordering the bootstrap estimates, the tenth smallest value ($j = (B + 1)\alpha = 10$) was found to be $\hat{D}_{(10)} = 53.6$ and the tenth largest value was $\hat{D}_{(390)} = 100.7$, giving an approximate 95% confidence interval for D of (53.6, 100.7) objects/ha. This compares with $\hat{D} \pm 1.96 \cdot \widehat{\text{se}}(\hat{D}) = (60.8, 100.4)$ha by the more traditional method. Assuming the distribution of $\hat{D}$ is log-normal and using the result of Burnham *et al.* (1987: 212), we obtain the interval (63.1,

102.9) objects/ha. Note that the lower limit is smaller for the bootstrap method. This is because cosine adjustments to the half-normal fit sometimes generated a fitted detection function with a flatter shoulder than that of the half-normal. If no adjustments to the half-normal fit are allowed, the bootstrap should duplicate the analytic method, except asymptotic normality is not assumed when setting confidence limits. Applying this with $B = 399$ gives $\widehat{\text{se}}(\hat{D}) = 10.8$ and an approximate 95% confidence interval for D of (62.5, 102.7), which is shifted slightly to the right of the symmetric analytic interval, reflecting the greater uncertainty in the upper limit, but agrees well with the interval calculated assuming the distribution of $\hat{D}$ is log-normal.

Variances of functions of the fitted density, such as $\hat{\rho}$ or $\hat{r}_{1/2}$, may be estimated using the methods of Section 3.4, or from the above bootstrap method, replacing the bootstrap estimate of density $\hat{D}_{(i)}$ by the appropriate estimate, such as $\hat{\rho}_{(i)}$ or $\hat{r}_{1/2_{(i)}}$. Adopting the analytic approach,

$$\frac{\partial \hat{\rho}}{\partial \hat{\sigma}^2} = 1/\sqrt{(2\hat{\sigma}^2)}$$

so that

$$\widehat{\text{se}}(\hat{\rho}) = \sqrt{(2\hat{\sigma}^6)}/\sqrt{\left\{\sum_{i=1}^{n} (r_i^2 - 2\hat{\sigma}^2)^2\right\}} = 0.61 \text{ m}$$

and

$$\widehat{\text{se}}(\hat{r}_{1/2}) = \widehat{\text{se}}(\hat{\rho}) \cdot \sqrt{(\log_e 2)} = 0.51 \text{ m}$$

By comparison, the bootstrap method yields $\widehat{\text{se}}(\hat{\rho}) = 1.12$ m, with 95% confidence interval (12.62, 16.84) m, and $\widehat{\text{se}}(\hat{r}_{1/2}) = 0.93$ m, with 95% confidence interval (10.51, 14.02) m. If no cosine adjustments are allowed, as above, we get $\widehat{\text{se}}(\hat{\rho}) = 0.66$ m, with 95% confidence interval (12.55, 15.00) m, and $\widehat{\text{se}}(\hat{r}_{1/2}) = 0.55$ m, with 95% confidence interval (10.45, 12.49) m. These results are in good agreement with the analytic results.

We noted earlier that the AIC value for the preferred analysis of the example data was almost the same as that using the uniform key with a single polynomial adjustment. However, the latter model gave an estimated density of 60.9 objects/ha, with 95% confidence interval [48.2, 77.0]. Thus the true parameter value, $D = 79.6$ objects/ha, is outside the confidence interval. The bootstrap option within DISTANCE was implemented with $B = 200$ replicates, to obtain a variance for $\hat{h}(0)$ that

allows for estimation of the number of polynomial terms required. It gave $\widehat{\text{se}}\{\hat{h}(0)\} = 0.000783$, compared with the analytic estimate of $\widehat{\text{se}}\{\hat{h}(0)\} = 0.000581$, which is conditional on a single term adjustment to the uniform key. Thus the variance is larger as expected, and the revised 95% confidence limit for D is [46.6, 79.3]. The true density is therefore still just outside the interval, probably because the uniform + polynomial model gives a negatively biased estimate of density for this data set. To attempt to improve the variance estimate corresponding to $\hat{D} = 60.9$, a component of variance corresponding to model misspecification bias should be estimated. We do this by generating 199 bootstrap samples, and analysing each resample by the three models of Table 5.1 that gave competitive AIC values, namely uniform + cosine, uniform + polynomial and half-normal + Hermite polynomial. In each resample, the bootstrap estimate of density is taken to be the estimated density under the model with the smallest AIC. Under this rule, the uniform + cosine model was selected in 49 of the 199 replicates, the uniform + polynomial model in 92, and the half-normal + Hermite polynomial model in the remaining 58. The 95% percentile confidence interval was [48.0, 94.5] objects/ha, which is wider than the intervals obtained by assuming that the selected model is the correct model, and comfortably includes the true parameter value, $D = 79.6$.

In the above bootstrap implementations, the sampling unit was taken to be the individual point. This is valid if points are randomly distributed through the study area, and provides a good approximation if points are arranged as a regular grid. To reduce travel time between points, transect lines are sometimes defined, and counts are made at regular points along each line. If the spacing between lines is similar to the distance between neighbouring points on the same line, then the point may still be taken as the sampling unit. However, if separation between lines is large, then the line should be taken as the sampling unit. Thus lines are selected with replacement until the number of lines in the resample is equal to the number in the real sample, or, if the number of points per line is very variable, until the number of points in the resample is as close as possible to the number in the real sample. If a line is selected, the data from all points on that line are included in the resample.

5.8 Estimation when the objects are in clusters

If point transects are used for objects that are sometimes recorded in clusters during the survey period, the recording unit should be the cluster, not the individual object, and analyses should be based on clusters. In this

section, various options for the analysis of clusters are considered. If it is assumed that (i) probability of detection is independent of cluster size and (ii) cluster sizes are accurately recorded, or alternatively that they are estimated without bias at all distances, then $E(s)$ may be estimated by the mean size of detected clusters, $\bar{s}$. Estimated cluster density is then

$$\hat{D}_s = \frac{n \cdot \hat{h}(0)}{2\pi k}$$

and estimated object density is

$$\hat{D} = \hat{D}_s \cdot \bar{s} = \frac{n \cdot \hat{h}(0) \cdot \bar{s}}{2\pi k}$$

Note that the formula for cluster density is identical to that for object density when the objects do not occur in clusters. The formula for the variance of $\hat{D}_s$ is also identical to that given for object density in Section 5.7.1. The variance of object density is now estimated by

$$\widehat{\text{var}}(\hat{D}) = \hat{D}^2 \cdot \left\{ \frac{\widehat{\text{var}}(\hat{D}_s)}{\hat{D}_s^2} + \frac{\widehat{\text{var}}(\bar{s})]}{\bar{s}^2} \right\} = \hat{D}^2 \cdot \left\{ \frac{\widehat{\text{var}}(n)}{n^2} + \frac{\widehat{\text{var}}[\hat{h}(0)]}{[\hat{h}(0)]^2} + \frac{\widehat{\text{var}}(\bar{s})}{\bar{s}^2} \right\}$$

where

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

In practice, larger clusters often tend to be more detectable than small clusters at greater distances, so that $E(s)$, and hence D, are overestimated. This is a form of size-biased sampling (Cox 1969; Patil and Ord 1976; Patil and Rao 1978; Rao and Portier 1985). Bias can be negative if the size of a detected cluster at a large distance from the observer tends to be underestimated. If either bias occurs, then the above method should be modified or replaced.

The simplest approach is based on the fact that size bias in detected clusters does not occur within a region around the point for which detection is certain. Hence, $E(s)$ may be estimated by the mean size of clusters detected within distance v of the point, 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 r_i from the point is replaced by s_i objects, each at distance r_i. Thus, the sampling unit is assumed to be the object rather than the

cluster. 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 final method estimates cluster density D_s conventionally, as does the first. Then, given the r_i, $E(s)$ is estimated by regression modelling of the relationship between s_i and r_i. All four approaches are illustrated in this section using program DISTANCE.

The data used to illustrate the four methods were simulated from a half-normal detection function without truncation, in which the scale parameter σ was a function of cluster size:

$$\{\sigma(s)\}^2 = \sigma_0^2\left(1 + b \cdot \frac{s - E(s)}{E(s)}\right)$$

where $\sigma_0 = 30$ m, $b = 0.75$ and $E(s) = 1.85$ for the population. (In Chapter 4, $\sigma(s)$ was assumed to be a linear function of s; for point transects, theoretical considerations suggest that it is more appropriate to assume $\{\sigma(s)\}^2$ is a linear function of s.) Cluster sizes s were generated by simulating values from the geometric distribution with rate $E(s) - 1$ and adding one, and a cluster of size s was detected with probability

$$g(r \mid s) = \exp\left[-\frac{r^2}{2\{\sigma(s)\}^2}\right]$$

The expected sample size was $E(n) = 96$, distributed between $k = 60$ points, with $\text{var}(n) = 2.65 \cdot E(n)$. True densities were $D_s = 283$ clusters/km^2 and $D = 1.85 \times 283 = 523$ objects/km^2. The bivariate detection function $g(r, s)$ is monotone non-increasing in r and monotone non-decreasing in s. The detected cluster sizes are not a random sample from the population of cluster sizes; the mean size of detected clusters $\bar{s}$ has expectation $> E(s)$.

A histogram of the untruncated distance data shows a rather long tail (Fig. 5.4). Truncation at 70 m deleted just under 10% of observations (eight from 92), and allowed the data to be modelled more reliably. The same four models were applied as for Section 5.6: uniform + cosine, uniform + polynomial, half-normal + Hermite polynomial and hazard-rate + cosine. All four models fitted the truncated data well. AICs for the four models were 691.8, 693.7, 692.6 and 693.1, which favour the uniform + cosine model. We therefore use it to illustrate methods of analysis of the example data.

The uniform + cosine model for the untruncated data required three cosine terms to adequately fit the right tail of the data (Fig. 5.4). By truncating the data, only a single cosine term is required (Fig. 5.5), and the size bias in the truncated sample of detected clusters is reduced. The fit of the model was good ($\chi^2_5 = 4.51$; $p = 0.48$). The estimated density of clusters was 258.1 clusters/km^2 ($\widehat{\text{se}} = 52.2$), compared with the true value of 283. The mean cluster size from the untruncated sample data was 2.293 ($\widehat{\text{se}} = 0.165$), which is biased high due to the size-biased sampling. The scatter plot of cluster size against detection distance (Fig. 5.6) shows wide scatter, but a significant correlation ($r = 0.272$). Truncation at $w = 70$ m reduced this correlation to 0.180. Multiplying the density of clusters by the uncorrected estimate of mean cluster size from data truncated at 70 m ($\bar{s} = 2.202$; $\widehat{\text{se}}(\bar{s}) = 0.168$), the density of individuals is estimated as 574.6 objects/km^2 with $\widehat{\text{se}} = 115.1$ and 95% confidence interval [389.6, 847.5], which comfortably includes the true density of 523 objects/km^2.

5.8.1 Standard method with additional truncation

Observed mean cluster sizes and standard errors for a range of truncation distances are shown in Table 5.4. The detection function $g(r)$ was estimated using a truncation distance of w, while a truncation distance of $v (v \leq w)$ was used to estimate mean cluster size. It seems that 70 m may be too large a truncation distance for unbiased estimation of mean cluster size, but an appropriate distance is difficult to determine, because mean cluster size does not stabilize as the truncation distance is reduced. Possible choices for truncation distance v range between 21.5 m, for which $\bar{s} = 1.650$, and 46.9 m, giving $\bar{s} = 2.030$. If strong size bias is

Table 5.4 Observed mean cluster sizes and standard errors for various truncation distances v. Probability of detection at the truncation distance for cluster size estimation, $\hat{g}(v)$, was estimated from a uniform + 1-term cosine model with $w = 70$ m (Fig. 5.5) for $v \leq 70$ m, and from a uniform + 3-term cosine model with $w = 120$ m (Fig. 5.4) for $v = 120$ m

Truncation distance, v(m)	n	$\bar{s}$	$\widehat{\text{se}}(\bar{s})$	$\hat{g}(v)$
120.0	92	2.293	0.165	0.005
70.0	84	2.202	0.168	0.07
46.9	67	2.030	0.183	0.30
36.8	52	2.135	0.228	0.50
31.9	38	2.079	0.243	0.60
27.0	31	1.806	0.199	0.70
21.5	20	1.650	0.232	0.80
14.9	10	2.000	0.422	0.90

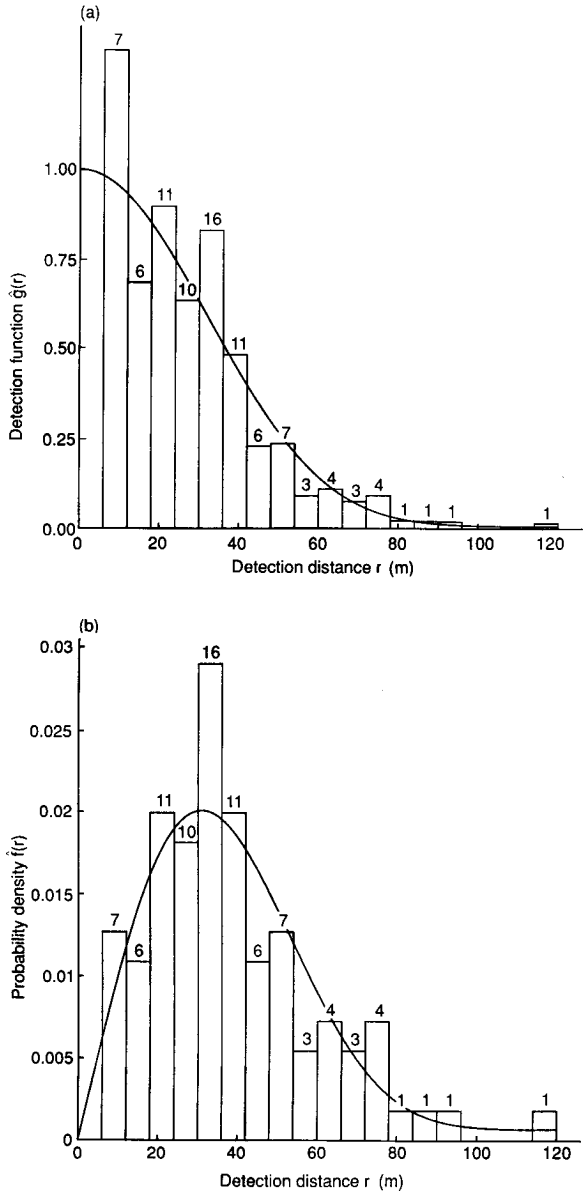

Fig. 5.4. Histograms of the example data using 20 distance categories for the case where cluster size and detection distance are dependent. The fit of a uniform + 3-term cosine detection function to untruncated data is shown in (a), in which frequencies are divided by detection distance, and the corresponding density function is shown in (b).

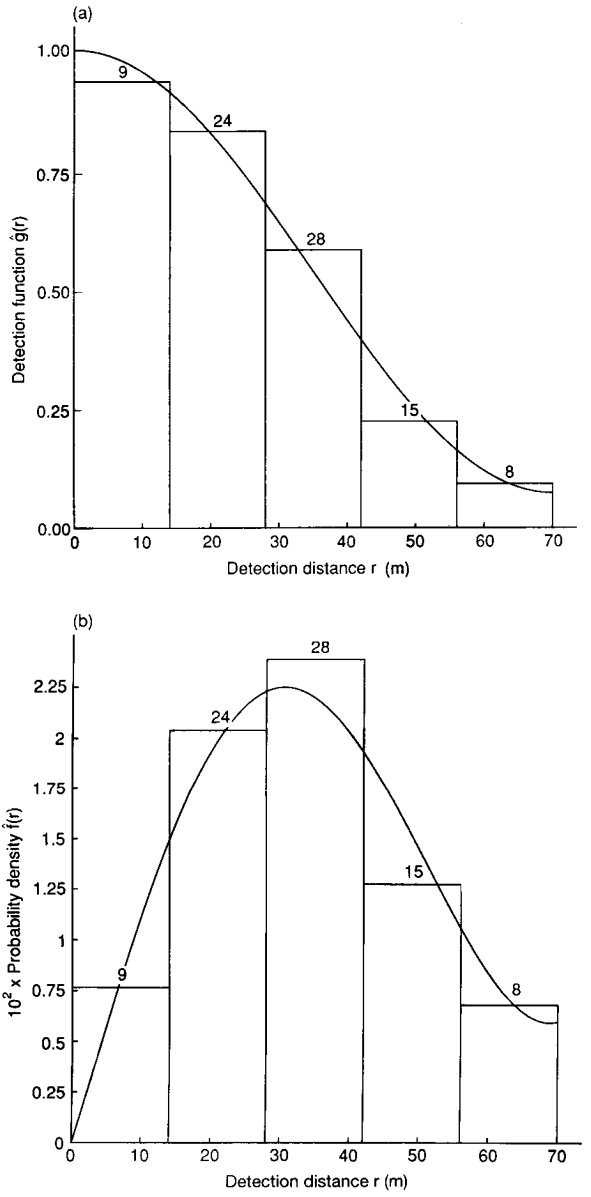

Fig. 5.5. Histograms of the example data using five distance categories and truncation at $w = 70$ m for the case where cluster size and detection distance are dependent. The fit of a uniform + 1-term cosine detection function is shown in (a), in which frequencies are divided by detection distance, and the corresponding density function is shown in (b).

suspected, a reasonable compromise might be $v = 27.0$ m, so that $\bar{s} = 1.806$ with $\widehat{\text{se}} = 0.199$. Replacing the estimates $\bar{s} = 2.202$ and $\widehat{\text{se}} = 0.168$ by these values, density is estimated as 471.2 individuals/km^2, with $\widehat{\text{se}} = 101.5$ and 95% CI [310.4, 715.3]. In view of the difficulty in selecting v, and the sensitivity of the estimate to the choice, another approach seems preferable in this instance.

5.8.2 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, invalidating analytic variance estimates and model selection procedures. Robust methods for variance estimation avoid the first difficulty, but model selection is more problematic. One solution 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. Adopting this strategy, a uniform + 1-term cosine model was fitted to the distance data truncated at 70 m, and the following estimates obtained. Number of objects detected, $n = 185$. Estimated density, $\hat{D} = 526.2$ objects/km^2, with analytic $\widehat{\text{se}} = 104.6$ and 95% confidence interval [355.3, 779.1]. These estimates are lower than those obtained assuming cluster size is independent of distance, and the point estimate is appreciably closer to the true density of 523 objects/km^2. Average cluster size can be estimated by the ratio of estimated object density (526.2) to estimated cluster density (258.1), giving 2.039.

5.8.3 Stratification

Stratification by cluster size can be an effective way of handling size bias. For the example data, if two strata are defined, one corresponding to individual objects and the other to clusters ($\geqslant$ two objects), sample sizes before truncation are 36 and 56 respectively. If the second stratum is split into clusters of size two and clusters of more than two individuals, the respective sample sizes in the three strata before truncation are 36, 27 and 29. The data were analysed for both choices of stratification.

Results are summarized in Table 5.5. As for the line transect example in the previous chapter, no precision is lost by stratification, despite the small samples from which $f(0)$ was estimated. The estimated densities are lower than that obtained by assuming cluster size is independent of detection distance, as would be expected if size bias is present. Both stratifications yield similar estimated densities, and they bracket the

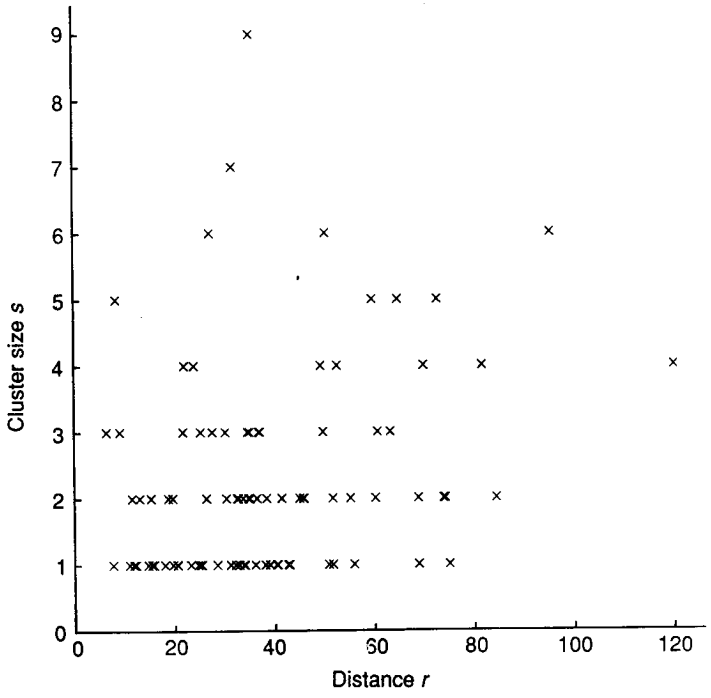

Fig. 5.6. Scatterplot of the relationship between cluster size and detection distance. The correlation coefficient is 0.272 ($w = \infty$).

estimate obtained by the previous method. The true density is 523 objects/km^2, very close to both estimates. 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 $\hat{E}(s) = 534.3/258.1 = 2.070$, and for three strata, $\hat{E}(s) = 524.0/258.1 = 2.030$. Both estimates are rather higher than the true mean cluster size of 1.85.

5.8.4 Regression estimator

Average cluster size can be estimated from a regression of cluster size on estimated detection probability. This procedure estimates the average cluster size for clusters close to the centreline, where detection is assumed to be certain, and thus size bias is reduced. The loss in precision in correcting for size bias using regression is generally small. The method

of regressing $z_i = \log_e(s_i)$ on $\hat{g}(x_i)$ (Section 3.6.3), applied to the example data, yields $\hat{E}(s) = 1.772$ and $\widehat{\text{se}}\{\hat{E}(s)\} = \sqrt{[\widehat{\text{var}}\{\hat{E}(s)\}]} = 0.125$. The corresponding density estimate is 462.2 individuals/km^2, with $\widehat{\text{se}} = 91.6$, cv = 19.8% and 95% confidence interval [313.8, 680.9]. The estimate $\hat{E}(s)$ is close to the true parameter value of 1.85. The resulting density estimate (462.2) is low relative to the true density (523), although the confidence interval comfortably includes the true value.

Table 5.5 Summary of results for different stratification options. Model was uniform with cosine adjustments; distance data were truncated at $w = 70$ m. True $D = 523$ objects/km^2

Cluster size	Sample size after truncation	Efective search radius (m)	$\hat{D}$	$\widehat{\text{se}}(\hat{D})$	95% CI for D
All	84	41.3	574.6	115.1	(389.6, 847.5)
1	35	39.2	120.8	24.7	
2–9	49	43.9	413.5	100.5	
All			534.3	103.5	(366.8, 778.3)
1	35	39.2	120.8	24.7	
2	24	41.6	147.1	42.2	
3–9	25	46.0	256.1	80.9	
All			524.0	94.5	(369.0, 744.1)

5.9 Assumptions

The assumptions of point transect sampling are discussed in Section 2.1. There has been considerable confusion on whether objects must be assumed to be randomly distributed, both in the literature and among biologists. If objects are distributed stochastically independently from each other, but with variable rate depending on location, then the assumption that points rather than objects are randomly located suffices unless the rate shows extreme variation over short distances (of the order of a typical detection distance). If the rate can change appreciably in a short distance or if the presence of one object greatly increases the likelihood that another object is nearby (thus violating the assumption that detections are independent events), then given random placement of points, reliable estimation may still be possible provided robust variance estimation methods are used and provided that the results of goodness of fit and likelihood ratio tests (which will tend to give spurious significances) are viewed with suspicion. The more serious the departure from random, independent detections, the larger the sample size required to yield reliable analyses. Robust empirical or resampling methods should always be used for estimating the variance of sample

size, as described in Section 5.7, to guard against the effects of clustered detections. The most extreme departures from a random distribution of objects are when the objects occur in well-defined clusters. In such cases, the above problems are avoided by taking the cluster rather than the object to be the sampling unit. Strict random placement of points can be modified. For example, stratification of the study area allows sampling intensity to vary between strata, or a regular grid of points may be randomly superimposed on the area. Use of a regular grid allows the biologist to control the distance between points.

Surveys should be designed to minimize departures from the assumption that probability of detection at the point is unity ($g(0) = 1$). For example, the assumption is likely to be more reasonable for songbirds if the recording time at each point is long (giving each bird time to be detected) or if surveys are carried out in early morning, when detectability may be an order of magnitude higher (Robbins 1981; Skirvin 1981). We do not concur with the argument that early morning should be avoided when carrying out point transects. The reasoning behind it is that bird detectability varies rapidly during the first hour or two of daylight. Although detectability may vary less later in the day, it will also be lower, and densities of some species may be appreciably underestimated. Whenever possible, survey work should be carried out when detectability is greatest, and survey design should allow for variation in detectability. Models that are robust to variable detectability (pooling robust) should be used to analyse the data.

Time of season also determines whether it is reasonable to assume that probability of detection at the point is unity. For multiple species studies, it may be necessary to carry out surveys more than once, say early and late in the season. For any given species, the data collected closest to the time that it is most detectable can then be used. For many songbirds, it may be practical to survey only territorial males.

For point transect sampling, we consider that it is necessary to assume that the detection function has a shoulder because we believe that reliable estimation is not possible if it fails, although small departures from the strict mathematical requirement that $g'(0) = 0$ need not be serious. Unlike line transect data, only a very small proportion of point transect distances is close to zero, because the area covered close to the point is small. Thus, there is a case for designing surveys to ensure that $g(r) = 1$ out to some predetermined distance. If there is an area about the point for which detection is perfect, then different point transect models will tend to give more consistent estimation. When $g'(0) = 0$ but $g''(0) < 0$, the stronger criterion of an area of perfect detection fails. Methods based on squaring detection distances (Burnham *et al.* 1980: 195) and the method due to Ramsey and Scott (1979, 1981b) may then

perform poorly. Even when the criterion is satisfied, but the distance up to which detection is certain is close to zero, such methods can be poor.

The mathematical theory assumes that random movement of objects does not occur. In line transect sampling, random movement prior to detection can be tolerated provided average speed of objects is appreciably less than (i.e. up to about one-third of) the speed of the observer (Hiby 1986). The problem is more serious for point transects, for which the observer is stationary. Bias occurs because probability of detection is a non-increasing function of distance from the point, so that objects moving at random are more likely to be detected when closer to the point, leading to overestimation of object density. As noted above, the assumption that $g(0) = 1$ is more plausible if recording time at each point is large, but bias arising from random object movement increases with time at the point; thus recording time at each point is a compromise, and is typically five to ten minutes for songbird surveys.

Response to the observer may take the form of movement towards or away from the observer, or of a change in the probability of detection of the object. Movement towards the observer has a similar effect on the data as random movement, and leads to overestimation of density (Fig. 5.7). Movement away from the observer tends to give rise to underestimation (Fig. 5.7), as does a decrease in detectability close to the point, if this is sufficient to violate the assumption that $g(0) = 1$. An increase in detectability, as when birds 'scold' the intruder, is generally helpful. However, if birds also move in response to the observer, or if females are seldom detected except very close to the point, the detection function might be difficult to model satisfactorily. The effects of response to the observer have been considered by Wildman and Ramsey (1985), Bibby and Buckland (1987) and Roeder *et al.* (1987).

Bibby and Buckland considered two 'fleeing' models. In the first, each object was assumed to maintain a minimum distance (its 'disturbance radius' r_d) between itself and the observer. The radius was allowed to vary from object to object, and was assumed to follow a negative exponential distribution. The detection function was assumed to be half-normal. If the data were to be analysed using a binomial half-normal model (Section 6.2.1) with the division between near and far sightings set at $c_1 = 30$ m (Chapter 6), and if 50% of detections would fall within c_1 in the absence of evasive behaviour, Bibby and Buckland calculated that the bias in $\hat{D}$ (evaluated by numeric integration) would be – 9% when the mean disturbance radius was 10 m, – 20% for 15 m, – 30% for 20 m and – 55% for 40 m. In this case, bias might be deemed 'acceptable' (< 10% in magnitude) if the mean disturbance radius was

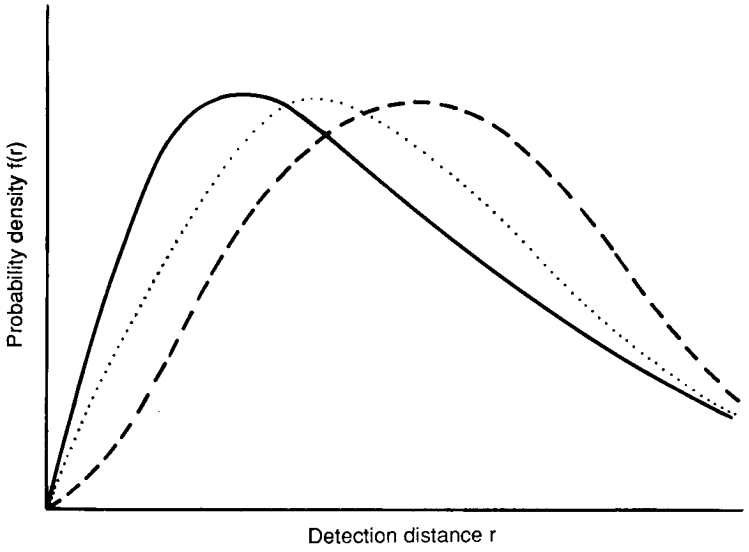

Fig. 5.7. Plots of the real probability density function (·····), the apparent function when there is movement away from the observer (– – –), and the apparent function when there is movement towards the observer (———). Estimated density of birds is proportional to the slope of the curve at zero distance from the point, so that density is overestimated when there is movement towards the observer and underestimated when movement is away from the observer.

of the order of one-third the median detection distance or less. In their second fleeing model, many objects close to a sample point become undetectable, because they either leave at the approach of the observer, moving beyond the range of detection, or take to cover, remaining silent until the observer has departed. The probability that an object at distance r is undetectable was modelled as half-normal. In otherwise identical circumstances to the first model, bias in $\hat{D}$ was found to be – 24% when the point at which 50% of objects become undetectable was 10 m, – 44% for 15 m, – 61% for 20 m and – 88% for 40 m. They concluded that species for which the second model applied were unsuitable for surveying by the point transect method, but considered that the first model, for which bias was less severe, would apply to most species of woodland songbird that show evasive behaviour.

Roeder *et al.* (1987) also considered two models for disturbance, in which the probability of disturbance was exponentially distributed, being one at distance zero. They then simulated data in which a 'disturbed'

object either moved exactly 10 m away from the observer (model 1) or hid (model 2). Their conclusions, based on analyses using the method due to Ramsey and Scott (1979, 1981b), the Fourier series method on squared distances and an order statistic method, were consistent with those reported above. In cases where there is an area of perfect detectability well beyond any effects arising from evasive behaviour, Wildman and Ramsey (1985) showed that their method is still valid under a model in which objects move away from the observer, and can be modified if objects close to the observer are known to hide.

The term 'doughnut' or 'donut' refers to a paucity of observations close to the point, and is generally attributed to object response of one of the above types to the observer. Wildman and Ramsey (1985) used data on the omao or Hawaiian thrush (*Phaeornis obscurus*) as a good example of this. In some instances, a poor choice of model can lead to erroneous identification of a doughnut; the empirical distribution function of detection distances is useful for assessing whether a doughnut really exists.

Distances are assumed to be measured without error (or to be assigned to defined distance intervals without error), but the assumption is less problematic than for line transects in two respects. First, only the observer-to-object distance is required for modelling. This is often easier to measure or estimate than the perpendicular distance of the object from a transect, especially if a detected object is not visible or audible, or has moved, by the time the observer reaches the closest point on the transect to it, or if densities are high, so that the observer may need to keep track of several detections simultaneously. Second, to reduce the problems inherent in estimating perpendicular distances for line transect sampling, sighting distances and sighting angles are often recorded. Effort is often concentrated ahead of the observer, so that measurement errors in the angles often give rise to recorded angles, and hence calculated perpendicular distances, of zero. Such data are notoriously difficult to model. Point transect data do not exhibit this problem; small observer-to-object distances are seldom recorded as zero, and few small distances occur, as the area surveyed close to a point is small.

In songbird point transect surveys on Arapaho National Wildlife Refuge, locations of detected birds were marked, and were later measured to the nearest decimetre (Knopf *et al.* 1988). Such accuracy is not usually possible; for example, up to 90% of detections are purely aural in woodland habitats (Reynolds *et al.* 1980; Scott *et al.* 1981; Bibby *et al.* 1985), so that the location of the bird must be estimated. Consistent bias in distance estimation should be avoided. If distances are overestimated by 10%, densities are underestimated by $100(1 - 1/1.1^2) = 17\%$; if they are underestimated by 10%, densities are overestimated by $100(1/0.9^2 - 1) = 23\%$. Bias in line transect density

estimates would be smaller (9% and 11% respectively). **Provided** distance estimation is unbiased on average, measurement errors must be large to be problematic. Permanent markers at known distances are a valuable aid to obtaining unbiased estimates, and good range finders are effective over typical songbird detection distances, at least when the habitat is sufficiently open to use them. Scott *et al.* (1981) suggest that range finders are accurate to ± 1% within 30 m, and to ± 5% between 100 m and 300 m, whereas trained observers are accurate to ± 10–15% for distances to birds that can be seen. In our experience, range finders are often less accurate than ± 5% at distances close to 300 m.

If most objects are located aurally, then the assumption that an object is not counted more than once from the same point may be problematic. For example, a bird may call at one location, move unseen to another location, and again call. It is seldom problematic if the same bird is recorded from different points, unless it is following the observer.

5.10 Summary

Relative to line transects, relatively few distances are recorded close to zero distance in point transect surveys. Thus estimation of the central parameter ($h(0)$ for point transects and $f(0)$ for line transects) is more difficult, and model selection more critical. This was seen for the first example, where estimation was satisfactory if the correct model was selected, but if the uniform + polynomial model was selected, underestimation occurred, even though the model selection criteria indicated that the model was good. One of the contributory factors to this result was that the true detection function was the half-normal, which does not have an area of perfect detection around the point, even though it has a shoulder. Expressing this mathematically, $g''(0) \neq 0$, even though $g'(0) = 0$. If field methods are adopted that ensure an area of perfect detectability, estimation is more reliable, and different models will tend to give very similar estimates of density. The hazard-rate key is best able to fit data that show a large area over which detection is perfect, because the hazard-rate detection function can fit a wide, flat shoulder. It performed relatively badly on the example data sets largely because it tended to fit a flat shoulder to the simulated data, which were generated from the half-normal, which possesses a rounded shoulder.

To estimate densities reliably from point transect sampling, design and field methods should be carefully determined, following the guidelines of Chapter 7, and the data should be checked for recording and transcription errors. Histograms of the distance data are a useful aid for gaining an understanding of any features or anomalies, and give an

indication of how much truncation is likely to be required. Several potential models should be considered, and model selection criteria applied to choose between them. Special software is essential if efficient, reliable analysis is to be carried out. Variance estimation methods should be chosen for their robust properties; the model that gives the most precise estimate is **not** the best model if either the estimate is seriously biased or the variance estimate ignores significant components of the true variance. A strategy for data analysis is outlined in Section 2.5.

Systematic error in estimated distances must be avoided. Observer training is essential if data quality is not to be compromised (Chapter 7). If more than one observer collects the data, analyses should be attempted that stratify by observer, to detect observer differences. The importance of this is illustrated in Section 8.7. It may prove beneficial to stratify analysis by other factors, such as species, location, habitat, month, year, or any factor that has a substantial impact on detection probabilities. Hypothesis testing may be used to determine which factors affect detection, thus reducing the amount of stratification and increasing parsimony. If the factor is ordinal or a continuous variable, it might enter the analysis as a covariate, so that its effect on detectability is modelled.

If objects occur in clusters, the location of the centre of each cluster and the number of objects in the cluster should be recorded. If clusters occur but are not well-defined, the observer should record each individual object, and its location, and use robust variance estimation methods. It is also useful to indicate which detected objects were considered to belong to the same cluster, so that a comparative analysis can be carried out by cluster. The location of the cluster can then be determined by the analyst, by calculating the geometric centre of the recorded locations.

The checklist of stages in line transect analyses given at the end of Chapter 4 may also be used for point transect analyses. In that checklist, replace 'line' by 'point', 'perpendicular distance' by 'detection distance', and 'Section 4.*' by 'Section 5.*' Also, the rule of thumb for selecting a truncation point w for detection distances is that $\hat{g}(w) \doteq 0.10$, or less satisfactorily, that roughly 10% of observations are truncated (Section 5.3).