6

Extensions and related work

6.1 Introduction

In this chapter, we consider extensions to the theory described in Chapter 3, and we describe distance sampling methods that are closely related to line and point transect sampling. We also examine models that do not fit into the key + adjustment formulation of earlier chapters. The material on these other models is not exhaustive, but is biased towards recent work, and models that may see future use and further methodological development. Most of the older models not described here are discussed in Burnham *et al.* (1980). One of the purposes of this chapter is to stimulate further research by raising some of the issues that are not satisfactorily handled by existing theory.

6.2 Other models

6.2.1 Binomial models

Binomial models are a special case of multinomial models, the theory for which is given in Section 3.4. We examine them briefly, since closed form estimators are available for some underlying models for the detection function; these are sometimes used as indices of abundance, to assess change in abundance with habitat (Section 8.9) or over time.

Line and point transect methods sometimes provide a quick and inexpensive alternative to census methods for generating population abundance indices of songbirds. In areas of thick cover, the observer may rely heavily on aural detection, with perhaps fewer than 10% of detected birds visible. Difficulty in both locating the bird and moving

through vegetation make measurement of each detection distance impractical, and the disturbance would also cause many birds to move or change their behaviour. Bibby *et al.* (1985) stated:

> Recording the distance at which each bird was detected would have been desirable but was not practicable when so many were heard and not seen. Overcoming this difficulty might have risked swamping the observer's acuity for other birds when an average of about nine birds was recorded at each five-minute session. A single decision as to whether or not each bird was within 30 m when first detected was easier to achieve in the field and sufficient to permit estimates of density.

Sometimes, therefore, birds are simply recorded according to whether they are within or beyond a specified distance c_1. To help classify those birds close to the dividing distance, permanent markers may be positioned on trees or bushes at distance c_1. Only single-parameter models may be fitted to such data, and it is not possible to test the goodness of fit of any proposed model. The data may be analysed using the multinomial method for grouped data. Because there are only two groups (with the second cutpoint $c_2 = \infty$), the sampling distribution is binomial. As for the models of Chapter 3, numeric methods will be required in general, but below we consider the half-normal binomial model for point transects (Buckland 1987a), for which analytic estimates are available.

$$\text{Define } g(r) = \exp\{-(r/\sigma)^2\}, \quad 0 \leq r < \infty$$

Then

$$v = \pi\sigma^2$$

and

$$f(r) = 2r \cdot \exp\{-(r/\sigma)^2\}/\sigma^2$$

Given the binomial likelihood, simple algebra yields the maximum likelihood estimate

$$\hat{h}(0) = 2 \cdot \log_e(n/n_2)/c_1^2$$

where n is the number of birds detected and n_2 is the number beyond distance c_1.

Thus,

$$\hat{D} = \frac{n \cdot \hat{h}(0)}{2\pi k} = \frac{n \cdot \log_e(n/n_2)}{c_1^2 \pi k}$$

where k is the number of plots sampled. Suppose for the example of Fig. 5.1 that distances had been recorded simply as to whether they were within or beyond 15 m. Then $c_1 = 15$ m, $k = 30$, $n = 144$ and $n_2 = 46$, yielding $\hat{D} = 0.00775$ birds/m^2, or 77.5 birds/ha.

The asymptotic variance of $\hat{h}(0)$ is

$$\widehat{\text{var}}\{\hat{h}(0)\} = \frac{4(1/n_2 - 1/n)}{c_1^4}$$

so that the variance of $\hat{D}$ may be estimated using the methods of Chapter 5. For the example, we obtain $\widehat{\text{se}}(\hat{D}) = 10.7$ birds/ha, compared with 11.1 birds/ha when exact distances are analysed.

Two measures of detectability are $r_{1/2}$, the point at which the probability of detection is one half, and ρ, the effective radius of detection. Further algebra yields

$$\hat{r}_{1/2} = \sqrt{\left\{\frac{2 \cdot \log_e 2}{\hat{h}(0)}\right\}}, \quad \text{with} \quad \widehat{\text{var}}(\hat{r}_{1/2}) = \frac{2 \cdot \log_e 2 \cdot (1/n_2 - 1/n)}{c_1^4 \cdot \{\hat{h}(0)\}^3}$$

while

$$\hat{\rho} = \sqrt{\left\{\frac{2}{\hat{h}(0)}\right\}}, \quad \text{with} \quad \widehat{\text{var}}(\hat{\rho}) = \frac{2 \cdot (1/n_2 - 1/n)}{c_1^4 \cdot \{\hat{h}(0)\}^3}$$

Thus for the example we have $\hat{r}_{1/2} = 11.7$ m, with $\widehat{\text{se}}(\hat{r}_{1/2}) = 0.6$ m, and $\hat{\rho} = 14.0$ m, with $\widehat{\text{se}}(\hat{\rho}) = 0.7$ m.

The efficiency of this binomial point transect model for estimating density relative to the half-normal model applied to ungrouped detection distances (Ramsey and Scott 1979; Buckland 1987a) is typically around 65%–80% (Buckland 1987a). This loss is relatively small; more serious is that robust models with more than a single parameter cannot be used on binomial data, and there is no information from which to test whether the form of the half-normal model is reasonable. However, bias from fitting an inappropriate model may be consistent between years, so that the method can be useful for providing an index of relative abundance over time at low cost.

Buckland (1987a) also derived analytic results for a linear binomial model for point transects, and found that density estimates for a variety of species are similar under the two models (Section 8.9). In practice, a detection function is unlikely to be approximately linear, so we give just the half-normal model here. The linear model had been considered earlier by Järvinen (1978), but only partially developed.

The choice of c_1 requires some comment. The value that minimizes the variance of $\hat{h}(0)$, of $\hat{r}_{1/2}$ and of $\hat{\rho}$ is $c_1 = 1.78/\sqrt{h(0)}$, which implies that roughly 80% of detections should lie within c_1. Buckland (1987a) finds that estimation is more robust when around 50% lie within c_1. Further, for simultaneously monitoring several species, an average value of 80% across all species may mean that few or no birds of quiet or unobtrusive species are detected beyond c_1. A practical advantage in selecting a smaller value for c_1 than the optimum is that the observer will be more easily able to determine whether a bird is within or beyond c_1.

As a safeguard, two cutpoints, c_1 and c_2, with $0 < c_1 < c_2 < \infty$, might be used, so that the sampling distribution is trinomial. The data could be analysed using the results for the general multinomial distribution in Section 3.4, but detection functions with at most two parameters could be used. Another option would be to use the above binomial model, first using cutpoint c_1, then using cutpoint c_2. If the two density estimates differed appreciably, this might be an indicator that the model is not robust. Otherwise the two estimates might be averaged. For surveys of several species, the first cutpoint might be used for quieter, more unobtrusive species, and the second for louder, more obvious species.

Järvinen and Väisänen (1975) developed three binomial models for line transects, in which the detection function was assumed to be linear, negative exponential or half-normal. The last of these is the most plausible, but a closed form estimator is not available for it. Program DISTANCE allows the user to implement this model using numeric methods. Otherwise, its limitations and advantages are very similar to those of the binomial half-normal model for point transects, described above.

Although the goodness of fit of a binomial model cannot be tested, the homogeneity of the binomial data can. Suppose each line or point transect is assigned to one of R groups, which might, for example, be R geographic regions or woods. Then an $R \times 2$ contingency table analysis may be carried out, where the frequencies are n_{ij}, $i = 1, \ldots, R$, $j = 1, 2$. Then n_{i1} is the total count within distance c_1 for group i, and n_{i2} is the total count beyond c_1. If a significant test statistic is obtained, then there is evidence that the detection function varies between groups. This method extends in the obvious way to multinomial models.

A variety of now outdated line transect methods is given in Burnham *et al.* (1980). In particular, a non-parametric binomial method once thought to have promise is the Cox method, derived by Eberhardt (1978a). We no longer recommend this method. However, as a matter of intellectual curiosity, we derived the analogue of the Cox method for point transect sampling. A linear detection function is assumed, so the model is very similar in concept to the linear binomial point transect model of Buckland (1987a). The difference is that the Cox method assumes that data are truncated at a distance for which the linear detection function is non-zero, whereas the method of Buckland, in common with the linear line transect model of Järvinen and Väisänen (1975), assumes data are untruncated in the field; an estimate of the point at which probability of detection becomes zero is provided by the model.

Let the distance data be grouped with the first two cutpoints being c_1 and c_2. Let the corresponding counts in these two intervals be n_1 and n_2 with total $n = n_1 + n_2$. Let k be the number of points sampled. The Cox estimator is derived by assuming that $g(r) = 1 + b \cdot r$ is an adequate model over the range $0 \leq r \leq c_2$. (It would be better to assume a quadratic form, $g(r) = 1 + b \cdot r^2$, but we use Eberhardt's formulation.) Of course the parameter b is negative. Based on just the counts in these first two intervals, we can get an estimate of b and hence an estimator of density, D. We do not provide the algebraic derivation here. The result is

$$\hat{D} = \frac{n}{k\pi c_2^2}\left[\frac{(c_2/c_1)^3 \cdot (n_1/n) - 1}{(c_2/c_1) - 1}\right]$$

Alternatively,

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

from which one can infer $\hat{h}(0)$. In the simple case of the two intervals having equal widths, Δ (i.e. $c_1 = \Delta$ and $c_2 = 2\Delta$), the result reduces to

$$\hat{D} = \frac{7n_1 - n_2}{4k\pi\Delta^2}$$

In Burnham *et al.* (1980: 169), the general Cox case was given for line transects. The $\hat{f}(0)$ in their publication is not in the same form as used here for $\hat{h}(0)$. For comparison we provide the results below.

Cox estimator for point transects:

$$\hat{h}(0) = \frac{2}{c_2^2}\left[\frac{(c_2/c_1)^3 \cdot (n_1/n) - 1}{(c_2/c_1) - 1}\right]$$

Cox estimator for line transects:

$$\hat{f}(0) = \frac{1}{c_2}\left[\frac{(c_2/c_1)^2 \cdot (n_1/n) - 1}{(c_2/c_1) - 1}\right]$$

For this same context of two cutpoints and truncation at c_2, if we assume the detection function has the generalized form $g(y) = 1 + b \cdot y^p$ for $0 \leq y \leq c_2$ and where p is a known integer ≥ 1, then relevant results for point and line transects are

$$\hat{h}(0) = \frac{2}{c_2^2}\left[\frac{(c_2/c_1)^{2+p} \cdot (n_1/n) - 1}{(c_2/c_1)^p - 1}\right]$$

and

$$\hat{f}(0) = \frac{1}{c_2}\left[\frac{(c_2/c_1)^{1+p} \cdot (n_1/n) - 1}{(c_2/c_1)^p - 1}\right]$$

Corresponding theoretical sampling variances are

$$\text{var}\{\hat{h}(0)\} = \frac{4}{c_2^4}\left[\frac{(c_2/c_1)^{2+p}}{(c_2/c_1)^p - 1}\right]^2 \cdot \frac{p_1(1-p_1)}{n}$$

and

$$\text{var}\{\hat{f}(0)\} = \frac{1}{c_2^2}\left[\frac{(c_2/c_1)^{1+p}}{(c_2/c_1)^p - 1}\right]^2 \cdot \frac{p_1(1-p_1)}{n}$$

where $p_1 = E(n_1)/n$, and is estimated by $\hat{p}_1 = n_1/n$.

6.2.2 Empirical estimators

Emlen (1971, 1977) developed a non-mathematical approach for line transect analysis of songbird data. He assumed that a characteristic

proportion of birds of any species will be detected in the surveyed area $2wL$. He called this proportion the coefficient of detectability. This corresponds to the parameter P_a defined in Section 3.1. The method typically uses data from only two to four distance categories, and an estimator of the product $E(n) \cdot f(0)$ is determined from a smoothed frequency histogram of perpendicular distances. The density estimate from this model is probably best considered as a rough index of relative abundance, and is not recommended (Burnham *et al.* 1980: 164).

Ramsey and Scott (1979, 1981b) developed a point transect model similar to Emlen's (1971, 1977) line transect model. Suppose cutpoints are defined at distances $c_0 = 0, c, 2c, \ldots, kc$, so that the truncation point $w = kc$. Let $A(0, i)$ be the area of the circle of radius ic, and let $A(i, j)$ be the area of the annulus with inner radius ic and outer radius jc. Thus, $A(i, j) = \pi c^2(j^2 - i^2)$. Let $n(i, j)$ be the number of birds counted within the annulus. Then the corresponding density $D(i, j)$ may be estimated by

$$\hat{D}(i, j) = \frac{n(i, j)}{A(i, j)}$$

The value i is chosen to be the smallest value such that the likelihood of differing densities within the areas $A(0, i)$ and $A(i, j)$ is at least four times the likelihood of equal densities for all $j > i$. That is, i is the smallest value that satisfies

$$[\hat{D}(0, i)]^{n(0, i)} \cdot [\hat{D}(i, j)]^{n(i, j)} \geq 4 \cdot [\hat{D}(0, j)]^{n(0, j)} \text{ for all } j > i$$

Having calculated i,

$$\hat{D} = \frac{n(0, i)}{A(0, i)}$$

with

$$\widehat{\text{var}}(\hat{D}) = \frac{\hat{D}}{A(0, i)}$$

This variance estimate is only valid if birds are randomly distributed throughout the study area, so is likely to be poor, but an empirical estimate may be obtained, for example using the bootstrap. The density estimate is only valid if there is an area of perfect detectability, assumed to extend out to distance ic.

Wildman and Ramsey (1985) developed the 'CumD' estimator, which is similar to the above, but estimates the distance out to which detection is certain without the need to group the data. The estimator is defined for both line and point transects, and observations are transformed to detection 'areas', defined as $a = 2Lx$ for line transects and $a = \pi r^2$ for point transects. These areas are ordered from smallest to largest, giving $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$, with empirical distribution function $F_n(a)$. Let $j(0) = a_{j(0)} = 0$, and let $j(1)$ be the largest integer such that

$$d_1 = \frac{j(1)}{a_{j(1)}} = \max\left\{\frac{j}{a_j} \mid j \geqslant \sqrt{n}\right\}$$

Then for $m = 2, 3, \ldots$ let $j(m)$ be the largest integer such that

$$d_m = \frac{j(m) - j(m-1)}{a_{j(m)} - a_{j(m-1)}} = \max\left\{\frac{j - j(m-1)}{a_j - a_{j(m-1)}} \mid j > j(m-1)\right\}$$

Straight lines linking the points $[a_{j(m)}, j(m)/n]$, $m = 0, 1, 2, \ldots$, form a convex envelope over $F_n(a)$, which is the isotonic regression estimate of $F(a)$ (Barlow *et al.* 1972) and yields an estimated detection function that is a non-increasing function of distance from the point. The slopes d_m are average estimates of density of detections within annuli of increasing distance from the point. These equate to estimates of object density if all objects within a given annulus are detected. Likelihood ratio tests of the equality of density between the first region and the next $m - 1$, $m = 2, 3, \ldots$, are used to provide a stopping rule. The smallest value of m, m^* say, is chosen such that the null hypothesis is rejected, and all objects within the corresponding radius a_{j^*}, where $j^* = j(m^*)$, are assumed to have been detected, yielding an estimate of density.

This innovative and intuitively appealing approach is computationally inexpensive and easily programmed. However, probability of detection should be at or close to unity for some distance from the point for the method to yield good estimates; because of the paucity of sightings close to the point, this distance must be appreciably greater for point transects than for line transects for comparable performance. Also, estimation of a distance up to which all objects are detected through hypothesis testing causes the method to underestimate density when sample size is small, as there are too few data for the tests to have much power. Bias in the method is therefore a strong function of sample size, at least for small samples. Investigation of how large sample size should be for the method to perform well would be useful. Wildman and Ramsey (1985) show that it must be very large if the true detection function, expressed as a

function of area, is negative exponential. For point transects, this corresponds to the half-normal detection function, when expressed as a function of distance, and bias is still of the order of 23% for a sample size of 10 000.

6.2.3 Estimators based on shape restrictions

Johnson and Routledge (1985) developed a non-parametric line transect estimator based on shape restrictions for which they found 'a general improvement in efficiency over existing estimators.' The method has not seen wide use, but the recent release of software TRANSAN (Routledge and Fyfe 1992) makes it more accessible, and we encourage further evaluation.

The density function $f(x)$ is constrained to be non-negative and to integrate to unity. In addition, Johnson and Routledge added the monotonicity constraint that $f(x)$ must be monotonic non-increasing and the shape constraint that $f(x)$ must have a concave shoulder, followed by a convex tail, separated by a single point of inflection. The range of concavity must be determined, or guessed, by the user, and it is suggested that the percentage of detections that fall within the point of inflection might be as high as 90% or below 50%.

The parameter $f(0)$ is estimated by grouping the distance data, and using the frequencies in a histogram estimator of $f(x)$. Let h_i represent the height of the ith histogram bar, $i = 1, \ldots, u$. Find the adjusted heights, $\tilde{h}_i$, by minimizing

$$\sum_{i=1}^{u} (h_i - \tilde{h}_i)^2$$

subject to the imposed constraints. Then $f(0)$ is estimated by $\hat{f}(0) = \tilde{h}_1$.

Johnson and Routledge used a bootstrap approach to quantify precision, but one that is more sophisticated than the general purpose bootstrap described in Section 3.7.4. For the single-parameter case, a guess can be made of say the lower $100(1 - 2\alpha)\%$ confidence limit for $f(0)$, and bootstrap resamples generated. The parameter $f(0)$ is estimated from each resample, and if the proportion exceeding the estimate from the true data is greater than α, then the current guess of the limit is estimated to be too large, and is reduced. The process is repeated until the true limit is located with adequate precision. Johnson and Routledge suggested that a more efficient search procedure might be developed. A general algorithm for evaluating confidence limits in this way for single parameter problems, which updates the estimated limit after each

resample, was given by Buckland and Garthwaite (1990). Its optimal properties were noted by Garthwaite and Buckland (1992).

Johnson and Routledge based their conclusions on estimator performance on a simulation study, in which the Fourier series and half-normal models were compared with the shape restriction estimator. If the procedures recommended in earlier chapters were implemented, more severe data truncation would have been carried out for some of their simulations, and the Fourier series and half-normal models would have been rejected as inappropriate in some. However, the shape restriction method proved to be robust to choice of truncation point and to the true underlying detection function, and therefore merits further investigation and development.

6.2.4 Kernel estimators

(a) *Line transect sampling* There are several methods for fitting probability densities using kernels. They are based on the concept of replacing a point (a detection distance here) by a distribution, centred on that point. This is done for all observations, and the distributions are summed, to provide the estimated density function. Buckland (1992a) compared the kernel estimator of Silverman (1982) with the Hermite polynomial model for fitting the deer data from survey 11 of Robinette *et al.* (1974). To force the algorithm to fit a symmetric density, differentiable at zero (and hence possessing a shoulder), each distance x was replaced by two, x and $-x$. The optimum window width for a normal distribution with standard deviation σ, i.e. $h = 1.06\sigma n^{-0.2}$, yielded a comparable but less smooth fit to the data than the Hermite polynomial model. The kernel method is far less computer-intensive than the methods recommended here, but the kernel estimate of $f(0)$ is highly sensitive to the choice of window width (Buckland 1992a). Further, the kernel method does not readily yield a variance for $\hat{f}(0)$, although the bootstrap may be used, either as described in Section 3.7.4 or using the more sophisticated approach of Garthwaite and Buckland (1992), noted in Section 6.2.3. A final disadvantage of the kernel method is that covariates cannot be incorporated, thus ruling out the methods of Section 3.8.2. One advantage of the kernel method is that observations have only a local effect on the fitted density. For parametric or semi-parametric methods, if the model fails to fit the tail of the distribution well, its fit at zero distance may be adversely affected.

Brunk (1978) developed a kernel method based on orthogonal series, an approach which is a close parallel to the key + adjustment formulation, especially if the adjustment terms are orthogonal to the key, as for the Hermite polynomial model. Buckland (1992a) found that Brunk's method

gave unstable estimation relative to the Hermite polynomial model when the data were simulated from a markedly non-normal distribution.

(b) *Point transect sampling* Quang (1990b) proposed a method based on kernel techniques. As in his line transect developments, he assumed that perfect detectability occurs somewhere, but not necessarily at zero distance; that is, that $g(r) = 1$ for some value of $r \geq 0$. Given $g(0) = 1$, it was noted in Chapter 3 that density estimation could be reduced to estimation of $h(0) = \lim_{r \to 0} f(r)/r$. Under Quang's formulation, this generalizes, so that the maximum value of the ratio $h(r) = f(r)/r$ over all r must be estimated. He used the kernel method with a normal kernel (Silverman 1986) to estimate $f(r)$ and hence $h(r)$.

The concept of maximizing $h(r)$ is also applicable to series-type models. Suppose a model is selected whose plotted detection function increases with r over a part of its range, thus indicating that objects close to the point are evading detection, either by fleeing or by remaining silent. Then $\hat{h}(0)$, which is the estimated slope of the density at zero, $\hat{f}'(0)$, may be replaced by the maximum value of $\hat{h}(r) = \hat{f}'(r)$ in the point transect equation for estimated density (Section 3.7.1).

6.2.5 Discrete hazard-rate models

The hazard-rate development of Chapter 3 assumed that either the sighting cue or the probability density of flushing time is continuous. Often this is not the case. For example whales that travel singly or in small groups may surface at regular intervals, with periods varying from a few seconds to over an hour, depending on species and behaviour, when the animals cannot be detected.

(a) *Line transect sampling* Schweder (1977) formulated both continuous and discrete sighting models for line transect sampling, although he did not use these to develop specific forms for the detection function.

Let $q(z, x) = \text{pr}\{\text{seeing the object for the first time} \mid \text{sighting cue at } (z, x)\}$

where z and x are defined in Fig. 1.5. Then if the ith detection is recorded as (t_i, z_i, x_i), where t_i is the time of the ith detection, the set of detections comprises a stochastic point process on time and space. The first-time sighting probability depends on the speed s of the observer so that

$$q(z, x \mid s) = Q(z, x) \cdot E\left\{ \prod_{i>1} [1 - Q(z_i, x_i)] \right\}$$

where $Q(z, x)$ is the conditional probability of sighting a cue at (z, x) given that the object is not previously seen; $Q(z, x)$ is thus the discrete hazard. Assuming that detections behind the observer cannot occur, then

$$g(x|s) = \int_0^\infty u(z, x)q(z, x|s)dz$$

where $u(z, x)$ is the probability that a sighting cue occurs at (z, x), unconditional on whether it is detected; $u(z, x)$ is a function of both object density and cue rate.

More details were given by Schweder (1977, 1990), who used the approach to build understanding of the detection process in line transect sampling surveys. In a subsequent paper (Schweder *et al.* 1991), three specific models for the discrete hazard were proposed, and the corresponding detection function for the hazard they found to be most appropriate for North Atlantic minke whale data is:

$$g(x) = 1 - \exp[-\exp\{a' + b' \cdot x + c' \cdot \log_e(x)\}] \qquad (6.1)$$

If we impose the constraints $b' \leq 0$ and $c' < 0$, this may be considered a more general form of the hazard-rate model of Equation 3.7, derived assuming a continuous sighting hazard:

$$g(x) = 1 - \exp[-(x/\sigma)^{-b}]$$

When $b' = 0$, Equation 6.1 reduces to Equation 3.7 with $a' = b \cdot \log_e(\sigma)$ and $c' = -b$. Thus a possible strategy for analysis is to use Equation 3.7 (the standard hazard-rate model) unless a likelihood ratio test indicates that the fit of Equation 6.1 is superior. Both models are examples of a complementary log-log model (Schweder 1990).

(b) *Point transect sampling* Point transects are commonly used to estimate songbird densities. In many studies, most cues are aural, and therefore occur at discrete points in time. Ramsey *et al.* (1979) defined both an 'audio-detectability function' $g_A(r)$ and a 'visual-detectability function' $g_V(r)$. (Both are also functions of T, time spent at the point seeking detections.)

Let $p(r, t) = \text{pr}\{$ object at distance r is not detected within time $t\}$

Then

$$g_A(r) = 1 - p(r, T) = 1 - \sum_{j=1}^{\infty} [1 - \gamma(r)]^j pr(j)$$

where j is the number of aural cues the object makes in time T, $pr(j)$ is the probability distribution of j, and $\gamma(r)$ is the probability that a single aural cue at distance r is detected. This assumes that the probability of detection of an aural cue is independent of time, the number of cues is independent of distance from the observer and the chance of detecting the jth cue, having missed the first $j - 1$, is equal to the chance of detecting the first cue. Hence the audio-detectability function is of the form

$$g_A(r) = 1 - \psi[1 - \gamma(r)]$$

where

$$\psi(s) = \sum_{j=0}^{\infty} s^j pr(j)$$

is the probability generating function of j.

The visual detectability function, $g_V(r)$, is modelled in a continuous framework, and yields Equation 3.8: $g(r) = 1 - \exp[-k(r)T]$. Ramsey *et al.* (1979) then combined these results to give

$$g(r) = 1 - \psi(1 - \gamma(r)) \cdot \exp[-k(r)T]$$

A detectability function may be derived by specifying (1) a distribution for the number of calls, (2) a function describing the observer's ability to detect a single call, and (3) the function $k(r)$ of visual detection intensities. Ramsey *et al.* considered possibilities for these, and plotted resulting composite detection functions. One of their plots shows an audio detection function in which detection at the point is close to 0.6 but falls off slowly with distance and a visual function where detection is perfect at the point, but falls off sharply. The composite detection function is markedly 'spiked' and would be difficult to model satisfactorily. This circumstance could arise for songbird species in which females are generally silent and retiring, and can be avoided by estimating the male population alone, or singing birds alone, if adults cannot easily be sexed. If data are adequate, the female population could be estimated in a separate analysis.

6.3 Modelling variation in encounter rate and cluster size

6.3.1 On the meaning of var(n)

We have concentrated our modelling and data analysis considerations on the detection function $g(y)$ and, where applicable, the mean cluster size $E(s)$. For both of these data components, we emphasize estimation of parameters as a way to extract and represent the structural (i.e. predictable) information in the data and thus make inferences about population abundance. However, it is also necessary to estimate the sampling variance of $\hat{D}$, var($\hat{D}$), which involves at least one additional set of parameters, namely var(n), which may vary over strata and time. The variance of the counts, var(n), is intended to summarize the non-structural or residual component in the counts. By definition, var(n) = $E[n - E(n)]^2$, thus we must consider whether the **expected** encounter rate is constant for each line or point. Expected encounter rate could vary over a sampled area, in which case there is structural information in the counts. We can go further with this idea by considering the information in the actual spatial positions (x–y coordinates) of detections in the sampled area, given the known distribution of effort (i.e. the locations of lines or points). This entails fitting a relative density distribution model to the large-scale spatial structure of object density as revealed through the spatial variation of encounter rate.

This added level of analysis would require fitting $E[n(x, y)]$, the expected encounter rate as a function of spatial position. Then var(n) is estimated from the residuals about this fitted model for encounter rate, denoted by $\hat{E}(n_i/l_i)$ in the case of line transect sampling. (To obtain results for point transects, replace l_i by 1 and L by k throughout this section.) If we have k replicate lines in a stratum, then var(n) should be estimated as

$$\widehat{\text{var}}(n) = L \frac{\sum_{i=1}^{k} l_i \cdot \left[\frac{n_i}{l_i} - \hat{E}\left(\frac{n_i}{l_i}\right)\right]^2}{k - 1} \tag{6.2}$$

If we assume there is no variation in encounter rate among complete lines within a stratum or time period, then we have

$$E\left(\frac{n_i}{l_i}\right) = \mu \tag{6.3}$$

which is constant for all lines. Proper design can ensure that this assumption is reasonable. Relevant design features are stratification, orientation of lines parallel to density gradients (i.e. perpendicular to density contours) and equal, or appropriate, line lengths. When Equation 6.3 is true, the appropriate estimator of μ is $\hat{\mu} = n/L$, and Equation 6.2 gives, within a stratum,

$$\widehat{\text{var}}(n) = L \frac{\sum_{i=1}^{k} l_i \cdot \left[\frac{n_i}{l_i} - \frac{n}{L}\right]^2}{k-1} \tag{6.4}$$

which we have already recommended.

The critical point is that if Equation 6.3 fails to the extent that there is substantial variation in per line encounter rate, with

$$\mu_i = E\left(\frac{n_i}{l_i}\right)$$

then $\widehat{\text{var}}(n)$ from Equation 6.4 is not appropriate as it includes both stochastic (residual) variation and the structural variation among $\mu_1, \ldots, \mu_k$. This latter variation does not belong in var(n), as it represents large-scale variation in the true object density over the study area. This variation is of interest in its own right, but it is difficult to model and estimate. We will return to this point in Section 6.3.3.

A final comment on the meaning of var(n) is in order; as stated above, var(n) is meant to measure the residual (stochastic) variation in detection counts, n. There are two components to this stochastic variation. First, there is always the small-scale, hence virtually unexplainable, spatial variation in locations of objects. Thus even if detection probability was one over the strips or plots in which counts are made, there would be a substantial component to var(n) due to the small-scale sampling variation in the number of objects in the sampled area, a. Second, when detection probability is not one, then there is the further stochastic variation in the counts due to the particular detections made given the number of objects in area a. Although possible, it is neither necessary nor useful to partition var(n) into these two components.

6.3.2 Pooled estimation of a common dispersion parameter b

Often we can assume that the expected encounter rate μ_i is constant for each replicate line or point within a stratum and time of sampling; thus we assume Equation 6.3. As noted above, this equation will hold under

proper design of the study. The subject of this section is efficient estimation of var(n); this is a concern when sample sizes per stratum are low.

The basic idea of efficient estimation of var(n), once the μ_i are appropriately modelled, is that we can model the structure of var(n) over strata and/or time. The idea of modelling variances is not new to statistics (e.g. Carroll and Ruppert 1988) and the practice is becoming increasingly common. As a starting point to any such modelling, we recommend the representation

$$\text{var}(n) = b \cdot E(n) \tag{6.5}$$

and then modelling the dispersion parameter b. However, the only case we consider here is for when b may be constant over strata and/or time. This is a common situation in our opinion.

Assume the data are stratified spatially and/or temporally into V distinct data sets, indexed by $v = 1, \ldots, V$. Within each data set, assume some replication exists, with line lengths

$$l_{vj}, j = 1, \ldots, k_v, v = 1, \ldots, V$$

and corresponding counts n_{vj}. Nominally, we must now estimate V separate count variances var(n_v), $v = 1, \ldots, V$. The problem is there may be sparse data for each estimate, due to little replication within data sets (small k_v) or few detections (small n_v, perhaps under ten). Experience, and some theory, suggests that the dispersion parameter b in Equation 6.5 will be quite stable and can be modelled, thereby reducing the number of dispersion parameters that must be estimated. If objects have a homogeneous Poisson spatial distribution by data set, then $b = 1$. This is not a reasonable assumption; we should expect $b > 1$, but still probably in the range 1 to say 4.

An accepted principle in data analysis is that we should fit the data by a plausible but parsimonious model, i.e. a model that fits the data well with few parameters (e.g. McCullagh and Nelder 1989). This principle applies to dispersion parameters as well as to structural parameters such as $f(0)$, $h(0)$, $E(s)$ and $E(n)$. Below we provide formulae for estimating a common dispersion parameter across all V data sets.

Separate estimates for each data set are found by applying Equations 6.4 and 6.5:

$$\widehat{\text{var}}(n_v) = L_v \frac{\sum_{i=1}^{k_v} l_{vi} \left[\frac{n_{vi}}{l_{vi}} - \frac{n_v}{L_v} \right]^2}{k_v - 1}$$

where $L_v = \sum l_{vi}$
and

$$\hat{b}_v = \frac{\widehat{\text{var}}(n_v)}{n_v}$$

with $k_v - 1$ degrees of freedom. If $b_v = b$ for all V data sets, then under a quasi-likelihood approach (McCullagh and Nelder 1989), the estimator of b is

$$\hat{b} = \frac{\sum_{v=1}^{V} (k_v - 1)\hat{b}_v}{\sum_{v=1}^{V} (k_v - 1)} \tag{6.6}$$

which has $\sum_{v=1}^{V} (k_v - 1)$ degrees of freedom.

The estimator $\hat{b}$ should have good properties because it is based on a general theory. There is, however, an obvious alternative moment estimator:

$$\hat{b} = \frac{\sum_{v=1}^{V} \widehat{\text{var}}(n_v)}{\sum_{v=1}^{V} n_v} = \frac{\sum_{v=1}^{V} n_v \hat{b}_v}{\sum_{v=1}^{V} n_v} \tag{6.7}$$

We performed a limited simulation comparison of these two methods (i.e. Equations 6.6 and 6.7), and failed to distinguish one as inferior, in terms of bias or precision.

The case of a common dispersion parameter and a pooled estimator is analogous to the analysis of variance assumption of homogeneity of error variance and use of the corresponding pooled estimator of error mean square. When b is so estimated, the squared coefficient of variation of $n = n_1 + \ldots + n_v$ is estimated by

$$[\text{cv}(n)]^2 = \frac{\hat{b}}{n}$$

and for any one stratum,

$$[\mathrm{cv}(n_v)]^2 = \frac{\hat{b}}{n_v}$$

These squared coefficients of variation are then used in formulae for sampling variances of $\hat{D}$ and $\hat{D}_v$; in particular, we can get the best possible variance estimator of an individual stratum density estimate based on

$$[\mathrm{cv}(\hat{D}_v)]^2 = \frac{\hat{b}}{n_v} + [\mathrm{cv}\{\hat{f}(0)\}]^2 + [\mathrm{cv}\{\hat{E}(s)\}]^2$$

Assuming constant $f(0)$ and $E(s)$ over the V strata, which the above formula implicity does, the limiting factor on the precision of $\hat{D}_v$ is just the sample size n_v.

Further comparative investigations of the two estimators of b would be useful. Use of such a pooled estimator is most compelling when sample sizes, n_v, are small, in which case weighting by sample size seems, intuitively, to be important. Yet in Equation 6.6, the weights ignore actual sample sizes. Perhaps when per survey sample sizes, n_v, are smaller than the number of replicate lines or points, k_v, Equation 6.7 would be better.

Further thoughts on this matter arise by considering an average density estimate over temporally repeated surveys on the same area. Then an average density over the repeated surveys is estimated as

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

where

$$L = L_1 + \ldots + L_v$$

The sampling variance on $\hat{\bar{D}}$ is provided by the usual formula but with $[\mathrm{cv}(n)]^2$ computed as

$$[\mathrm{cv}(n)]^2 = \frac{\sum_{v=1}^{V} \widehat{\mathrm{var}}(n_v)}{n^2} = \frac{\hat{b}}{n}$$

where $\hat{b}$ is estimated from Equation 6.7, thus supporting use of Equation 6.7 for temporal stratification. However, this argument does not apply

to spatial stratification. (In computing this $\hat{\bar{D}}$, one should consider whether true density varies by time; if it does, then either $\bar{D}$ may not be relevant, or its variance should include the among D_v variation, which the above cv(n) does not do.)

(a) *An example* In 1989 and 1990, Ebasco Environmental, under contract to the US Minerals Management Service, conducted 13 consecutive aerial line transect surveys for marine mammals offshore of the states of Oregon and Washington, USA (Green *et al.* 1992). The same set of 32 parallel transects (i.e. $k_v = 32$ for all v) was flown each survey during an 18-month period. A given survey took about a week; surveys were a month or more apart in time. Two species of dolphin were of particular interest and generated enough detections to allow density estimation: Risso's dolphin (*Grampus griseus*) and Pacific white-sided dolphin (*Lagenorhynchus obliquedens*). Most of the detections occurred during the eight spring and summer surveys (two in each season in both 1988 and 1989). Table 6.1 presents summary results for $\widehat{\text{var}}(n)$ and $\hat{b}$ from these eight surveys.

Table 6.1 Encounter data and dispersion parameter estimates from the study reported on in Green *et al.* (1992) for Pacific white-sided and Risso's dolphins for the spring and summer surveys (indexed by v) in 1988 and 1989 (surveys 5, 6, 7 and 13 were in autumn and winter, and spring survey 9 targeted grey whales). $k_v = 32$ for each survey

Pacific white-sided dolphin				Risso's dolphin			
v	n_v	$\widehat{\text{var}}(n_v)$	$\hat{b}_v$	v	n_v	$\widehat{\text{var}}(n_v)$	$\hat{b}_v$
1	2	3.28	1.64	1	6	10.31	1.72
2	14	58.46	4.18	2	11	37.33	3.39
3	6	21.29	3.55	3	7	10.51	1.50
4	5	3.81	0.76	4	10	25.30	2.53
8	3	5.49	1.83	8	6	7.05	1.18
10	5	33.47	6.69	10	20	66.71	3.34
11	2	0.93	0.47	11	5	21.46	4.29
12	3	2.23	0.74	12	1	0.72	0.72
Totals	40	128.96	19.86		66	179.39	18.67

Table 6.1 shows that individual survey estimates of the dispersion parameter are quite variable, ranging from 0.47 to 6.69 for Pacific white-sided dolphin and from 0.72 to 4.29 for Risso's dolphin. However, the corresponding sample sizes of total per survey counts are small, ranging from 1 to 20, and averaging, per survey, 5 and 8.25 for Pacific

white-sided and Risso's dolphin, respectively. Because of the small sample sizes for these data, most of the variation in the eight estimates of b for each species is surely sampling variation and not variation in true dispersion over time. We believe it is appropriate and desirable to compute and use a single $\hat{b}$ for each species in this case. Equation 6.7 yields $\hat{b} = 128.96/40 = 3.22$ and $179.39/66 = 2.72$ for Pacific white-sided and Risso's dolphins, respectively. Both estimates are close to the 'default' value $b = 3$ suggested by Burnham *et al.* (1980: 35–6) for when no estimates are available (such as in the initial planning of a study).

(b) *Basis for the theory* The derivation of $\hat{\mu}_v$, $\hat{b}_v$ and $\hat{b}$ (from Equation 6.6) can be carried out in a quasi-likelihood framework (McCullagh and Nelder 1989: 323–8). The starting point is the model $n_{vi} = \mu_v \cdot l_{vi} + \varepsilon_{vi}$ with $E(\varepsilon_{vi}) = 0$ and $\text{var}(\varepsilon_{vi}) = b_v \cdot E(n_{vi}) = b_v \cdot \mu_v \cdot l_{vi}$ for $i = 1, \ldots, k_v$ and $v = 1, \ldots, V$. A special case of the model is to use $b_v = b$ for all v. Given independence over all i and v, then the optimal estimator of μ_v is obtained as the solution to the equation

$$\sum_{i=1}^{k_v} - l_{vi} \cdot \left[\frac{n_{vi} - \mu_v \cdot l_{vi}}{b_v \cdot \mu_v \cdot l_{vi}} \right] = 0$$

The solution is

$$\hat{\mu}_v = \frac{n_v}{L_v}$$

and this is true regardless of whether or not the dispersion parameter b varies by stratum. Direct application of quasi-likelihood theory gives the estimate of separate dispersion parameters as

$$\hat{b}_v = \frac{1}{k_v - 1} \cdot \sum_{i=1}^{k_v} \frac{(n_{vi} - \hat{\mu}_v \cdot l_{vi})^2}{\hat{\mu}_v \cdot l_{vi}}$$

$$= L_v \cdot \frac{\sum_{i=1}^{k_v} l_{vi} \left[\frac{n_{vi}}{l_{vi}} - \frac{n_v}{L_v} \right]^2}{(k_v - 1) \cdot n_v}$$

Applying the same theory to the special case of a constant dispersion parameter produces the result

$$\hat{b} = \frac{1}{\sum_{v=1}^{V}(k_v - 1)} \cdot \left[\sum_{v=1}^{V} \sum_{i=1}^{k_v} \frac{(n_{vi} - \hat{\mu}_v \cdot l_{vi})^2}{\hat{\mu}_v \cdot l_{vi}} \right]$$

This is equivalent to

$$\hat{b} = \frac{\sum_{v=1}^{V} \left(\frac{L_v}{n_v}\right) \sum_{i=1}^{k_v} l_{vi} \left(\frac{n_{vi}}{l_{vi}} - \frac{n_v}{L_v}\right)^2}{\sum_{v=1}^{V}(k_v - 1)}$$

$$= \frac{\sum_{v=1}^{V} \frac{1}{n_v} \cdot (k_v - 1) \cdot \widehat{\text{var}}(n_v)}{\sum_{v=1}^{V}(k_v - 1)}$$

$$= \frac{\sum_{v=1}^{V}(k_v - 1) \cdot \hat{b}_v}{\sum_{v=1}^{V}(k_v - 1)}$$

In principle there are ways to test $H_0 : b_v = b$ for all v. However, the motivation for getting a pooled estimate of an assumed constant dispersion parameter is strongest with sparse data, in which case the tests we are aware of are not reliable. Testing $H_0 : b_v = b$ in this distance sampling context with sparse data is an area in need of research.

6.3.3 Modelling spatial variation in encounter rate

Total numbers, or density, in an area can be reliably estimated even if there are predictable trends in density over the area (though care must then be taken with the spatial allocation of lines or points). An example of predictable trends would be a consistent year-to-year density gradient with distance from coastline in some marine mammals, or a biologically significant association of, for example, some dolphin species with measurable oceanographic features such as surface temperature. The distance sampling methods presented in this book can be embedded in point process sampling theory to allow density surfaces to be fitted to the spatial coordinates of detection locations and even to relate such surfaces to measurable covariates. The information for this modelling

is contained in the spatial locations of detections as represented in an x–y coordinate system and in any covariates measured at those locations and elsewhere over the study area.

When important trends in density exist within strata, the main effect on theory presented here is with respect to estimation of var(n) using Equation 6.4. Improved estimation of var(n) requires modelling encounter rate to give reliable estimation of

$$\mu_{vi} = E\left(\frac{n_{vi}}{l_{vi}}\right)$$

and then use of Equation 6.2 by stratum. With carefully designed studies, Equation 6.4 will be reliable. However, we think there is substantial benefit to be gained by the addition of biological information and understanding, made possible by modelling the density of the population over the study area. Thus, while the estimation of var(n) is our motivation for mentioning point process modelling of encounter rate, the benefits to be gained go beyond improved sampling variance estimators.

Some basic theory for embedding distance sampling in a point process model of the population over the study area has been developed by Schweder (1974, 1977) and Burdick (1979). We give here our own view of how one can conceptualize this modelling of encounter rate; some simplifications are made below compared to a completely general theory. Again, we consider line transect sampling; theory simplifies for point transect sampling because points may be treated as dimensionless.

Let $D(x, y)$ represent the intensity function for a point process model of objects over area A. (One can think of $D(x, y) \cdot dx \cdot dy$ as the expected local density about point x, y.) The density parameter D that we have focused on in this book can be defined by

$$\frac{1}{A}\int\int_A D(x, y)dx\,dy = D$$

Note that D is really an average density over the study area; also, technically the above double integral gives $E(N/A)$, whereas $D = N/A$. The symbol A is given a dual role here, both as the scalar size of the area and as a symbol for the set of points defining the study area. In a point process model, the expected number of points in any area $a \subset A$ is

$$E(N|a) = \int\int_a D(x, y)dx\,dy$$

For one realization of the process, the probability of finding one or more points in the area a is $1 - \exp[-E(N|a)]$.

The relevant surveyed areas for line transects are the sample strips of length l_i and width $2w$, thus changing our symbolism some, we can write

$$E(n_i|l_i) = \int\int_{2wl_i} D(x, y)\, p(x, y)\, dx\, dy$$

The above still denotes an integral over an area. Now we have added $p(x, y)$, which is the conditional probability of actually detecting an object at coordinates (x, y), given an object is there. To simplify the formulation of the problem, we translate the local coordinate system for the above double integral so that the x-coordinate is the transect centreline, and the y-coordinate is perpendicular to that line. (Given a straight line, this is a linear translocation-rotation coordinate transformation.) Thus

$$E(n_i|l_i) = \int_0^{l_i}\int_{-w}^{w} D(x, y)\, p(x, y)\, dy\, dx$$

The next step is critical. In the above coordinate representation, for a given y, variations of $p(x, y)$ in x (i.e. along the line of travel) either do not exist or are irrelevant (in the absence of covariates to explain such spatial variations). Therefore, we can now replace $p(x, y)$ by the detection function $g(y)$. In practice, the scale will always be such that l_i is at least an order of magnitude larger than w. Hence we can safely model the local intensity at any point x in this strip over which integration occurs as the intensity which applies on the transect centreline; hence intensity is assumed to be independent of y. That is, in the strip of area $2wl_i$, we assume the model $D(x) = D(x, y)$. Under these simplifying assumptions, the above double integral becomes

$$E(n_i|l_i) = \int_0^{l_i}\int_{-w}^{w} D(x)\, g(y)\, dy\, dx$$

$$= \left[\int_{-w}^{w} g(y)\, dy\right] \cdot \left[\int_0^{l_i} D(x)\, dx\right]$$

$$= 2wl_i \left[\frac{1}{w}\int_{-w}^{w} g(y)\,dy\right] \cdot \left[\frac{1}{l_i}\int_0^{l_i} D(x)\,dx\right]$$

$$= 2wl_i \cdot P_a \cdot \bar{D}_i$$

where $\bar{D}_i$ is the average density along the ith transect. Summing over transects, we get

$$E(n \mid L) = \bar{D} \cdot 2wL \cdot P_a$$

where

$$\bar{D} = \frac{\sum_{i=1}^{k} 2wl_i \cdot \bar{D}_i}{2wL}$$

is the average density along the sample of lines used.

It is the above $\bar{D}$ that line transect methods actually estimate (Burdick 1979). Either the design of the line placement must produce $\bar{D} \equiv D$ (random line placement has the purpose of achieving $E(\bar{D}) = D$; expectation is with respect to randomization of line placement), or we must model $D(x, y)$, fit the model from the sample data of spatial coordinates of detected objects, and compute the overall $\hat{D}(= \hat{N}/A)$ from

$$\frac{1}{A}\int\int_A \hat{D}(x, y)\,dx\,dy = \hat{D} \tag{6.8}$$

Also, from $\hat{D}(x, y)$ and $\hat{P}_a$ we can then get

$$\hat{E}(n_i \mid l_i) = \hat{P}_a \int\int_{2wl_i} \hat{D}(x, y)\,dx\,dy$$

$$= 2wl_i \cdot P_a \cdot \hat{\bar{D}}_i$$

If $\bar{D}_i = \bar{D}$ for all $i = 1, \ldots, k$, then per line encounter rates are constant and we have

$$\hat{E}\left(\frac{n_i}{l_i}\right) = \frac{n}{L}$$

and Equation 6.4 is valid.

The additional information from distance sampling, which is relevant to $D(x, y)$, is in the set of coordinates of detection locations, (x_{ij}, y_{ij}), $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, and any covariates taken at these locations. To extract this spatial information, these points are treated, conceptually, as a sample of size n from the probability density function defined by

$$d(x, y) = \frac{D(x, y)}{2wL \cdot \bar{D}}$$

over the disjoint areas represented by the k lines traversed. Using some form of parametric model for $d(x, y)$, one fits the model to these (x, y) data by standard statistical methods, thereby getting $\hat{d}(x, y)$. This is not a trivial undertaking, but it is possible. The normalization of $d(x, y)$ to integrate to one over the sample area of transects is necessary for identifiability reasons in the fitting of $\hat{d}(x, y)$. Using $\hat{\bar{D}}(\equiv \hat{D})$ obtained from standard line transect analyses, one obtains the desired rescaling:

$$\hat{D}(x, y) = 2wL \cdot \hat{\bar{D}} \cdot \hat{d}(x, y)$$

If $D(x, y)$, hence $d(x, y)$, is taken as constant over the entire study area, then $d(x, y)$ integrates over the study area to $A/(2wL)$, and from Equation 6.8, we get $\hat{D} = \hat{\bar{D}}$. However, if $D(x, y)$ varies substantially over the study area and lines are poorly placed, this approach used in conjunction with Equation 6.8 could give a much less biased estimate of $D = N/A$.

We suggest simplifying the process of relating the (x, y) detection location data to $D(x, y)$ by projecting each location perpendicularly onto the line and using that point as the recorded detection location. At the scale (much larger than w) over which important variations occur in density, this redefined detection location is acceptable. Detection locations now become distances along the lines, and the problem is effectively reduced to one dimension and numerical scaling-integrations become one-dimensional line integrals. Now all the locational data (x_{ij}, y_{ij}), $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, fall on a (disjoint) 'line' in the study area. Thus the pdf $t(x, y)$, which is really one-dimensional, of a detection location is

$$t(x, y) = \frac{D(x, y)}{\gamma}$$

where the normalizing γ is the line integral

$$\gamma = \sum_{i=1}^{k} \int_{l_i} D(x, y)\, dx\, dy$$

Note that γ is not identifiable from the (x, y) location data alone. Any parametric model for $D(x, y|\underline{\theta})$ now generates the parametric likelihood

$$\mathcal{L}(\underline{\theta}) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} t(x_{ij}, y_{ij}|\underline{\theta})$$

Standard numerical MLE and model selection methods can be applied; at each iteration, γ must be recomputed by numerical line integration. Once the MLE $\hat{\underline{\theta}}$ is found, then γ is estimated as $L \cdot \hat{D}$ from the usual line transect estimation of D and we can get

$$\hat{D}(x, y) = t(x, y|\hat{\underline{\theta}}) \cdot L \cdot \hat{D}$$

(A similar, but by no means identical, development of basic theory is possible for point transects.) More sophistication can be added if the parameters $\underline{\theta}$ affecting density are modelled as functions of covariates recorded at the locations of detections (and available for a grid of points over A).

In many data sets, even when $E(s)$ is also estimated, we see that the contribution of $\widehat{\text{var}}(n)$ to $\widehat{\text{var}}(\hat{D})$ is large, often greater than 50%, and sometimes in excess of 70%. We suspect that Equation 6.3 often fails and a more detailed analysis of the data to estimate varying encounter rates would be useful. The scientifically critical part of this procedure is what to use as a model for $D(x, y)$, or equivalently, $t(x, y)$. The technically difficult part is computing the integrals that are needed, and fitting $t(x, y)$ (or $d(x, y)$, but that is unnecessarily harder). These integrals and the fitting will generally be done by numerical methods. Additional sampling variance is introduced by $\hat{D}(x, y)$, so there must be a worthwhile reduction in the bias of $\widehat{\text{var}}(n)$ and/or bias of $\hat{D}$, and the detection of important spatial trends in density, to justify this additional analysis. General software and methods for these analyses need to be developed. We expect to see this subject area implemented for practical use in the next five to ten years. This spatial modelling in terms of x–y coordinates is a necessary first step to the incorporation of spatially varying covariates that affect object density.

6.3.4 Modelling variation in cluster size

The modelling strategies for encounter rate outlined above may also be applied to cluster size. Spatial and temporal variation in mean cluster size is common, and as with encounter rate, this structural variation should if possible be modelled. A simple method of achieving this is to stratify in space and time before estimating mean cluster size. When sample sizes within strata are small, a common dispersion parameter $c = \text{var}(s)/E(s)$ might be assumed. Suppose the stratification yields V data sets. If $\hat{E}(s_v) = \bar{s}_v$, $v = 1, \ldots, V$, then the variance of $\hat{E}(s_v)$ is estimated by the sample variance of observed group sizes, $\widehat{\text{var}}(s_v)$, divided by n_v. The dispersion parameter is estimated by

$$\hat{c} = \frac{\sum_v \left\{ (n_v - 1) \cdot \dfrac{\widehat{\text{var}}(s_v)}{\hat{E}(s_v)} \right\}}{\sum_v (n_v - 1)}$$

Then

$$\widetilde{\text{var}}(\bar{s}_v) = \frac{\widetilde{\text{var}}(s_v)}{n_v} = \frac{\hat{c} \cdot \bar{s}_v}{n_v}$$

If size bias in the sample of detected clusters is suspected, then $E(s_v)$ and $\text{var}(s_v)$ should be estimated by one of the methods outlined in Section 3.6 before application of the above equation for $\hat{c}$. That method yields estimators $\hat{E}(s_v)$ and $\widehat{\text{var}}[\hat{E}(s_v)] = d_v \cdot \widehat{\text{var}}(s_v)$ for some value d_v, from which the variance of $\hat{E}(s_v)$ is estimated by

$$\widetilde{\text{var}}[\hat{E}(s_v)] = \hat{c} \cdot \hat{E}(s_v) \cdot d_v$$

Modelling the spatial variation in cluster size may be seen as an alternative to dividing an area into geographic strata. The latter is an attempt to create sub-areas in which spatial variation in cluster size is small, whereas the former allows mean cluster size to vary as a continuous function through the area. Similarly, temporal variation in cluster size may be modelled. Having fitted a surface for mean cluster size, using perhaps generalized linear or generalized additive modelling techniques, mean cluster size can be estimated for the study area as a whole, or for any part of it, and if temporal variation is also modelled, mean cluster size can also be estimated at different points in time.

6.3.5 Discussion

The spatial models described in general terms in Section 6.3.3 and alluded to above are also applicable to the parameters $f(0)$ (line transects), $h(0)$ (point transects) and g_0. Similarly, estimation of these parameters by individual strata, combined with the assumption of a common dispersion parameter, are options available to the analyst. However, these parameters are unlikely to vary spatially to the same degree that encounter rate and mean cluster size do. Additionally, estimation of $f(0)$, $h(0)$ or g_0 is bias-prone when samples are small, whereas estimation of encounter rate and mean cluster size are less problematic. For these two reasons, the case for spatial modelling of the detection process, or for estimating $f(0)$, $h(0)$ or g_0 separately within strata, is less compelling than for encounter rate and mean cluster size.

In principle, it is possible to model the density surface, allowing for spatial and temporal variation in individual parameters, together with the effects of environmental conditions on parameters, effects of cluster size or observer on probability of detection, and so on. In practice, considerable software development would be necessary, and if the principle of parsimony was ignored, implementation of such general models would be prevented by numerical difficulties. Section 6.8 on a full likelihood approach lays out the philosophy and structure around which more general modelling could be developed. A simpler, if less comprehensive, strategy is to fit a spatial and, where relevant, temporal model for each parameter in turn. By fitting these models independently, with inclusion of covariates such as environmental factors where required, a spatial surface can be estimated for each of the parameters encounter rate, mean cluster size, $f(0)$ or $h(0)$, and, where relevant, g_0. Density can then be estimated at any point in the study area (and at any time in the study period, if temporal variation is modelled) by combining the estimates from each surface at that point. Abundance can then be estimated for any selected part or the whole of the study area by evaluating the density estimate at a grid of points, and implementing numerical integration. Variances may be estimated using resampling methods, to avoid the assumption that the surfaces for the different parameters are independently estimated.

6.4 Estimation of the probability of detection on the line or point

For both line and point transects it is usual to assume that the probability of detection on the centreline or at the point is unity; that is,

$g_0 = 1$. In practice the assumption is often violated. For example, whales that travel in small groups or that dive synchronously may pass directly under a survey vessel without being detected, or birds in the canopy of high forest directly above an observer may be unrecorded unless they call or sing. As shown in Section 3.1, it is easy to include the component g_0 in the general formula for line and point transects; far more difficult is to obtain a valid estimate of g_0. Most of the methodological development for estimating g_0 has arisen out of the need to estimate the size of cetacean stocks from line transect surveys, so that the effects of commercial or aboriginal takes on the stocks may be assessed. A summary of the evolution of ideas, mostly within the Scientific Committee of the International Whaling Commission, follows.

(a) *Line transect sampling* Early attempts to estimate g_0 were based on the models of Doi (1971, 1974). These were exceptionally detailed models, incorporating the effects of whale dive times and blow times, whale aggregation, response to the vessel, vessel speed, observer height above sea-level, physiological discrimination of the observers, number of observers on duty, binocular specification and angular velocity of eye scanning. The models gave rise to estimates of g_0 with very high estimated precision, but different model assumptions led to estimates that differed appreciably from each other (Best and Butterworth 1980; Doi *et al.* 1982, 1983). In other words, by making many detailed assumptions, the estimator has high precision but at the expense of high bias, and the validity of the approach is questionable.

Butterworth (1982a), using the approach of Koopman (1956), developed a continuous hazard-rate model very similar to that described in Section 3.2, and used it to examine mathematical conditions under which $g_0 < 1$. Butterworth *et al.* (1982) used this formulation to derive a formula for g_0 that was a function of vessel speed:

$$g_0(v) = 1 - \exp(-\alpha/v)$$

where v is vessel speed and α depends on the form of the hazard function. They argued that, if the whales were stationary and vessel speed zero, there would be infinitely many chances to detect whales on the centreline, so that $g_0(0) = 1$. If the true hazard is such that $g_0(v) < 1$ for $v > 0$, and if the specific hazard is independent of vessel speed, then the ratio of estimated whale density assuming $g_0 = 1$ at two different speeds will estimate the ratio of g_0 values at those speeds:

$$\frac{\hat{D}_{v_1}}{\hat{D}_{v_2}} = \frac{\hat{g}_0(v_1)}{\hat{g}_0(v_2)} = \frac{1 - \exp(-\alpha/v_1)}{1 - \exp(-\alpha/v_2)}$$

Solving this equation for α and substituting in the above equation for $g_0(v)$ with $v = v_2$ say yields an estimate of $g_0(v_2)$. Surveys of minke whales in the Antarctic are generally carried out on board vessels travelling at 12 knots.

Butterworth *et al.* (1982) reported on variable speed experiments in which $v_1 = 6$ or 7 knots and $v_2 = 12$ knots. They obtained estimates of $g_0(12)$ close to 0.7, but precision was poor, and estimates did not differ significantly from 1.0. Despite using a hazard-rate formulation, they assumed that the detection function was negative exponential, and Cooke (1985) criticized this; if the form is negative exponential at one speed, then it can be shown mathematically under the above model that the form cannot be negative exponential at the other speed. Cooke (1985) proposed a method based on the ratio of sightings rate at the two speeds. However, he noted that expected precision of estimates from this approach is low even when all the assumptions of the method are satisfied, and listed other reservations about the approach. Hiby (1986) also showed that random whale movement at a speed of 3 knots generated large bias in the sighting rate of a vessel travelling at 6 knots. This bias could be corrected for if the true average speed of movement was known, but he questioned whether it could be reliably estimated. Butterworth (1986) applied Cooke's approach, using four different methods of confidence interval estimation, to two data sets, both separately and combined. He found that most of the 95% intervals for $g_0(12)$ spanned the entire range [0,1], and in every case, the upper limit exceeded 1.0. In one case the lower limit also exceeded 1.0. Given the unresolved difficulties, the method has not been used again.

Zahl (1989) continued development of methods to analyse variable speed data, and put the use of variable numbers of observers on the sighting platform into the same theoretical framework. In common with Schweder (1990), he argued that discrete hazard-rate models are more appropriate than continuous models for whale data. He developed such a model in conjunction with a generalization of Cooke's (1985) method. While precision was improved (Zahl, unpublished), it remained poor, and he did not address the problem of random whale movement.

Butterworth *et al.* (1982) also described a parallel ship experiment. Although designed for examining whether whale movement was affected by the presence of a vessel, they noted that the expected proportion of sightings seen from both platforms in such an experiment can be estimated from the fitted detection function. Their estimates were inconsistent with results from the variable speed experiment, which they attributed to the use of the negative exponential for modelling perpendicular distances. Butterworth *et al.* (1984) extended the method to provide estimates of g_0, assuming a generalized exponential form for the

detection function. Buckland (1987c) generalized their results to provide g_0 estimates from parallel ship data using any model for the detection function, and allowing a different detection function for each vessel. He then analysed the data assuming the continuous hazard-rate model of Section 3.2. This resolved some of the inconsistencies observed in parallel ship data, but the observed duplicate sightings distribution departed significantly from the distribution predicted by the method. If some whales exhibit behaviour that makes them particularly visible relative to other whales, then the duplicate sightings proportion may be higher than anticipated at greater distances, and the observed data show such an effect. Two additional problems remained unresolved. The first is identification of duplicate detections (i.e. whether a sighting made by one platform corresponds with one made by the other), especially in areas of high whale density; the second is assessment of whether estimates of g_0 from parallel ship experiments are valid for correction of abundance estimates derived from normal survey data.

Schweder (1990) proposed new parallel ship experiments, in which one vessel is not only to one side of but also behind the other. He showed using a discrete hazard-rate model (Section 6.2.5) that sightings of cues from the two platforms cannot be considered independent. By placing one vessel behind the other, whales that are below the surface when the first vessel passes may be visible to the second vessel. Results from experiments advocated by Schweder, together with further methodological development, are given in Schweder *et al.* (1991), who estimated g_0 for North Atlantic minke whales. They mapped out surfacings as recorded by one observer in terms of relative position to the other, coding duplicate sightings as 1 and those missed by the reference observer as 0. The hazard probability of sighting was estimated from these data, and integrated multiplicatively, using stochastic simulation. The surfacing pattern of minke whales was estimated from monitoring two whales to which a VHF transmitter had been attached. This allowed them to estimate g_0 without the assumption that the observers detect animals independently. Instead, the assumption of independence is transferred to individual surfacings; conditional on an animal surfacing at a given location, the probability of detection of that surfacing is assumed to be independent between platforms. They obtained $\hat{g}_0 = 0.43$, with 95% confidence interval (0.32, 0.54). This interval took account of uncertainty in whether detections were duplicates.

Several authors have noted that, if a negative exponential model is assumed for the detection function and a correction factor e is defined to allow for deviations of the true detection function from this model, then although neither e nor g_0 can be estimated robustly or with good precision, their ratio, called the 'eh' factor, where $h = 1/g_0$, can

(Butterworth *et al.* 1984; Cooke 1985; Kishino *et al.* 1988). However, the method is still sensitive to the assumption that sightings from the two platforms are independent. Following a comparison of the performance of the negative exponential, exponential power series, hazard-rate, Fourier series and Hermite polynomial models (Buckland 1987b), the Scientific Committee of the International Whaling Commission adopted the hazard-rate model in preference to the negative exponential model.

The variable speed and parallel ship methods both require special experiments. These take the vessels away from survey work, and g_0 during the experiments may not be typical of g_0 values during normal survey mode; for example experiments are carried out in areas of high whale density, so that sample size is adequate. The field procedure most widely used currently is the 'independent observer' method; an additional observation platform is used (for example a second crow's nest vertically below the main one), and observer(s) search independently of the observer(s) on the primary platform. Observers on one platform are not advised of sightings made from the other. Few resources beyond those needed for normal survey mode are required, and the method is therefore often incorporated into normal survey mode.

The methods and problems of analysis are similar to those for a parallel ship experiment with inter-ship distance set to zero, and there is again a strong case for using discrete hazard-rate models in any future methodological development. Identification of duplicate cues is simpler than for two ships, since the bridge can more easily coordinate information from the two platforms, and matches are more easily made when both detections are observed from almost the same position. Exact recording of times of cues aids the identification of duplicate sightings, especially when the same cue is seen from both platforms; if feasible, a whale detected from one platform should be located say from the bridge and followed so that if it is later detected from the other platform, it may be more easily identified as a duplicate.

Two further methods of examining independent observer data have arisen from the Southern Hemisphere minke whale subcommittee of the International Whaling Commission. The first uses one platform to confirm the positions of a sample of schools, and then plots the proportion of these sightings detected by the other platform by distance from the centreline, which should provide an empirical fit to the detection function (with $g(0) \leq 1$) for the second platform. Butterworth and Borchers (1988) describe this approach, and apply it assuming a negative exponential detection function (their 'DNE' method). The second method does not require that duplicate sightings are identified, and is discussed in Hiby and Hammond (1989). It uses information on pairs

of sightings from between and within platforms that are definitely not duplicates to prorate sightings that may be duplicates, without the necessity to identify whether any specific pair of sightings is a duplicate. D.L. Borchers (personal communication) has noted a theoretical shortcoming of this approach, and recommends that it not be used.

One of the most troublesome aspects of estimating g_0 is that different sources of heterogeneity can give rise to substantial bias. Bias in $\hat{g}_0$ might be positive or negative, depending on the type of heterogeneity, and how it is modelled. Observer heterogeneity can arise through different sighting efficiencies for different observers, and through variable sighting efficiency of a single observer through time. Platform heterogeneity is similar in nature. The same observer may have different sighting efficiencies from different platforms, and the relative efficiency of different platforms may vary with environmental or other factors. Environmental heterogeneity affects the efficiency of both the observer and the platform, and environmental variables such as sea state and temperature are likely to affect behaviour of the whales. Individual animals will in any case exhibit heterogeneous behaviour, which leads to too many duplicate detections from double-counting methods, and hence to negatively biased estimates of abundance. Because of the confounding between the various sources of heterogeneity, it is usually not possible to model heterogeneity adequately even when carefully designed experiments are carried out to estimate g_0.

Traditional methods of handling heterogeneity include stratification and covariance analysis, and both are potentially useful for reducing the effects of heterogeneity on estimates of g_0. Generalized linear modelling provides a framework for implementing both approaches. For example, stratification by observer can be achieved by introducing one parameter per observer, and sea state (Beaufort scale) may be incorporated as a regression variable (covariate). The method can be taken further. Suppose parameters are defined for each observer and each platform. Then interaction terms between observer and platform may be introduced. This method is more reliable if each observer is on duty for an equal time on each platform, according to an appropriate design. If observer performance is thought to vary with time, the time that the observer has been on duty (or function(s) of that time) may be entered as a covariate. However, problems arise in practice because different covariates may be highly correlated with each other, and because there may be considerable confounding between stratification factors. Further, the quality of data from which g_0 might be estimated is often poor, so that a realistic model may have more parameters than the data can support, and unquantifiable bias may arise either through practical difficulties in data collection or through inappropriate model specification.

The methods developed for handling heterogeneity in closed population mark–recapture by Huggins (1989, 1991) and Alho (1990) are relevant whenever data from independent platforms are used to estimate g_0. Indeed, probability of detection as a function of distance, estimated by $\hat{g}(x)$, is itself a covariate in this framework. The method is illustrated in a slightly different context in Section 6.12; adding $\hat{g}(x)$ as a covariate allows it to be applied here.

We consider each source of heterogeneity, and use simple examples to illustrate the effects on estimation. For these examples, it is supposed that each detection function is flat and equal to g_0 out to some distance d, and that g_0 is estimated by the Petersen (1896) two-sample mark–recapture estimate. Although this approach is simplistic, it serves to illustrate concepts. Thus we have

$$\hat{g}_{i0} = n_i/\hat{N} \quad \text{for platform} \quad i,\ i = 1, 2 \tag{6.9}$$

$$\hat{g}_0 = 1 - \prod_{i=1}^{2} [1 - \hat{g}_{i0}] = \hat{g}_{10} + \hat{g}_{20} - \hat{g}_{10}\hat{g}_{20} \tag{6.10}$$

for both platforms combined where n_i = number of detections from platform i and

$$\hat{N} = \frac{n_1 n_2}{n_{12}}, \text{ with } n_{12} = \text{number of detections made from both platforms.}$$

Suppose that there is a single observer on each of two platforms, one of whom sees all the whales on the centreline ($g_{10} = 1.0$) while the other sees only one half ($g_{20} = 0.5$). Suppose further that in the first half of the experiment, 50 whales (or whale schools) pass within distance d of the vessel, and the observers see 50 and 25 of these whales respectively. For the second half of the survey, they switch platforms, and again 50 whales pass, of which the first observer sees 50 and the second observer 25. Then if g_0 is estimated by platform using Equation 6.9, $n_1 = n_2 = 75$ and $n_{12} = 50$, so that $\hat{g}_{10} = \hat{g}_{20} = 2/3$, and $\hat{g}_0 = 8/9$. In fact, $g_0 = 1$, so that abundance is overestimated by $100 \times (9 - 8)/8 = 12.5\%$. In this case, the problem may be solved by applying Equation 6.9 to observers instead of platforms, giving $n_1 = 100$ and $n_2 = n_{12} = 50$, so that $\hat{g}_{10} = 1.0$, $\hat{g}_{20} = 0.5$ and $\hat{g}_0 = 1.0$. However, the above argument may now be reversed; if for a given observer g_0 is less for one platform than the other, bias arises for exactly the same reason. A solution is to estimate g_0 separately for each observer on each platform. This allows for an observer effect, a platform effect and an interaction between them. Data may be too sparse to support such an approach; Equation 6.9 and generalizations of it are unstable for small values of n_{12}. Another solution is to assume there is no interaction effect, so that g_0 is assumed to be an additive function

of an observer and a platform effect. This solution may prove satisfactory when the time spent by each observer on each platform is subject to a randomized experimental design. However, it assumes that g_0 for a single observer or platform remains constant throughout the survey.

Platform heterogeneity arising from heterogeneity in environmental conditions would be problematic if for example sighting conditions were better from one platform in some conditions and from the other in different conditions. Thus reflected sunlight may cause one vessel in a parallel ships survey to miss many whales that pass between the vessels and are detected by the other vessel. This will lead to negatively biased estimates of g_0 and positively biased estimates of abundance, since each ship will be affected in this way during different periods of the survey; from a theoretical point of view, it is identical to the problem of the above example, for which platform efficiency changes when the observers swap platforms. However, it is more difficult to design a survey to achieve balance for environmental effects, and appropriate parameterization is problematic.

Observer efficiency may vary over time due to factors such as mood, health, comfort, time on duty, etc. Both observer and platform efficiency will vary with environmental conditions. Bias from these sources will tend to be less; at any given time, implications are similar to the cases considered above, but over time, bias changes. If no observer or platform is consistently more efficient than another, average bias from this source may tend to be small. Additionally, environmental conditions might be introduced into the model as covariates, so that g_0 is related to environmental conditions by a regression model, which might be a generalized linear or non-linear model. Again, data inadequacies may severely constrain the model options.

The fourth class of heterogeneity is heterogeneous behaviour of animals. If some whales are more easily detected than others, then double-counting methods suffer the same bias as two-sample mark–recapture experiments on populations with heterogeneous trappability. Obvious whales are likely to be seen by both observers, whereas unobtrusive whales may be missed by both. The number of whales is therefore negatively biased and $\hat{g}_0$ is positively biased. Suppose that of 160 whales passing within distance d of the vessel, 80 are certain to be seen from each of two platforms and $g_0 = 0.25$ for both platforms for the remaining 80 whales. Assuming independence between platforms, the expected numbers of whales detected are $n_1 = n_2 = 100$, $n_{12} = 85$. Equations 6.9 and 6.10 yield $g_{10} = g_{20} = 0.85$ and $g_0 = 0.9775$. In fact, $g_{10} = g_{20} = 0.625$ and $g_0 = 0.71875$, so that abundance is underestimated by 26.5%. This problem might be partially solved by stratifying detections by animal behaviour, or by type of cue.

Both heterogeneity in whale behaviour and cues that occur only at discrete points in time generate positive bias in the g_0 estimate if the effects are not allowed for, whereas heterogeneity in observer ability and in ease of detection from the respective platforms may lead to negative bias. Schweder's (1990) methods, referenced above, remove the bias that arises from heterogeneity in whale behaviour due to differential diving. If sufficient data are available, the effects of observer heterogeneity might be removed by adding a separate parameter for each observer to the model from which g_0 is estimated, for example using generalized linear models (Gunnlaugsson and Sigurjónsson 1990), although it must still be assumed that a single observer is consistent in ability relative to other observers both within and between watch periods. If there is more than one observer on each platform, another option is to select teams of observers so that each team is of comparable ability. Data analysis should then include a test of whether it is reasonable to assume that each team was equally efficient at detecting whales. It may also be necessary to introduce platform-specific parameters, or at least to test whether such parameterization is required.

Methodological development to solve these difficult problems is continuing, and an innovative paper by Hiby and Lovell (unpublished) proposes a sophisticated approach which allows for response by the whale to the vessel and does not assume stochastic independence between the platforms. The approach uses data on duplicate cues rather than duplicate animals, and can be used in conjunction with either line transect sampling or cue counting methodology. However, it does require surfacing rate to be estimated, and whether the effects on the sightings data of both $g_0 < 1$ and response to the vessel are simultaneously quantifiable has yet to be assessed.

We provide below theory from Buckland (1987c), which allows estimation of g_0 given independent detections from two platforms a distance d apart; $d > 0$ corresponds to parallel ship surveys, where the ships are separated by a distance d, and $d = 0$ corresponds to independent observer platforms on the same vessel or aircraft. The theory for the latter case is also given in Hiby and Hammond (1989). A continuous sighting cue is assumed. The method is subject to potentially serious biases, especially if time periods between cues are not short (say over five minutes for shipboard surveys). Also given is a method described by Buckland and Turnock (1992), which is an extension of ideas utilized by the DNE method (Butterworth and Borchers 1988) mentioned above. The latter approach is more robust to the effects of animal heterogeneity and of animal movement.

Suppose the two platforms are labelled A and B, with B a distance d to the right of A (Fig. 6.1). If both platforms are on the same vessel, then $d = 0$. Suppose a cluster of whales is detected at perpendicular

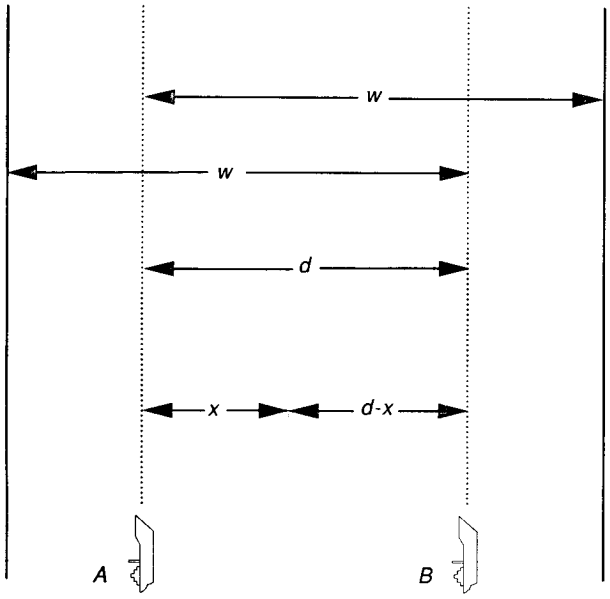

Fig. 6.1. Illustration of parallel platform method. The platforms are a distance d apart, so that a detection at distance x from the centreline of platform A is a distance $d - x$ from the centreline of platform B. If the detected object is to the left of A's centreline, x is negative. A truncation distance of w from the centreline of the farther platform is used.

distance x from A, where x is negative if the detection is to the left of A, and positive if it is to the right. Let the probability that A detects the cluster be $g_A(|x|)$. In terms of notation used elsewhere in this book, $g_A(|x|)$ corresponds to $g(x) \cdot g_0$; thus $g_A(0)$ is not assumed to be unity, but is the value of g_0 for platform A. Further, let the probability that B detects the cluster be $g_B(|d - x|)$, assumed to be independent of $g_A(|x|)$. Then the probability that the cluster is detected by both platforms is $g_A(|x|) \cdot g_B(|d - x|)$.

Now suppose that detected clusters within a distance w of each platform are analysed, where $d/2 < w < \infty$. Define

$$\mu_A = \int_{d-w}^{w} g_A(|x|) \cdot dx$$

$$\mu_B = \int_{d-w}^{w} g_B(|d - x|) \cdot dx$$

and

$$\mu_{AB} = \int_{d-w}^{w} g_A(|x|) \cdot g_B(|d-x|) \cdot dx$$

Further, let

$$v_A = \frac{\mu_A}{g_A(0)}, \; v_B = \frac{\mu_B}{g_B(0)} \quad \text{and} \quad v_{AB} = \frac{\mu_{AB}}{g_A(0) \cdot g_B(0)}$$

Standard line transect analysis of the data from each platform, or of the pooled data from both platforms if the detection function can be assumed to be the same for both, yields estimates of $g(|x|)/g(0)$ for each platform. Hence v_A, v_B and v_{AB} are estimated using numerical or analytic integration. If n_A clusters are detected within the strip of width $2w - d$ (Fig. 6.1) from platform A, n_B from B and n_{AB} from both, then n_{AB}/n_B estimates μ_{AB}/μ_B.

Hence $$\hat{g}_A(0) = \frac{\hat{v}_B \cdot n_{AB}}{\hat{v}_{AB} \cdot n_B}$$

Similarly, $$\hat{g}_B(0) = \frac{\hat{v}_A \cdot n_{AB}}{\hat{v}_{AB} \cdot n_A}$$

Variances for these estimates may be found for example using the bootstrap.

If w is chosen such that $g(|x|)$ is more or less constant for $0 \leq x \leq w$, then $v_A = v_B = v_{AB} = 2w - d$, so that

$$\hat{g}_A(0) = \frac{n_{AB}}{n_B} \quad \text{and} \quad \hat{g}_B(0) = \frac{n_{AB}}{n_A}$$

which may be obtained directly from the two-sample mark–recapture estimate of Petersen (1896).

The following method, due to Buckland and Turnock (1992), is more robust. Suppose an observer or team operates normal line transect sampling techniques from a primary platform (platform A), and an independent observer or team searches a wider area from a secondary platform (platform B), ahead of the normal area of detectability for the primary platform, making no attempt to detect most animals close to the line. The method does not require the assumption that probability

of detection on the centreline is unity. If detections from the secondary platform are made before animals move in response to the presence of the platforms, the method is unbiased when responsive movement occurs before detection from the primary platform. Further, if sighting distances and angles of secondary detections are measured without error, the method is unbiased when sighting distances or angles recorded by the primary platform are subject to bias or error. For example, if the primary platform is a ship and the secondary platform is a helicopter hovering above a detection, ship's radar may be used.

Data from the primary platform are used to estimate the encounter rate (number of detections per unit distance), while data from the secondary platform allow the effective width of search from the primary platform to be estimated. The secondary platform may be thought of as confirming the position of a sample of animals, and the proportions of these detected by the primary platform allow estimation of the detection function, without the necessity to assume $g_0 = 1.0$.

The probability of detection of an animal from one platform should be independent of whether the animal is detected by the other. Detections made by observers on the secondary platform should be at least as far ahead of the primary platform as the maximum distance at which animals are likely to move in response to the presence of the observation platforms. Any secondary detections that occur at shorter distances should be truncated before analysis. Secondary observers should also search out as far as the greatest distance perpendicular to the transect line from which animals would be able to move into the normal detectable range of the primary platform. For animals that are only visible at regular points in time, such as whales with a regular dive cycle, the normal detectability range of the secondary observers should exceed the distance travelled by the primary observers during the course of a single, complete cycle. Conceptually, duplicate detections are expected to be sighted from the secondary platform first, but if the above conditions are met, the analysis may include duplicates first sighted from the primary platform.

Of those animals that pass within the normal detectability range of the secondary observers, the proportion actually detected need not be high, although if few duplicate detections occur, precision will be poor. If the secondary platform cannot be in operation throughout the survey, it should operate during representative, preferably random, subsets of the survey effort.

Secondary observers need not detect all animals on the centreline of either platform. However, the perpendicular distance of each secondary detection from the centreline of the primary platform must be recorded. It is also necessary to determine whether any animal detected by the

secondary platform is also detected by the primary platform. Thus, any animal detected by one platform should be monitored by that platform, or by a third platform, until either the other platform detects the animal or it passes beyond the area searched by the other platform. If animals occur in groups, group size should be recorded by both platforms, either to help identify duplicate detections or, if duplicates are reliably identified and animal groups are well-defined, to validate group size estimates.

Suppose for the moment that both platforms operate throughout the survey. The sightings data from the primary platform are analysed independently of the data from the secondary platform, to yield a conventional line transect density estimate $\hat{D}_A$, calculated assuming no movement and $g_0 = 1$. If animals occur in groups, we define these to be group densities (number of animal groups per unit area), rather than animal densities. The estimates may be biased either because of movement in response to the observation platforms or because probability of detection on the centreline is less than unity. However, for the subset of duplicate detections, the position of the animals prior to any response to the platforms is known. A detection function may therefore be fitted to the distances, as recorded by the secondary platform, of duplicate detections from the centreline of the primary platform. In addition, a detection function, relative to the centreline of the primary platform, is fitted to all secondary detections. An asymptotically unbiased density estimate, $\hat{D}_u$, is calculated as follows.

Let $g_A(x')$ = probability that an animal detected from the secondary platform at perpendicular distance x' from the centreline of the primary platform is subsequently detected from the primary platform, with $g_A(0) \leqslant 1.0$

w = truncation distance for perpendicular distances x'

$f_A(x')$ = probability density of perpendicular distances, prior to responsive movement, of animals subsequently detected by the primary platform

$$= g_A(x')/\mu, \quad \text{with} \quad \mu = \int_0^w g_A(x')\,dx'$$

n_B = number of secondary detections

n_{AB} = number of detections made from both platforms (duplicate detections)

n_A = number of primary detections

$f_B(x')$ = probability density of perpendicular distances from the primary platform centreline for secondary detections

$f_{AB}(x')$ = probability density of perpendicular distances from the primary platform centreline for duplicate detections, as recorded by the secondary platform

$f(x)$ = probability density of perpendicular distances x recorded from the primary platform

L = length of transect line

Then the conventional (biased) estimate of density is

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

and the asymptotically unbiased estimate is given by

$$\hat{D}_u = \frac{n_A \cdot \hat{f}_A(0)}{2L \cdot \hat{g}_A(0)}$$

where

$$\hat{f}_A(0) = \frac{\hat{g}_A(0)}{\int_0^w \hat{g}_A(x')\,dx'}$$

and

$$\hat{g}_A(x') = \frac{n_{AB} \cdot \hat{f}_{AB}(x')}{n_B \cdot \hat{f}_B(x')}$$

The probability densities $f_{AB}(x')$ and $f_B(x')$ are estimated by standard line transect methods. The critical assumptions of the method are as follows.

1. No animals beyond the range of detectability of the secondary platform are able to move into the range of detectability from the primary platform.
2. It is always possible to determine whether an animal detected by the secondary platform is also detected by the primary platform.
3. Given that an animal passes the secondary platform at perpendicular distance x', its probability of detection from the primary platform is independent of whether it was detected by the secondary platform.

4. Perpendicular distances of animals detected by the secondary platform from the centreline of the primary platform are measured exactly, or at least without bias.

If the secondary platform is not in continuous operation, the above procedure is carried out on data collected while both platforms were in operation and a correction factor is calculated as

$$c = \frac{\hat{D}_u}{\hat{D}_A}$$

Density for the entire survey area is estimated by $c \cdot \hat{D}$, where $\hat{D}$ is estimated from the sightings data from the primary platform for the full survey, calculated assuming $g_0 = 1$. To estimate the variance of $\hat{D}_u$ or of c analytically, the correlation between the estimated densities $\hat{f}_B(x')$ and $\hat{f}_{AB}(x')$ and between n_{AB} and both n_B and n_A must be accounted for, and a robust method should be used to estimate the variance in sample sizes. Variance can be estimated more simply and more robustly by applying a resampling method. For example, bootstrap variances may be obtained by resampling from the sightings and effort data from both platforms by day or by search leg (Section 3.7.4).

In the presence of random or responsive movement, $\hat{g}_A(0)$ is not a valid estimate of g_0, since animals close to the centreline when detected by the secondary platform may have moved away from it before detection by the primary platform, and similarly, some away from the centreline may approach it. Thus $\hat{g}_A(0)$ is biased low for g_0 (Buckland and Turnock, 1992). In this case, the method provides a single correction for both sources of bias; stronger assumptions are required to separate the two components of the correction.

The above methods were applied to Dall's porpoise data. The fitted densities $\hat{f}_B(y)$ and $\hat{f}_{AB}(y)$, estimated assuming a hazard-rate model, are shown in Fig. 6.2. The estimate of $g_A(0)$ was 0.597, and the multiplicative correction factor was 0.130 ($\widehat{\text{se}} = 0.050$; 95% percentile confidence interval [0.075, 0.262]). Thus the corrected density estimate is less than one seventh the uncorrected estimate. For these data, the combination of strong attraction of porpoise towards the ship and the close approach most porpoise were able to make before detection by the shipboard observers led to an estimate of porpoise density that was an order of magnitude too high.

The approach outlined above is relatively insensitive to observer, platform and environment heterogeneity, provided the secondary platform is in operation continuously, or at least during a representative sample of time periods in the survey. Animal heterogeneity is more

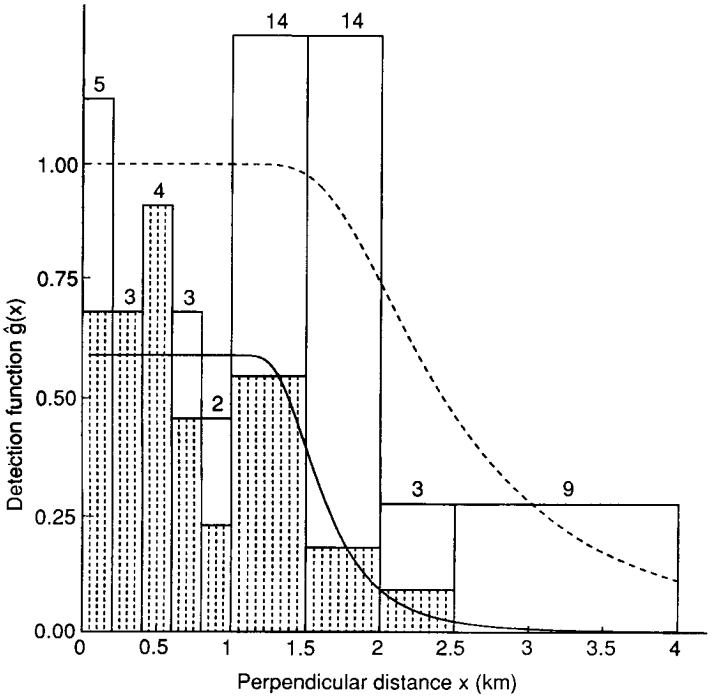

Fig. 6.2. Fitted densities to all helicopter sightings (upper curve) and to duplicate sightings (lower curve), Dall's porpoise, 1984. The hatching indicates the number of duplicate detections in each perpendicular distance interval (as recorded by the helicopter), and the open bars correspond to detections made by the helicopter alone.

problematic, but if the method of detection from the secondary platform is such that the probability of detection from the primary platform is independent of, or only weakly dependent on, whether an animal was detected by the secondary platform, then estimation should be reliable. It may prove necessary to stratify by behaviour of the animal or, if animals occur in groups, by group size to satisfy this requirement.

The field methods of our example, for which a helicopter searches ahead of a survey ship, illustrate one application of the method. If it is not practical to use a helicopter, but normal survey mode is to search with the naked eye, then a simpler solution might prove effective. Suppose an independent observer platform is available on the same vessel or vehicle as the primary platform. Instruct the independent observer to scan through binoculars, searching at distances beyond those typically scanned by the other observers. It does not matter that the

restricted field of view may cause the observer to miss many animals close to the centreline. He or she should follow any detection until either the other observers detect it or it has passed abeam. If his or her average detection distance is substantially greater than that of the other observers, the method might prove satisfactory for the case of diving species, provided the dive cycle is sufficiently short that each animal group surfaces within detectability range at least once. He or she should strive to detect animals beyond the maximum distance they are likely to respond significantly to the vessel, even at the expense of reducing the overall number of detections made. He or she should concentrate effort ahead of the vessel, because the above method corrects for non-uniform effort, and effort searching abeam is wasted, since the other observers are unlikely to detect any animals at large perpendicular distances. If normal searching mode is through hand-held binoculars, the secondary observer could use powerful, tripod-mounted binoculars. By ensuring that the secondary observer searches beyond the normal detectability range from the primary platform, bias from heterogeneity between animals is reduced, especially if the mode of searching from the secondary platform is very different from that from the primary, as would be the case if the secondary platform is a helicopter and the primary platform is a ship or is ground-based.

In practice, it may prove difficult to operate a secondary platform, especially in poor sighting conditions. Even if sufficient detections can be made at distances beyond the range over which animals respond to the observer, it may not be possible to track detected animals, to determine whether they are also detected by the primary platform. In designing line transect surveys, priority should be given to ensuring that g_0 is as close to unity as possible and that detections are made prior to responsive movement. Only if g_0 might be appreciably less than unity or if substantial responsive movement prior to detection is suspected should the methods outlined above be considered.

The methods of Huggins (1989, 1991) and Alho (1990), developed for mark–recapture models, provide a flexible framework for estimation of g_0, which may prove superior to the above methods. As noted earlier, their application is illustrated in a similar context in Section 6.12. By first fitting a detection function to the pooled perpendicular distance data, the estimated probability of detection for an object at distance x, $\hat{g}(x)$, can be included as a covariate. The method used in Section 6.12 then yields an estimate of the probability of detection unconditional on x, and hence of g_0. We encourage further development of this approach.

(b) *Point transect sampling* Estimation of g_0 has seldom been considered for point transect sampling, although in their hazard-rate for-

mulation, Ramsey *et al.* (1979) note that g_0 need not be unity. In particular, if only aural cues are recorded, then birds that do not call or sing will not be detected, irrespective of their position. The cue counting method for estimating whale numbers (described later) is very similar theoretically to point transects, and g_0 estimation is more important in this context. Problems are similar to the line transect case, and are well described by Hiby and Ward (unpublished), who propose a model that allows for the discrete nature of cues and yields estimates of g_0.

6.5 On the concept of detection search effort

The detection of objects in distance sampling requires some type of active search effort. This will often be visual, so that observers must have some visual search pattern. Koopman (1980) discusses ideas on the search and detection process. We suggest that it is useful to consider some concept of search effort, and we pursue this suggestion here for line transects. (Detections are often by aural cues in point transects, in which case it is not clear to us how to model search effort other than as time spent at the point.)

Conceptually, searching effort has its own distribution about the centreline for line transects. Can we separate this concept of search effort from some concept of 'innate' detectability? To a limited (but useful) extent, we think the answer is 'yes'. Let $e(x)$ be relative searching effort at distance x, and let E be total absolute effort over all perpendicular distances. Then the perpendicular distance distribution of total effort is $E(x) = E \cdot e(x)$. Total absolute effort, E, is conceptual because we do not precisely know what constitutes total effort, given that there are subjective aspects to the detection process; we do not know how to measure E on a meaningful scale. However, relative effort at distance $x \cdot dx$ could be the relative time spent searching at perpendicular distance $x \cdot dx$. This measure of $e(x)$ is sensible and could be measured, in principle. Usually, we require that total effort E is sufficient to ensure $g(0) = 1$. Therefore, we will use the norm $e(0) = 1$ to scale $e(x)$. We maintain, and assume, that $e(x)$ should be non-increasing in perpendicular distance x.

We consider here some useful heuristic thinking, while recognizing that this is not the best mathematical approach. Use of a hazard-rate approach is coherent, but is not required for the points we wish to make. For the detection function, write

$$g(x) = d(x) \cdot e(x) \tag{6.11}$$

where $d(x)$ is some innate, or standard, detection function, as for example for some optimal effort, $e_0(x)$. By assumption, both $d(0)$ and $e(0)$ are unity and both functions are non-increasing in x. Based on Equation 6.11, $g'(x) = d'(x) \cdot e(x) + d(x) \cdot e'(x)$, so that $g'(0) = d'(0) + e'(0)$. It follows that if both effort and innate detectability have a shoulder, then so does $g(x)$. However, if effort is poorly allocated perpendicular to the line, then we can get $g'(0) < 0$, i.e. no shoulder in the distance data, even when $d'(0) = 0$, which means that a shoulder is innately possible with a proper search effort design.

For line transect surveys in which there is visual searching for objects, especially aerial and ship surveys, histograms of the detection distances all too commonly have the shape of Fig. 6.3. It seems unlikely that the innate detectability would drop off this sharply; it is more likely that the data reflect an inadequate distribution of search effort or another field problem, such as heaping at zero distance or attraction of animals to the vessel before detection. We focus on effort here.

In order to pursue this idea mathematically, we need to be able to conceptualize innate detectability, $d(x)$. Although we may want to think of $d(x)$ as detectability under some optimal searching pattern $e_0(x)$, it is not possible to define an actual detection function, $g(x)$, free of some implicit or explicit underlying detection effort distribution. For $x \leq w$,

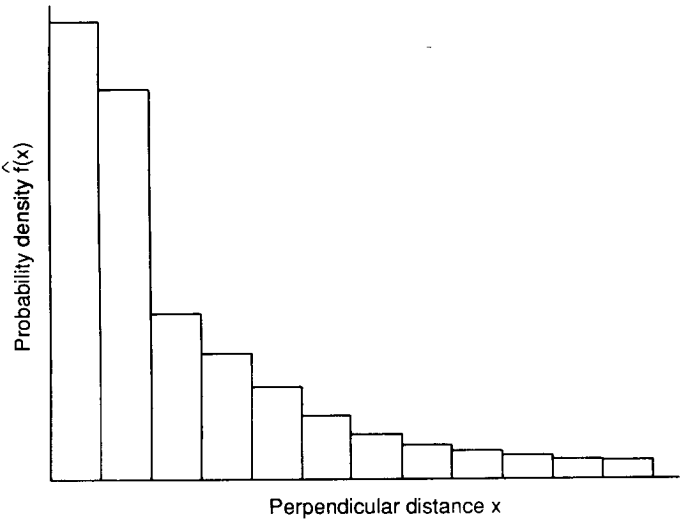

Fig. 6.3. Line transect data exhibiting a shape that is encountered too often; the idealized histogram estimator of the density function $f(x)$ suggests a narrow shoulder followed by an abrupt drop in detection probability, and a long, heavy tail.

we might allow $e(x)$ to be distributed as uniform $(0, w)$, but this becomes unreasonable for large w, and impossible as $w \to \infty$. Still, for small to moderate w, we could define $e_0(x)$ to be uniform. Then for a survey with this effort function, $g(x)$ would reflect the innate detectability of the object at perpendicular distances $x \leq w$.

Our intuition that detectability should not fall off sharply with increasing distance should be applied to $d(x)$. In most line transect sampling with which we have experience, the assumptions that $d(0) = 1$ and that $d(x)$ has a shoulder, i.e. $d'(0) = 0$, seem reasonable to us. (Many marine mammal surveys are exceptions to the first assumption (Section 6.4), and potentially to the second.) When the observed data appear not to exhibit a shoulder, we should bear in mind that the data really came from the detection function

$$g(x) = d(x) \cdot e(x)$$

and hence the probability density function of the observed perpendicular distance data is

$$f(x) = \frac{d(x) \cdot e(x)}{\int_0^w d(x) \cdot e(x)\, dx}$$

If $g(x)$ is as shown in Fig. 6.4, the data may primarily reflect effort $e(x)$, not innate detectability $d(x)$. For any data set, we would like to know the general nature of the effort distribution $e(x)$ to assess our faith in the assumptions that $g(x)$ has a shoulder and satisfies $g(0) = 1$.

Desirable patterns for search effort should be addressed at the design stage, and observers should be trained to follow them; Fig. 6.4 suggests that the search pattern was poor. We suspect that in aerial and ship surveys, there are often two distinct search modes occurring simultaneously: (1) intense scanning of the area near the centreline for much of the time, and (2) occasional scans at greater distances and over large areas, with more lateral effort. This may occur because one observer 'guards' the centreline, searching with the naked eye, while another scans a wider area with binoculars, or a single observer may search mostly with the naked eye, with occasional scans using binoculars. Data then come from the composite probability density function, as indicated in Fig. 6.5. In this case, most choices of histogram interval will obscure the shoulder.

For the innate detectability, $d(x)$, a shoulder should exist. Assume that an object on the centreline is moved just off the line. In an aerial transect

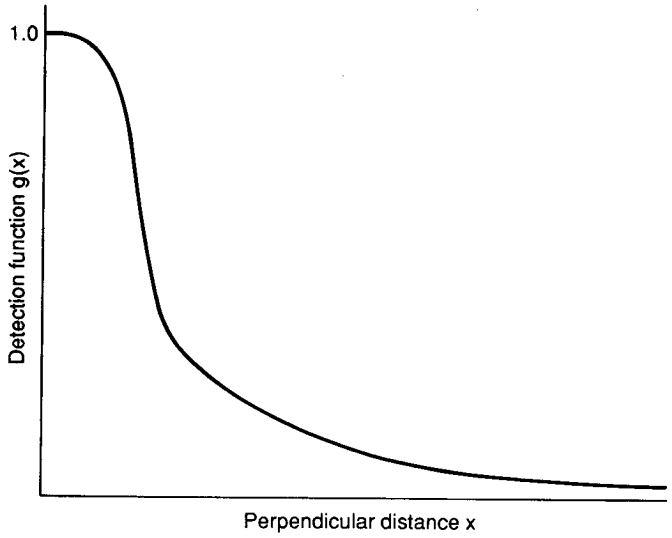

Fig. 6.4. An undesirable relative effort function $e(x)$ can give rise to the detection function shown here, and hence to data that exhibit the features of Fig. 6.3. Relative effort should be expended to ensure that the detection function has a wider shoulder relative to the tail than is shown here.

survey (Fig. 6.6), the maximum angle of declination to the object would change from 90° to perhaps 89° or 88°. Assuming the observer's view is not obstructed, the perceived properties of the object and detection cues will barely change. There is continuity operating, so that $g(x)$ will be almost the same at $x = 0$ as at a small increment from zero. Given continuity, we maintain that it is not reasonable for $d(x)$ to be spiked (i.e. $d'(x) < 0$).

We turn our attention now to considering what an optimal $e_0(x)$ might be. We consider an aerial survey (Fig. 6.6), although the theoretical approach applies more generally. In Fig. 6.6, the angle of declination ψ is also the angle of incidence of vision, with $\pi/2 \geq \psi \geq 0$. If objects are assumed to be essentially flat and detection probability is proportional to the perceived area of an object, then the same object when moved further away shows less area and so is less detectable. The best you could achieve in this case is an innate detectability $d(x)$ proportional to

$$\cos(\theta) = \cos\left[\tan^{-1}\left(\frac{x}{h}\right)\right]$$

This only allows for the loss in perceived object area due to the oblique view of the object as ψ decreases. Using heuristic arguments, if we add the effect of perpendicular distance off the centreline and generalize the result, we get the **form**

$$d(x) = [1 - e^{-\lambda O}] \cdot \frac{\cos\left[\tan^{-1}\left(\frac{x}{\sigma}\right)\right]}{1 + \left(\frac{x}{\sigma}\right)^2}, \quad 0 < x \text{ and for some scale factor } \sigma,$$

as a plausible innate detection function. Here, O is the true area of the object. We would want λO such that $d(0) \doteq 1$, in which case the tail behaviour of $d(x)$ (i.e. as x gets large) is

$$d(x) \doteq \frac{1}{1 + \left(\frac{x}{\sigma}\right)}$$

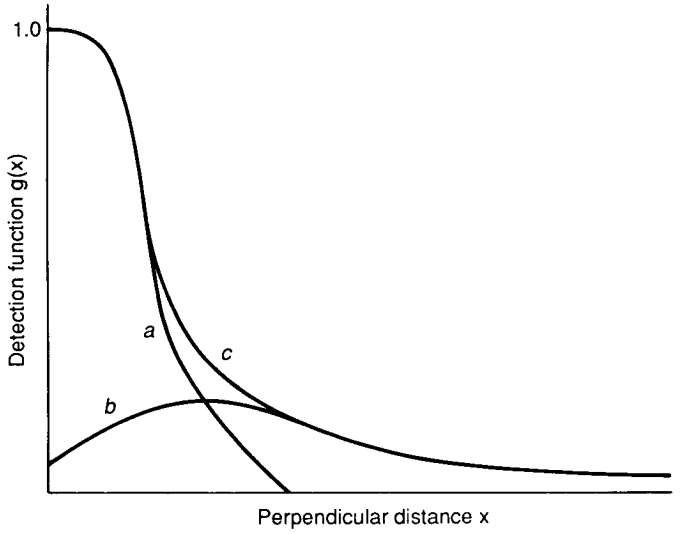

Fig. 6.5. The detection function c, which is the same as in Fig. 6.4, can arise from a mixture of curves a and b corresponding perhaps to two observers, one (a) 'guarding' the transect line and the other (b) scanning laterally; such minimal overlap of effort is undesirable.

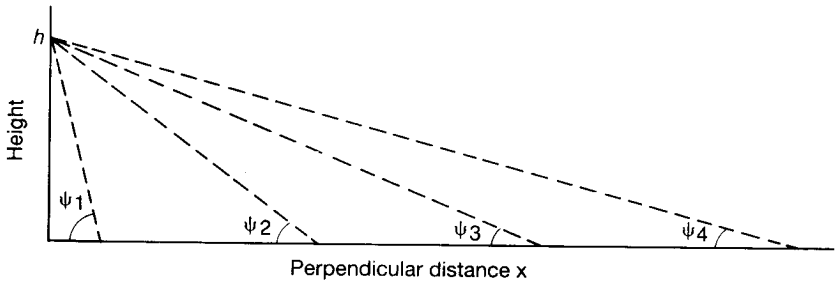

Fig. 6.6. The distance h represents (eye) height of an observer, and detections at various perpendicular distances x, indicated by the dashed lines, occur for angles of declination, ψ. For some types of visual cue, the cue strength depends critically upon ψ.

This is a very slow drop-off in detection probability. In fact, it cannot be used as a basis for general theory because it corresponds to $\int d(x)\,dx \to \infty$. This $d(x)$ does, however, give some sort of plausible upper bound on innate detectability, hence on possible effort. That is, we have reason to expect for any effort $e(x)$, properly scaled on x,

$$e(x) \leqslant \frac{\cos[\tan^{-1}(x)]}{1 + x^2}$$

Also, note that in this simple situation, innate detectability would have a definite shoulder.

Motivated by an aerial line transect mode of thinking, we could express our effort in terms of a distribution on the angle ψ (Fig. 6.6). To make derivations easier, we focus on u, $0 \leqslant u \leqslant 1$, where

$$u = \frac{2}{\pi} \cdot \psi = \frac{2}{\pi} \cdot \tan^{-1}\left(\frac{x}{h}\right)$$

and define $q(u)$ as the pdf of u. For greater generality, we use $u = \frac{2}{\pi} \cdot \tan^{-1}\left(\frac{x}{\sigma}\right)$; now the distribution of effort is proportional to

$$\frac{q[u(x)]}{\frac{dx}{du}}$$

where dx/du is evaluated at $u = \frac{2}{\pi} \cdot \tan^{-1}\left(\frac{x}{\sigma}\right)$, giving

$$e(x) \propto \frac{\frac{2}{\pi} \cdot q\left[\frac{2}{\pi} \cdot \tan^{-1}\left(\frac{x}{\sigma}\right)\right]}{\sigma\left[1 + \left(\frac{x}{\sigma}\right)^2\right]}, \qquad 0 \leqslant x$$

The proportionality constant is determined by the convention that $e(0) = 1$.

Let the effort be uniformly distributed over ψ, so that $q(u) \equiv 1$ for all u and

$$e(x) = \frac{1}{\left[1 + \left(\frac{x}{\sigma}\right)^2\right]}, \qquad 0 \leqslant x$$

If effort is uniform on $\cos(\psi)$, then we spend more time looking away from the centreline, and the result is

$$e(x) = \frac{\cos\left[\tan^{-1}\left(\frac{x}{\sigma}\right)\right]}{1 + \left(\frac{x}{\sigma}\right)^2}$$

It is interesting that if either ψ or $\cos(\psi)$ is uniform, the tail behaviour of the induced effort distribution is

$$e(x) \to \frac{1}{\left[1 + \left(\frac{x}{\sigma}\right)^2\right]} \to \left(\frac{\sigma}{x}\right)^2$$

Note that for the hazard-rate model of $g(x)$ for large x

$$g(x) \to \left(\frac{\sigma}{x}\right)^b$$

which of course includes the case of $b = 2$. Because effort decreases at large perpendicular distances, we would expect the applicable b to be $\geqslant 2$.

It is also useful to consider total effort, E, and its likely influence on $g(x)$ near $x = 0$. Now Equation 6.11 must be replaced by the more coherent form (justified by a hazard-rate argument)

$$g(x) = 1 - e^{-E \cdot e(x)} \tag{6.12}$$

with $e(0) = 1$, but where $e(x)$ is not mathematically identical to the $e(x)$ function considered above. Consider what happens at $x = 0$ as a function of total effort, E: $g(0) = 1 - \exp(-E)$. Some values of $g(0)$ as E increases are as follows:

E	$g(0)$
1	0.6321
2	0.8647
4	0.9817
8	0.9997
10	1.0000

It is obvious upon reflection, as the above illustrates, that if effort is inadequate, detection probability on the line can be less than unity even if innate detection probability at $x = 0$ is one. More interesting is what might happen to a shoulder under inadequate detection effort. Analytically from Equation 6.12, we have

$$g'(x) = \{1 - g(x)\} \cdot E \cdot e'(x)$$

and with E finite, we can write

$$g'(0) = (1 - g(0)) \cdot E \cdot e'(0)$$

If total effort is large enough to achieve $g(0) = 1$, we are virtually sure that $g'(0) = 0$, regardless of the shape of the relative effort, $e(x)$ (provided $e(x)$ is not pathologically spiked at $x = 0$, with $e'(0) = -\infty$). Also if $e(x)$ has a shoulder, then $g'(0) = 0$. This could occur with insufficient total effort, E, to ensure $g(0) = 1$, hence the presence of a shoulder in the data is no guarantee that $g(0) = 1$.

The case in which relative effort has no shoulder is interesting. As noted above, it is possible that $e'(0) < 0$ and yet $g'(0) = 0$. As an example, consider the spiked relative effort $e(x) = e^{-x}$ for $E = 15$, so that

$$g(x) = 1 - e^{-15e^{-x}}$$

and

$$g'(x) = (-e^{-15e^{-x}})(15e^{-x})$$

A few values are given below:

x	$g(x)$	$g'(x)$
0.0	1.0000	− 0.00001
0.1	1.0000	− 0.00002
0.5	1.0000	− 0.00102
1.0	0.9960	− 0.02215
1.5	0.9648	− 0.11778

Even though effort is spiked at $x = 0$, $g(x)$ has a distinct shoulder. However, if total effort is decreased, this shoulder will vanish and $g(0) = 1$ will fail.

The result illustrated above is due to a threshold effect. Once effort is large enough to achieve $g(0) = 1$, more effort cannot push $g(0)$ higher, but it can increase $g(x)$ for values of $x > 0$. We conclude that if there is sufficient total effort expended, then a shoulder is expected to be present even with a spiked relative effort function. The converse is disturbing: if total effort is too little, we can expect $g(0) < 1$, and there may be no shoulder. We emphasize the implications of guarding the centreline; if this is done, then as total effort decreases, more of the relative effort is likely to go near the centreline. This forces $e(x)$ to decrease more quickly, ultimately becoming spiked. The end result might be that we would have $g(0) < 1$, and $g(x)$ might be spiked (no shoulder). The data analysis implications are that if $\hat{g}(x)$ is, or is believed to be, spiked, then there is a basis to suspect that $g(0)$ is less than one. Conversely, if there is a shoulder, then there is a greater chance that $g(0) = 1$.

6.6 Fixed versus random sample size

6.6.1 Introduction

Theory and application of distance sampling has been almost exclusively in terms of fixed line lengths (and a fixed number of replicate lines) or fixed time spent at each point, for a fixed number of points. This