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

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

approach means that line lengths $l_1, \ldots, l_k$ (and k itself), and of course L, are *a priori* fixed measures of sampling effort; that is, they are known before traversing the transects. It is then sample size n (overall and per line) that is random. Similarly, for point transects, k is fixed, time spent at each point is fixed and number of detections is a random variable. In principle it is possible to do the reverse: fix the total sample size to be achieved and traverse the line(s) until that predetermined n is reached, or count at a point until a predetermined sample size is reached. This sampling scheme results in L, or time at the point, being a random variable.

The purpose of this section is to provide some results comparing the cases of random and fixed n, under simplistic but tractable assumptions, and to comment upon this alternative design. We conclude that the two schemes (fixed L and random n, or fixed n and random L), under some idealized conditions, are not importantly different in their statistical properties. Primarily, field (i.e. applied) considerations dictate the choice between sampling schemes.

A common example contrasting fixed and random effort sampling is provided by the (positive) binomial and negative binomial distributions. For the binomial distribution, sample size is fixed at n and we record the number of successes, $\tilde{y}$, in n independent binary trials. For the negative binomial, we fix the number of successful trials (y) and sample the binary events until y successes occur, so that the number of trials, $\tilde{n}$, is random. (The added notation, '~', is needed here to indicate which variable is random.) The corresponding probabilities, expectations and variances are given below for the positive and negative binomial cases respectively:

$$\Pr\{\tilde{y} = i | n\} = \binom{n}{i} p^i (1-p)^{n-i} \qquad \begin{aligned} E(\tilde{y}) &= np \\ \operatorname{var}(\tilde{y}) &= np(1-p) \end{aligned}$$

$$\Pr\{\tilde{n} = i + y | y\} = \binom{i+y-1}{y-1} p^y (1-p)^i \qquad \begin{aligned} E(\tilde{n}p) &= y \\ \operatorname{var}(\tilde{n}p) &= E(\tilde{n}p)(1-p) \end{aligned}$$

Despite the differences in the two sampling schemes, the sampling variances are essentially the same. In particular, with reference to a fixed n under the direct (binomial) sampling approach, if we could select y for the inverse sampling such that $y = np$, then both sampling methods would have the same sampling variance.

Moreover, the respective MLEs and their variances are, for the positive binomial,

$$\hat{p} = \frac{\tilde{y}}{n}, \qquad \text{var}(\hat{p}) = \frac{p(1-p)}{n}$$

and for the negative binomial,

$$\hat{p} = \frac{y}{\tilde{n}}, \qquad \text{var}(\hat{p}) = \frac{p(1-p)}{E(\tilde{n})}$$

Thus, again, if we design the inverse sampling so that $E(\tilde{n}) = n$, there is no important large sample difference between the two approaches.

Another example more related to distance sampling is use of randomly placed quadrats versus a sample of random points, with the data being distance to the nearest plant (e.g. Patil *et al.* 1979b). In quadrats, the area is fixed and counts are random. In nearest neighbour sampling, the plant count is fixed but the area sampled is random. Under an appropriate matching of the effort expended under the two schemes, the corresponding density estimates have almost equal large sample sampling variances when plants are randomly distributed (Holgate 1964).

We surmise that this relationship holds for most positive and negative sampling schemes, i.e. there exist pairs of schemes such that the sampling variance of the parameter estimator is almost the same under either approach. In line transects, we have either L as fixed and n as random, or we fix n and traverse a random line length until n detections are made. To be consistent with the usual definitions of direct (positive) and indirect (negative, or inverse) sampling, we label these as below:

Positive case	n fixed	$\tilde{L}$ random
Negative case	$\tilde{n}$ random	L fixed

Comparability of sampling variances requires that comparable effort be used in both schemes; this translates into the pair of relationships $E(\tilde{L}\,|\,n) = L$ and $E(\tilde{n}\,|\,L) = n$.

6.6.2 *Line transect sampling with fixed n and random $\tilde{L}$, under Poisson object distribution*

We examine here some properties of $\hat{D}$ under such comparable schemes assuming a homogeneous Poisson distribution of objects, a constant detection function everywhere in the sampled area (spatially invariant $g(x)$), and independent detections. For the (usual) L-fixed case under these assumptions, $\tilde{n}$ has a Poisson distribution with mean $2LD/f(0)$. For $\tilde{L}$ random, we assume a random starting point for the line and we

move along it until n detections are made. Thus, there are $n-1$ random inter-observational segments, of length $\tilde{l}_i$, which add to $\tilde{L}$. The first segment is from the starting point to the point perpendicular to the first detection. In general, the ith segment of length $\tilde{l}_i$ is the distance travelled between points perpendicular to detections $i-1$ and i, where $i=0$ is defined to be the starting point. Assuming that the number of objects in any area of size a, including the total area ($a=A$), is Poisson with mean aD, then it can be shown that $\tilde{l}$ is an exponentially distributed random variable with mean $E(\tilde{l}) = f(0)/(2D)$. The pdf of $\tilde{l}$ is

$$f_{\tilde{l}}(\tilde{l}) = \frac{2D}{f(0)} \cdot \exp\left[-\frac{2\tilde{l}D}{f(0)}\right]$$

By the assumptions we have made here, the $\tilde{l}_1, \ldots, \tilde{l}_n$ are independent. Because $\tilde{L}$ is the sum of independent exponential random variables, it is known that

$$\tilde{L} = \sum_{i=1}^{n} \tilde{l}_i$$

is distributed as a gamma $[n, f(0)/(2D)]$ distribution, so it has pdf

$$f_{\tilde{L}}(\tilde{L}) = \frac{\left[\frac{2D}{f(0)}\right]^n \cdot \tilde{L}^{n-1} \cdot \exp\left[-\frac{2\tilde{L}D}{f(0)}\right]}{(n-1)!}$$

It is also easily established that

$$E(\tilde{L} \mid n) = \frac{n \cdot f(0)}{2D}$$

which leads to the estimator

$$\tilde{D} = \frac{n \cdot \hat{f}(0)}{2\tilde{L}} \tag{6.13}$$

$\hat{f}(0)$ is computed conditional on n exactly as in the case of fixed L and random $\tilde{n}$, so $\hat{f}(0)$ is the same estimator in either sampling scheme.

Compare the estimator in Equation 6.13 to that when $\tilde{n}$ is random:

$$\hat{D} = \frac{\tilde{n} \cdot \hat{f}(0)}{2L} \tag{6.14}$$

Under a sampling theory approach, the two estimators have different expressions for small sample bias. For random $\tilde{L}$, from Equation 6.13,

$$E(\tilde{D}) = \frac{n}{2} \cdot E\left[\frac{\hat{f}(0)}{\tilde{L}} | n\right]$$

Given a Poisson distribution of objects and constant $g(x)$, it is reasonable to assume that $\tilde{l}$ and x are independent. Then the above becomes

$$E(\tilde{D}) = \frac{n}{2} \cdot E\left[\frac{1}{\tilde{L}} | n\right] \cdot E[\hat{f}(0) | n]$$

Under the gamma distribution of $\tilde{L}$,

$$E\left[\frac{1}{\tilde{L}} | n\right] = \left[\frac{1}{n-1}\right] \cdot \left[\frac{2D}{f(0)}\right]$$

which yields

$$E(\tilde{D}) = \frac{n}{n-1} \cdot D \cdot \frac{E[\hat{f}(0) | n]}{f(0)}$$

When L is fixed,

$$E(\hat{D}) = D \cdot \frac{E[\hat{f}(0) | n]}{f(0)}$$

so there is little difference between the two sampling schemes for large n in this example. For $\tilde{L}$ random, the bias associated with $1/\tilde{L}$ could be eliminated by using

$$\tilde{D} = \frac{(n-1) \cdot \hat{f}(0)}{2\tilde{L}}$$

This adjustment for bias when n is fixed and $\tilde{L}$ is random seems generally appropriate.

An asymptotic formula for the variance of $\tilde{D}$ in Equation 6.13 is

$$\text{var}(\tilde{D}) = \left[\frac{f(0) \cdot n}{2E(\tilde{L}|n)}\right]^2 \cdot \left[\frac{\text{var}(\tilde{L}|n)}{[E(\tilde{L}|n)]^2}\right] + \left[\frac{n}{2E(\tilde{L}|n)}\right]^2 \cdot \text{var}[\hat{f}(0)|n]$$

$$= D^2 \cdot \left[[\text{cv}(\tilde{L})]^2 + [\text{cv}\{\hat{f}(0|n)\}]^2\right] \tag{6.15}$$

For the Poisson distribution of objects and a spatially invariant $g(x)$, so that $\tilde{l}$ is exponential, we have

$$\text{var}(\tilde{l}) = [E(\tilde{l})]^2 = \left[\frac{f(0)}{2D}\right]^2$$

and

$$\text{var}(\tilde{L}|n) = n \cdot \left[\frac{f(0)}{2D}\right]^2$$

Using these results and Equation 6.15 gives

$$\text{var}(\tilde{D}) = D^2 \cdot \left[\frac{1}{n} + [\text{cv}\{\hat{f}(0|n)\}]^2\right]$$

The variance of $\hat{D}$ in Equation 6.14 is

$$\text{var}(\hat{D}) = D^2\left[\frac{1}{E(\tilde{n} \mid L)} + [\text{cv}\{\hat{f}(0|\tilde{n})\}]^2\right]$$

Under comparable effort, in which case $E(\tilde{L}|n) = L$ and $E(\tilde{n}|L) = n$, it is clear that for large samples, $\text{var}(\tilde{D}) \doteq \text{var}(\hat{D})$.

The condition under which the two sampling schemes have almost the same variance for estimated density is that the coefficients of variation for $\tilde{n}$ and $\tilde{L}$ are equal:

$$\frac{\text{var}(\tilde{n}|L)}{[E(\tilde{n}|L)]^2} = \frac{\text{var}(\tilde{L}|n)}{[E(\tilde{L}|n)]^2}$$

This relationship holds for the above case.

6.6.3 Technical comments

Assuming independent detections, then a general formula for the cumulative distribution function of $\tilde{l}$ is

$$F_{\tilde{l}}(\tilde{l}) = 1 - \sum_{i=0}^{\infty}\left[1 - \frac{1}{w \cdot f(0)}\right]^i \cdot \Pr\{\tilde{N} = i | a = 2w\tilde{l}\} \qquad (6.16)$$

The probability of moving distance $\tilde{l}$ and detecting no objects is $1 - F_{\tilde{l}}(\tilde{l})$. Assume the area examined for detections is of width $\pm w$ about the line. Then the unconditional probability of detecting an object is $1/[w \cdot f(0)]$. The event that there are no detections in distance $\tilde{l}$ happens if there are no objects in the area of size $2w\tilde{l}$, or if there is one object but it is not detected, or there are two objects and both remain undetected, and so forth. The joint probability that there are i objects in the area of size $2w\tilde{l}$ and all are undetected is (under the assumptions made)

$$\left[1 - \frac{1}{w \cdot f(0)}\right]^i \cdot \Pr\{\tilde{N} = i | a = 2w\tilde{l}\}$$

For $i = 0, 1, 2, \ldots$, these events are mutually exclusive, hence we get Equation 6.16. For the Poisson case,

$$\Pr\{\tilde{N} = i | a = 2w\tilde{l}\} = \frac{\exp(-2w\tilde{l}D) \cdot (2w\tilde{l}D)^i}{i!}$$

Thus

$$\begin{aligned} F_{\tilde{l}}(\tilde{l}) &= 1 - \sum_{i=0}^{\infty}\left[1 - \frac{1}{w \cdot f(0)}\right]^i \cdot \frac{\exp(-2w\tilde{l}D) \cdot (-2w\tilde{l}D)^i}{i!} \\ &= 1 - \exp[-2w\tilde{l}D] \cdot \exp\left\{\left[1 - \frac{1}{w \cdot f(0)}\right] \cdot (2w\tilde{l}D)\right\} \\ &= 1 - \exp[-2\tilde{l}D/f(0)] \end{aligned}$$

which is the cdf of an exponential distribution. Notice also that w drops out of $F_{\tilde{l}}(\tilde{l})$ in this example.

Closed form results can also be derived if a negative binomial distribution is assumed for $\Pr\{\tilde{N} = i | a = 2\tilde{l}w\}$ (Burnham *et al.* 1980: 197). However, we have not perceived a simple way to derive results for the random $\tilde{L}$ case without making strong assumptions about $\Pr\{\tilde{N} = i | a = 2w\tilde{l}\}$ and independence of detections.

Finally, we describe how to find $\Pr(\tilde{n}|L)$, as this is needed to compare the two schemes. Let $\tilde{N}$ be the number of objects in the searched strip of area $2wL$. The event $\tilde{n} = i$ arises as the sum of mutually exclusive events: $\tilde{N} = i$ and all i objects are detected, $\tilde{N} = i + 1$ and only i are detected, $\tilde{N} = i + 2$ and only i are detected, and so forth. The probability formula is

$$\Pr(\tilde{n}|L) = \sum_{\tilde{N}=i}^{\infty} \Pr\{\tilde{n} = i|\tilde{N}\} \cdot \Pr\{\tilde{N}|a = 2wL\} \tag{6.17}$$

For example, under the assumptions of a spatially constant detection function and independent detections, $\Pr\{\tilde{n} = i|\tilde{N}\}$ is binomial:

$$\Pr\{\tilde{n} = i|\tilde{N}\} = \binom{\tilde{N}}{i} \cdot \left[\frac{1}{w \cdot f(0)}\right]^{i} \cdot \left[1 - \frac{1}{w \cdot f(0)}\right]^{\tilde{N}-i}$$

For the Poisson case,

$$\Pr\{\tilde{N}|a = 2wL\} = \frac{\exp(-2wLD) \cdot (2wLD)^{\tilde{N}}}{\tilde{N}!}$$

Applying Equation 6.17 with these distributions gives

$$\Pr(\tilde{n}|L) = \frac{\exp[-2LD/f(0)] \cdot [2LD/f(0)]^{i}}{i!}$$

which is a Poisson distribution.

6.6.4 Discussion

Having fixed n and random $\tilde{L}$ is often not a practical design in line transect sampling. In particular, when planes or helicopters are used, you cannot set out to fly a random distance; $\tilde{L}$ cannot exceed the fuel capacity of the plane. For most methods of traversing the line(s), the distance to travel must be specified in advance. This also allows an

accurate cost estimate for a study, which is generally necessary. Further, in most studies, if effort ceased when the required sample size was reached, the observer would have to return to some point at the study boundary, and it would be wasteful not to seek and record detections for the maximum possible time in the study area.

Representative coverage of the area being sampled is also an important consideration. Lines, or points, are allocated so as to achieve a representative sampling of the area. This is critical to allow valid inference from the sample to the entire area. If $\tilde{L}$ was random, one might finish before completing the *a priori* determined sample of lines, or finish the sample of lines and still need to sample more. With random line length, it is difficult to assure a representative sample over the area of interest; thus there is more danger of substantial bias in $\tilde{D}$ due to unbalanced spatial coverage.

Point transect sampling is potentially more amenable to a fixed n strategy. The fixed n can be set on a per point basis. Then a representative sample of k points can be selected and every point can be visited. The amount of time at each point will vary. An upper bound could be put on time, leading to a mixed strategy: stay at a point until n detections are made, or until the maximum time is reached. There is potential to develop the theory for such a strategy. However, it is still not especially practical to ask a recorder to be aware of when n total detections are made, and then to stop effort. Also, should this n be a total for all species, or for one target species?

We have not here presented any theory for point transects with fixed n and random time, as that theory is more difficult to conceptualize. For the typical application to birds, detections depend on cue generation, which would have their own temporal distribution. This adds another level of complexity to the case of a scheme with fixed n and random time.

Even for a fixed line length scheme, there is information in the interobservational distances as defined here. For example, the $\tilde{l}_i$ may be used to assess the spatial distribution of objects (Burnham *et al.* 1980: 196–8). Under the (unlikely) hypothesis of a Poisson spatial distribution and constant $g(x)$, $\tilde{l}$ is an exponential random variable. There are many tests available for the null hypothesis that a random sample is from an exponential pdf. The distribution of $\tilde{l}$ under other object distributions can be determined by methods presented here. The information contained in the $\tilde{l}_i$ is reduced in practice, because they are likely to be serially correlated. However, if independence can be assumed, the information in the $\tilde{l}_i$ might be used to provide better estimates of the residual variation in the $\tilde{n}$; this subject may be worthy of study. The concept is that

$$\widehat{\text{var}}(\tilde{n}|L) = \left(\frac{n}{L}\right)^2 \cdot \widehat{\text{var}}(\tilde{L}|n)$$

might hold true. Given independence of the interobservational distances,

$$\widehat{\text{var}}(\tilde{L}|n) = \frac{\sum_{i=1}^{n} (\tilde{l}_i - \tilde{L}/n)^2}{n-1}$$

thus giving an alternative estimator of $\text{var}(\tilde{n}|L)$. The above variance estimator for $\tilde{L}$ also provides the basis for an empirical estimator of $\text{var}(\tilde{D})$ for the fixed n scheme if such sampling is practical and valid.

A final important comment is required about the relative merits of random and fixed line lengths. Often, lines are *a priori* of fixed (frequently equal) length, in which case all the fixed length (and random n) theory holds. However, designs or field practice often result in unequal line lengths, for example when lines are placed at random but then cross from one side to the other of an *a priori* defined area, or when bad weather causes effort termination during a ship survey, so that a transect is shorter than intended. Either of these instances gives the appearance of some stochasticity in line length, hence one might consider that the set of lines has a random component that should be accounted for in variances (and perhaps biases) of estimates (Seber 1979). We disagree with this thinking; it is entirely appropriate to condition on achieved line length in these and other cases, provided the stochastic variations in length are unrelated to the density of objects, or if it is not possible to fit a model that relates variation in line length to object density.

It does not follow that random line length theory applies, simply because the survey design or field protocol results in varying line lengths. The theory applies only if there is information about density D in the probability distribution of line lengths. Even in the case of randomly placed lines running across a predefined area, there is no information about object density in the probabilistic distribution of line length by itself. Moreover, in this case the line lengths are known before they are ever traversed, hence there is every reason in theory to consider line length as a fixed ancillary (i.e. it affects the precision of $\hat{D}$ but contains no direct information about D) in all the usual designs and field protocols.

Once we consider line lengths as known and fixed prior to data collection, or after the fact we condition on actual line lengths when appropriate, then some potential statistical methods are not relevant. For example, it is not relevant to apply finite population sampling ratio

estimation theory to encounter rate; such ratio estimation theory leads to slightly different formulae for var(n) than we give in this book. The key point here is that line lengths l_i may often differ; this does not make them random in any sense that concerns us, especially if we know the actual line lengths before they are traversed. It is proper to take these line lengths as fixed unless there is a probability distribution on possible line lengths which depends on the density parameter of interest. The latter is only likely to be true under the scheme of fixed n, in which random linear effort continues until n detections are made, or under adaptive sampling schemes, in which sampling effort increases when areas of high density are found.

6.7 Efficient simulation of distance data

6.7.1 The general approach

To produce simulated distance data requires the Monte Carlo generation of sample size n, detection distances $y = x$ or r, and for the clustered case, cluster sizes s. The efficient way to do this is first to generate sample size according to some discrete distribution, $p(n)$, then generate n distances and cluster sizes based on the bivariate sampling distribution of distance and s. The alternative is first to generate spatially distributed clusters, and independently for each cluster, a cluster size s. Then determine for each cluster whether it gets detected according to some detection function, $g(y \mid s)$. This method is indirect and inefficient. The purpose of this section is twofold: to show how to simulate distance data directly and to outline explicitly how the simulated data of Chapters 4 and 5 were generated.

The following general approach is recommended. First, decide on a detection function; it will be bivariate if cluster size is to vary and there is to be size-biased detection. Otherwise, $g(y)$ depends only on distance. For the clustered case, decide on a probability distribution, $\pi(s)$, of cluster sizes in the yet-to-be-sampled population. Also select a sampling distribution for sample size, $p(n)$, such as Poisson or negative binomial. It is then possible to specify the exact parameterization of $p(n)$. To simulate data for k replicate lines or points, first generate independent sample sizes $n_1, \ldots, n_k$ according to $p(n)$. If objects do not occur in clusters, just generate n_i independent distances, $i = 1, \ldots, k$, according to the probability density function of distances, $f(y)$. This function is determined by, and computed from, $g(y)$, and the context (line or point transects). If cluster size varies but detection is independent of size, then for each generated distance y, produce independently a value of the random variable s according to $\pi(s)$.

The case of size-biased detection requires a two step process of either generating y from its marginal density function, then s from the size-biased distribution $\pi^*(s|y)$, or the reverse (which we recommend): generate an observed cluster size from the marginal size-biased distribution of detections, $\pi^*(s)$, then generate y from $f(y|s)$. The detailed theory for these distributions, given $g(y|s)$ and $\pi(s)$, is in Section 3.6.6.

The distribution of n, $p(n)$, must have, at least implicitly, $E(n)$ as one of its parameters, where

$$E(n) = \frac{2LD_s}{f(0)} \quad \text{(line transects)}$$

$$E(n) = \frac{2\pi k D_s}{h(0)} \quad \text{(point transects)}$$

(π by itself refers to 3.14159...). The density D_s denotes density of clusters. We use $f(0)$ and $h(0)$ in these representations to facilitate the case without truncation ($w = \infty$).

We now summarize quantities that must be specified to simulate distance data. These quantities are interrelated and hence cannot be independently set; in particular, we recommend that either $E(n)$ or sampling effort (L or k) is specified, and the other quantity is computed. Constants to be specified are w and k, and for line transects, line lengths l_i, $i = 1, \ldots, k$, which sum to L. Separately specified parameters are D_s and $E(s)$, and any additional parameters in $p(n)$ other than $E(n)$. Fundamental distributions to specify are $p(n)$ and $\pi(s)$. Finally, there is the detection function, $g(y|s)$, which, in conjunction with $\pi(s)$, determines the sampling distributions of y and s. In general we would need to compute $f(0)$, $h(0)$, $\pi^*(s)$, $f(x|s)$ and $f(x|s)$ numerically. From Section 3.6.6, formulae necessary in simulation of line transect data are

$$f(0|s) = \frac{1}{\int_0^w g(x|s)\,dx} \tag{6.18}$$

$$f(0) = \frac{1}{\sum_{s=1}^{\infty} \frac{\pi(s)}{f(0|s)}} \tag{6.19}$$

$$\pi^*(s) = \left[\frac{f(0)}{f(0|s)}\right]\pi(s) \tag{6.20}$$

and

$$f(x|s) = f(0|s) \cdot g(x|s) \tag{6.21}$$

Formulae necessary in simulation of point transect data are

$$h(0|s) = \frac{1}{\int_0^w r \cdot g(r|s)\,dr} \tag{6.22}$$

$$h(0) = \frac{1}{\sum_{s=1}^{\infty} \frac{\pi(s)}{h(0|s)}} \tag{6.23}$$

$$\pi^*(s) = \left[\frac{h(0)}{h(0|s)}\right]\pi(s) \tag{6.24}$$

and

$$f(r|s) = h(0|s) \cdot r \cdot g(r|s) \tag{6.25}$$

If the distribution of n is assumed to be Poisson, then

$$p(n) = \frac{e^{-E(n)}[E(n)]^n}{n!} \tag{6.26}$$

A useful parameterization of the negative binomial is

$$p(n) = \frac{\Gamma(\theta + n)}{\Gamma(\theta)\,\Gamma(n+1)}(1-\tau)^n \cdot \tau^{\theta}, \quad 0 < \theta,\ 0 < \tau < 1,\ 0 \leqslant n \tag{6.27}$$

which has

$$E(n) = \theta \cdot \frac{1-\tau}{\tau}$$

and

$$\text{var}(n) = \frac{E(n)}{\tau}$$

In point transects all with a fixed observation time, τ and θ can be the same over different points (within a stratum). For line transects, the l_i usually vary, and we recommend keeping τ constant while letting θ vary by line length, because this gives coherent results: $n_1 + \ldots + n_k$ is then distributed as negative binomial with parameters τ and $\theta_1 + \ldots + \theta_k$. Thus, we can arbitrarily vary the line lengths and still have sample sizes (individually and totals) as negative binomial random variables. Under this strategy, $1/\tau$ is the variance inflation factor relative to a Poisson (random) spatial distribution of objects. For the case of $\tau = 1$, use the Poisson distribution for Monte Carlo generation of sample sizes.

Consider the line transect case with one long line (i.e. ignore replicate lines) of length L and objects not clustered. We would first specify $g(x)$, then compute $f(0)$, by numerical integration if need be. It is important to keep straight the units of measurement in a simulation, because with real data, detection distances and line length are often in different units, such as metres and kilometres. Also, the units of $f(0)$ are the reciprocal of the distance units used for x. Say we get $f(0) = 10$, with units on x taken as kilometres. Then effective strip width is 0.1 km or 100 m.

In this hypothetical example, next we specify $D_s = 2$ clusters/km^2 and $E(n) = 70$. Now we determine L from $E(n) = 2LD_s/f(0)$: $70 = 2 \cdot 2 \cdot L/10$, or $L = 175$km. If we want n to be Poisson, then we generate it from a Poisson with mean 70. Given n, generate $x_1, \ldots, x_n$ from the pdf $f(x) = f(0) \cdot g(x)$.

We illustrate the approach in more detail for the simulation generation of examples in Chapters 4 and 5.

6.7.2 The simulated line transect data of Chapter 4

The half-normal bivariate detection function may be taken as

$$g(x \mid s) = \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma(s)}\right)^2\right]$$

where we model the scale parameter, σ, as a function of cluster size (c.f. Quinn 1979; Drummer and McDonald 1987; Ramsey *et al.* 1987; Otto and Pollock 1990). In particular, the form $\sigma(s) = \sigma \cdot s^{\alpha}$ has been much used. We think this is a reasonable model for data analysis; however, for simplicity of theory, we used a linear form for line transects:

$$\sigma(s) = \sigma_0\left[1 + b \cdot \frac{s - E(s)}{E(s)}\right] \tag{6.28}$$

subject to the constraint

$$b \leq \frac{E(s)}{E(s) - 1}$$

and assuming $0 \leq b$, although to a limited extent, negative values of b are mathematically possible. This form is also suggested by us because it puts the problem into the framework of generalized linear models (McCullagh and Nelder 1989). The case $b = 0$ corresponds to detection probability independent of cluster size. For $w = \infty$, $f(0 \mid s)$ and $f(x \mid s)$ are closed form:

$$f(0 \mid s) = \left[\sqrt{\frac{2}{\pi}}\right] \cdot \frac{1}{\sigma(s)}$$

$$f(x \mid s) = \left[\sqrt{\frac{2}{\pi}}\right] \cdot \frac{1}{\sigma(s)} \cdot \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma(s)}\right)^2\right] \tag{6.29}$$

Applying Equation 6.19, we have

$$f(0) = \left[\sqrt{\frac{2}{\pi}}\right] \cdot \frac{1}{\sum_{s=1}^{\infty} \pi(s) \cdot \sigma(s)} = \left[\sqrt{\frac{2}{\pi}}\right] \cdot \frac{1}{E\{\sigma(s)\}}$$

The form of $\sigma(s)$ in Equation 6.28 is convenient because we can explicitly evaluate its expectation with respect to $\pi(s)$; in fact, $E\{\sigma(s)\} = \sigma_0$ for any value of the parameter b. Thus, for any extent of size bias under this model,

$$f(0) = \left[\sqrt{\frac{2}{\pi}}\right] \cdot \frac{1}{\sigma_0}$$

From Equation 6.20,

$$\pi^*(s) = \left[\frac{f(0)}{f(0 \mid s)}\right] \pi(s) = \frac{\sigma(s)}{\sigma_0} \pi(s)$$

$$= \left[1 + b \cdot \frac{s - E(s)}{E(s)}\right] \pi(s) \quad s = 1, 2, \ldots \tag{6.30}$$

The expected value of s in the sample of detected clusters is

$$E^*(s) = \sum s \cdot \pi^*(s) = E(s) + b \cdot \frac{\text{var}(s)}{E(s)}$$

A simple choice for $\pi(s)$ is to let $s - 1$ have a Poisson distribution with mean $E(s) - 1$:

$$\pi(s) = \frac{e^{-[E(s)-1]} \cdot [E(s) - 1]^{s-1}}{(s-1)!} \qquad s = 1, 2, \ldots$$

We used this $\pi(s)$ in the Chapter 4 examples and then created a table (in the computer) of the values of $\pi^*(s)$, from Equation 6.30, for specified $E(s)$, b and σ_0. Then a value of a detected cluster size was generated by standard Monte Carlo methods. Given a value of s, then x was generated according to the distribution of Equation 6.29. This was done by generating a χ^2 variate on 1 df and calculating $x = \sigma(s) \cdot \sqrt{\chi^2}$.

The distribution of sample detections in Chapter 4 was negative binomial, set up with $\tau = 0.4$, so that the variance inflation factor is 2.5. The choice of 12 replicate lines was arbitrary. Other choices made: $\sigma_0 = 10\text{m}$ and $E(n) = 96$. It was then convenient to use $L = 48\text{km}$ and keep the encounter rate constant over replicate lines (whose lengths varied). These choices and decisions produce, by design, the result

$$E(n_i) = \theta_i \frac{1 - \tau}{\tau} = 2l_i \qquad i = 1, \ldots, 12$$

which implies $\theta_i = 64l_i/48$.

In metres, $f(0) = 0.0798$, hence with a conversion factor to units of per km^2, density of clusters is

$$D = \frac{96 \times 0.0798}{2 \times 48} \times 1000 = 79.8 \text{ clusters/km}^2$$

The simulation was set up in such a way that density of individuals is $79.8 \cdot E(s) = 239.4$ regardless of the value of b; b only determines the degree of size bias. Values of b used were 0 and 1 (with the same set of $n_1, \ldots, n_{12}$).

6.7.3 The simulated size-biased data of Chapter 5

The generation of the simulated data for point transects with size bias used the same half-normal bivariate detection function as for the line transect case:

$$g(r|s) = \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma(s)}\right)^2\right]$$

However, the relevant formulae are now Equations 6.22–6.25. In particular, we have for $w = \infty$,

$$h(0|s) = \frac{1}{[\sigma(s)]^2}$$

and

$$f(r|s) = \frac{r}{[\sigma(s)]^2} \cdot \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma(s)}\right)^2\right] \tag{6.31}$$

Using Equation 6.23 with the above expression for $h(0|s)$, we have

$$h(0) = \frac{1}{\sum_{s=1}^{\infty} \pi(s) \cdot \{\sigma(s)\}^2} = \frac{1}{E[\{\sigma(s)\}^2]}$$

Thus we choose to parameterize $\sigma(s)$ as

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

subject to the constraint

$$b \leqslant \frac{E(s)}{E(s) - 1}$$

This model gives

$$h(0) = \frac{1}{\sigma_0^2}$$

and

$$h(0|s) = \frac{1}{\sigma_0^2 \cdot \left[1 + b \cdot \frac{s - E(s)}{E(s)}\right]}$$

Hence, from Equation 6.24,

$$\pi^*(s) = \left[\frac{h(0)}{h(0|s)}\right]\pi(s) = \frac{\sigma(s)}{\sigma_0} \cdot \pi(s)$$

$$= \left[1 + b \cdot \frac{s - E(s)}{E(s)}\right]\pi(s) \qquad s = 1, 2, \ldots$$

which is the identical $\pi^*(s)$ in the line transect case in Equation 6.30.

For the Chapter 5 size-biased example data, we choose $\pi(s)$ to be the geometric distribution,

$$\pi(s) = (1 - \beta)^{s-1} \cdot \beta \qquad 0 < \beta < 1 \qquad s = 1, 2, \ldots$$

For this distribution, $E(s) = 1/\beta$. From the above expression for $\pi^*(s)$, we have for this example

$$\pi^*(s) = \left[1 + b \cdot \frac{s - E(s)}{E(s)}\right] \cdot (1 - \beta)^{s-1} \cdot \beta \qquad s = 1, 2, \ldots \qquad (6.32)$$

Note that $b = 1$ corresponds to considerable size bias and gives the simple form

$$\pi^*(s) = s \cdot (1 - \beta)^{s-1} \cdot \beta^2 \qquad s = 1, 2, \ldots$$

For the example in Section 5.8, we used $E(s) = 1.85$, hence $\beta = 0.54054$, and $b = 0.75$. These values serve to specify $\pi^*(s)$ completely. Also, we set $\sigma_0 = 30$m, which, together with $b = 0.75$, serves to specify $g(r|s)$, $h(0|s)$ and $f(r|s)$. The latter is given by Equation 6.31; that density function has cumulative distribution function

$$f(r|s) = 1 - \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma(s)}\right)^2\right]$$

Consequently, for this example, a random r, given an s drawn from Equation 6.32, was produced as

$$r = -2 \cdot \sigma(s) \cdot \log_e(1 - u)$$

where u is a uniform pseudo-random variable on the interval 0 to 1.

The variation in counts, n, was generated from the negative binomial distribution with variance inflation (dispersion) factor set at 2.65, so that $\tau = 0.37736$. The encounter rate per point was set at 1.6 and then k was set at 60 points to give, overall, $E(n) = 96$. In terms of Equation 6.27 and associated results, this means that on a per point basis (which is how the random counts are generated), we specified

$$E(n_i) = \theta \frac{1 - \tau}{\tau} = 1.6$$

and

$$\text{var}(n_i) = \frac{E(n_i)}{\tau} = 2.65\, E(n_i) \qquad i = 1, \ldots, k$$

Hence, τ is as above and $\theta = 0.52416$. As a consequence of the choice of model and parameters in this example, the density of clusters is

$$\frac{96 \times [1/30]^2}{2\pi \times 60} \times 1\,000\,000 = 283 \text{ clusters/km}^2$$

and density of individuals is $1.85 \times 2.83 = 523$ objects/km^2.

6.7.4 Discussion

We have recommended simulating the pair (y, s) by generating s from its marginal distribution $\pi^*(s)$, and then y from the conditional distribution, $f(y|s)$. The algebra for this was straightforward in the above two examples. Consider, however, the reverse process for the point transect example above: generate r from $f(r)$ then s from $\pi^*(s|r)$. The relevant formulae are

$$f(r) = \frac{1}{\sigma_0^2}\left[\sum_{s=1}^{\infty}(1-\beta)^{s-1} \cdot \beta \cdot r \cdot \exp\left\{-\frac{1}{2}\left(\frac{r}{\sigma(s)}\right)^2\right\}\right]$$

and

$$\pi^*(s|r) = \frac{(1-\beta)^{s-1} \cdot \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma(s)}\right)^2\right]}{\sum_{s=1}^{\infty}(1-\beta)^{s-1} \cdot \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma(s)}\right)^2\right]}$$

Use of these formulae would entail much more computing than the use of $\pi^*(s)$ and $f(r|s)$. Heuristically, this is because there is only a finite (and small, usually) number of possible values for s, whereas infinitely many values of r can occur. Therefore, with each new r, one must recompute the entire distribution $\pi^*(s|r)$ before s can be generated.

In some real applications, cluster sizes potentially range from one to thousands, for example dolphin surveys on some species. To simulate the essence of such applications, it is not necessary for s to vary over this set of values. A set of a hundred (or fewer) values should suffice (e.g. s taking values 1–10, 15, 20, 30, 50, 75, 100–900 by 100, 1000–5000 by 500). Keeping the range set of s small will greatly speed up simulations.

Closed form expressions underlying simulations will be the exception. Even in the above line transect examples, if we take w as finite, numerical integration must be used to find the necessary quantities given by Equations 6.18 to 6.21. Expect to use numerical integration; fortunately for purposes of simulation, the numerical methods need not be highly sophisticated.

Sometimes we only want to explore statistical properties of estimators of $f(0)$, $h(0)$, $g(x|s)$, $g(r|s)$ and $E(s)$, and not properties of $\hat{D}_s$ and $\hat{D}$. In this event, it is not necessary to generate a random sample size n for each replicate. In fact, it is better to fix n and do, say, 1000 replicates at that n, and repeat the process over a set of values of n.

6.8 Thoughts about a full likelihood approach

6.8.1 Introduction

In principle, analysis of distance data could be based on a full likelihood, $\mathcal{L}(D,\underline{\theta})$, for all data components. The focus is on average density D in the study area; we represent all the other parameters by $\underline{\theta}$. These other parameters appear in the probability components for n, $y = x$ or r, and s. An advantage of having a full likelihood is that it allows the computation of profile likelihood intervals for D. The disadvantages are the need to specify probability models for n and s. We have avoided assuming any probability model for n by using an empirical estimator of var(n), and getting confidence intervals assuming $\hat{D}$ is log-normally distributed. Similarly, a point estimator and sampling variance of $E(s)$ can be obtained in a regression framework, so no probability model is required for $\pi(s)$ or $\pi^*(s)$ (the distribution of s in the entire population, and in the detected sample, respectively). Probability models, and likelihood inference, have only been used for the distance part of the data,

and we use generalized approaches to ensure robust inference for n and s. The purpose of this section is to demonstrate how a full likelihood approach could be used, to make some comparisons of profile likelihood intervals to traditional and log-based intervals, and to comment on other advantages of a full parametric (likelihood) approach.

6.8.2 Full likelihood for line transects: simple examples

(a) *Half-normal $g(x)$, Poisson n* Assume that objects are spatially distributed as a homogeneous Poisson process, that the detection function is half-normal, that $w = \infty$, and that objects are single entities (i.e. we take $s = 1$). Then data from replicate lines may be collapsed into just the total count, n, for total line length L, and the perpendicular distances $x_1, \ldots, x_n$. The detection function $g(x)$ and pdf $f(x)$ are

$$g(x) = \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}$$

and $$f(x) = \left[\sqrt{\frac{2}{\pi\sigma^2}}\right] \cdot \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\} \qquad 0 \leqslant x < \infty,\ 0 < \sigma$$

The probability distribution of n is

$$\Pr(n) = \frac{\exp\{-2LD/f(0)\} \cdot \{2LD/f(0)\}^n}{n!}$$

(Section 6.6). We have

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

so there are only two parameters, σ and D, although it is sometimes simpler to leave $f(0)$ in the formulae. The full likelihood is

$$\mathcal{L}(D, \sigma) = \Pr(n) \cdot \left[\prod_{i=1}^{n} f(x_i)\right]$$

$$= \frac{\exp\{-2LD/f(0)\} \cdot \{2LD/f(0)\}^n}{n!} \cdot \prod_{i=1}^{n}\left[\left(\sqrt{\frac{2}{\pi\sigma^2}}\right) \cdot \exp\left\{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2\right\}\right]$$

We define

$$T = \sum_{i=1}^{n} x_i^2$$

and simplify $\mathcal{L}(D,\sigma)$ by collapsing terms where possible and by dropping some multiplicative constants, giving

$$\mathcal{L}(D, \sigma) = \exp\{-LD\sigma\sqrt{(2\pi)}\} \cdot \{LD\sigma\sqrt{(2\pi)}\}^n \cdot \left[\frac{1}{\sigma^n} \cdot \exp\left\{-\frac{T}{2\sigma^2}\right\}\right]$$

or simplified as much as possible,

$$\mathcal{L}(D,\sigma) = \exp\left\{-\left(LD\sigma\sqrt{(2\pi)} + \frac{T}{2\sigma^2}\right)\right\} \cdot D^n \tag{6.33}$$

Note that we ignore multiplicative constants in the likelihood function.

The joint MLEs from Equation 6.33 are

$$\hat{\sigma} = \sqrt{\frac{T}{n}}$$

and

$$\hat{D} = \frac{1}{L}\sqrt{\frac{n}{2\pi T}} \equiv \frac{n \cdot \hat{f}(0)}{2L}$$

Standard likelihood theory can be used to derive the theoretical variance of $\hat{D}$, which may be expressed in a variety of ways:

$$\text{var}(\hat{D}) = \frac{1.5D \cdot f(0)}{2L} = D^2 \cdot \left[\frac{1.5}{E(n)}\right] = D^2 \cdot \left[\frac{1}{E(n)} + \frac{1}{2E(n)}\right]$$

$$= D^2\left[\{\text{cv}(n)\}^2 + \{\text{cv}(\hat{f}(0))\}^2\right]$$

Thus, these results are all exactly the same as what are derived by the 'hybrid' method of using $E(n) = 2LD/f(0)$, the Poisson variance of n, and a likelihood only for the distance data.

In general, the hybrid approach with empirical estimation of var(n) will be almost fully efficient for $\hat{D}$ and is more robust as no distribution

need be assumed for n. One advantage of a full likelihood approach is the possibility of using a profile likelihood interval for D. Such intervals can be expected to perform better than $\hat{D} \pm 1.96\,\widehat{\text{se}}(\hat{D})$ or log-based intervals because the likelihood function encodes information about the sampling distribution of $\hat{D}$, thus allowing for non-normality of $\hat{D}$ (or $\log_e(\hat{D})$). Below, we give some general explanation of profile likelihoods, then derive the profile log-likelihood function for D for the above example.

(b) *Profile likelihood intervals* Let $\mathcal{L}(D, \underline{\theta})$ be the full likelihood such as that given in Equation 6.33. The profile likelihood is symbolically

$\mathcal{L}(D, \hat{\underline{\theta}}(D))$ = maximum value of $\mathcal{L}(D, \underline{\theta})$ over $\underline{\theta}$ for any given value of D

This is then a function of just the single parameter D. For computing profile likelihood intervals, it is convenient to use the following function:

$$\phi(D) = 2\,\{\log_e \mathcal{L}(\hat{D}, \hat{\underline{\theta}}) - \log_e \mathcal{L}(D, \hat{\underline{\theta}}(D))\} \tag{6.34}$$

In Equation 6.34, $\mathcal{L}(\hat{D}, \hat{\underline{\theta}})$ is equivalent to $\mathcal{L}(\hat{D}, \hat{\underline{\theta}}(\hat{D}))$, where $\hat{D}$ is the MLE of D. The function $\phi(D)$ is a pivotal quantity, asymptotically distributed as a single degree of freedom chi-square, χ^2_1. This approximate distribution of $\phi(D)$ holds better at small sample sizes than the assumed normality of $\hat{D}$ underlying the use of $\hat{D} \pm 1.96\,\widehat{\text{se}}(\hat{D})$. A $100(1 - \alpha)\%$ profile confidence interval for D is given as the set of all values of D such that $\phi(D) \leqslant \chi^2_1(\alpha)$, where $\chi^2_1(\alpha)$ is the $1 - \alpha$ percentile point of the χ^2_1 distribution (3.84 for a 95% confidence interval). We only need the interval endpoints, which are the two solutions to the equation

$$\phi(D) = 2\,\{\log_e \mathcal{L}(\hat{D}, \hat{\underline{\theta}}) - \log_e \mathcal{L}(D, \hat{\underline{\theta}}(D))\} = \chi^2_1(\alpha) \tag{6.35}$$

Barndorff-Neilson (1986) described the theory underlying the method, including ways to improve on the approximation of $\phi(D)$ as a χ^2_1 random variable.

(c) *Profile formulae, half-normal $g(x)$, Poisson n* Starting with the likelihood in Equation 6.33, we first must find the maximum for σ given any fixed value of D. The steps are summarized below:

$$\log_e \mathcal{L}(D, \sigma) = -\,LD\sigma\sqrt{(2\pi)} - \frac{T}{2\sigma^2} + n \cdot \log_e(D)$$

and

$$\frac{\partial \log_e \mathcal{L}(D, \sigma)}{\partial \sigma} = -LD\sqrt{(2\pi)} + \frac{T}{\sigma^3} = 0$$

From the above equation,

$$\hat{\sigma}(D) = \left[\frac{T}{LD\sqrt{(2\pi)}}\right]^{1/3}$$

The joint MLEs $\hat{D}$ and $\hat{\sigma}$ are given above, hence finding the expression in Equation 6.34 is now merely algebraic manipulation:

$$\log_e \mathcal{L}(D, \hat{\sigma}(D)) = -\frac{3n}{2}\left[\frac{D}{\hat{D}}\right]^{2/3} + n \cdot \log_e(D)$$

from which

$$\log_e \mathcal{L}(\hat{D}, \hat{\sigma}) = -\frac{3n}{2} + n \cdot \log_e(\hat{D})$$

Reduced to a very simple form, we get for this example

$$\phi(D) = 3n \cdot \left[(D/\hat{D})^{2/3} - 1 - \log_e \{(D/\hat{D})^{2/3}\}\right] \tag{6.36}$$

To get a profile likelihood interval for D, we substitute the values of $\hat{D}$ and n in Equation 6.36, tabulate $\phi(D)$ for a range of D, and pick off the two solutions to Equation 6.35. It can be useful to plot $\phi(D)$, as is shown in another context by Morgan and Freeman (1989).

Below we look at some numerical examples comparing different confidence intervals. However, first we determine Equation 6.34 explicitly for a few more examples. These, and the above, are overly simplistic compared to real data, but only very simple cases lead to analytical, or even partially analytical, solutions for $\phi(D)$.

(d) *Negative exponential* $g(x)$, *Poisson* n The negative exponential $g(x)$ is not a desirable detection function, but for $w = \infty$ and a Poisson n, we can derive closed form results for this case. Some formulae are

$$g(x) = \exp\left\{-\frac{x}{\lambda}\right\}$$

$$f(x) = \frac{1}{\lambda} \cdot \exp\left\{-\frac{x}{\lambda}\right\} \qquad 0 \leqslant x < \infty,\ 0 < \lambda$$

and $f(0) = 1/\lambda$. The full likelihood is

$$\mathscr{L}(D, \lambda) = \frac{\exp\{-2LD/f(0)\} \cdot \{2LD/f(0)\}^n}{n!} \cdot \prod_{i=1}^{n} \left[\frac{1}{\lambda} \cdot \exp\left\{-\frac{x_i}{\lambda}\right\}\right]$$

Defining

$$T = \sum_{i=1}^{n} x_i$$

we simplify $\mathscr{L}(D, \lambda)$ to

$$\mathscr{L}(D, \lambda) = \exp\left\{-\left(2LD\lambda + \frac{T}{\lambda}\right)\right\} \cdot D^n$$

The joint MLEs are $\hat{\lambda} = \dfrac{T}{n}$ and $\hat{D} = \dfrac{n^2}{2LT} \equiv \dfrac{n \cdot \hat{f}(0)}{2L}$

From likelihood theory,

$$\operatorname{var}(\hat{D}) = \frac{D}{L\lambda} = D^2 \cdot \left[\frac{2}{E(n)}\right] \equiv D^2\left[\{\operatorname{cv}(n)\}^2 + \{\operatorname{cv}(\hat{f}(0))\}^2\right]$$

Fixing D in $\mathscr{L}(D, \lambda)$, we find $\hat{\lambda}(D)$ as follows:

$$\log_e \mathscr{L}(D, \lambda) = -2LD\lambda - \frac{T}{\lambda} + n \cdot \log_e(D)$$

and

$$\frac{\partial \log_e \mathscr{L}(D, \lambda)}{\partial \lambda} = -2LD + \frac{T}{\lambda^2} = 0$$

so that

$$\hat{\lambda}(D) = \sqrt{\frac{T}{2LD}}$$

Finally, we derive

$$\phi(D) = 4n \cdot \left[(D/\hat{D})^{1/2} - 1 - \log_e \{(D/\hat{D})^{1/2}\} \right] \tag{6.37}$$

(e) *Negative exponential* $g(x)$, *Poisson* n, *Poisson* s To the above example, we add the feature of varying cluster size, but with detection probability independent of cluster size, s. Let s be Poisson with mean κ. The parameter D is the density of individuals, not clusters. In the Poisson model, as given above, for counts of clusters, the density parameter is cluster density, D_s, not D. To parameterize this likelihood component in terms of density of individuals, we must replace D_s by D/κ. The full likelihood for this model is

$$\mathcal{L}(D,\lambda,\kappa) = \frac{\exp\{-2LD\lambda/\kappa\} \cdot \{2LD\lambda/\kappa\}^n}{n!} \cdot \prod_{i=1}^{n} \left[\frac{1}{\lambda} \cdot \exp\left\{ -\frac{x_i}{\lambda} \right\} \cdot \left\{ \frac{\exp(-\kappa) \cdot \kappa^{s_i}}{s_i!} \right\} \right]$$

Using $\bar{s}$ to denote mean cluster size and $\bar{x} = T/n$, this likelihood can be reduced to

$$\mathcal{L}(D, \lambda, \kappa) = \exp\left\{ -\left(\frac{2LD\lambda}{\kappa} + \frac{T}{\lambda} + n\kappa \right) \right\} \cdot D^n \cdot \kappa^{n(\bar{s}-1)}$$

and

$$\log_e \mathcal{L}(D, \lambda, \kappa) = -\frac{2LD\lambda}{\kappa} - \frac{T}{\lambda} - n\kappa + n \cdot \log_e(D) + n(\bar{s} - 1) \cdot \log_e(\kappa)$$

The hybrid and full likelihood results agree here; in particular,

$$\hat{D} = \frac{n \cdot \bar{s}}{2L\bar{x}}$$

$$\text{var}(\hat{D}) = D^2 \left[\frac{3}{E(n)} \right]$$

To get the profile likelihood, we need $\hat{\lambda}(D)$ and $\hat{\kappa}(D)$. Closed form results do not seem to exist. However, from the two partial derivatives set to zero,

$$\frac{\partial \log_e \mathcal{L}(D, \lambda, \kappa)}{\partial \lambda} \quad \text{and} \quad \frac{\partial \log_e \mathcal{L}(D, \lambda, \kappa)}{\partial \kappa}$$

we can derive the equations

$$\hat{\kappa}(D) = \frac{\bar{x}}{\hat{\lambda}(D)} + \bar{s} - 1 \tag{6.38}$$

and

$$\hat{\lambda}(D) = \sqrt{\left\{\frac{n\bar{x}}{2LD} \cdot \left[\frac{\bar{x}}{\hat{\lambda}(D)} + \bar{s} - 1\right]\right\}} \tag{6.39}$$

The function $\phi(D)$ can be written as

$$\phi(D) = 6n \cdot \left[\hat{\kappa}(D) - \bar{s} - \log_e\left[\left(\frac{D}{\hat{D}}\right)^{1/3}\right] - (\bar{s} - 1) \cdot \log_e\left[\left(\frac{\hat{\kappa}(D)}{\bar{s}}\right)^{1/3}\right]\right] \tag{6.40}$$

To compute $\phi(D)$, choose a value of D, solve Equation 6.39 iteratively (easily done as $\hat{\lambda}(D)$ is a stable fixed point), compute $\hat{\kappa}(D)$, then compute Equation 6.40 (also using $\hat{D}$, which is closed form).

6.8.3 Full likelihood for point transects: simple examples

(a) *Negative exponential $g(r)$, Poisson n* Results for point transects can be obtained for a couple of simple cases. Here we assume a Poisson distribution for n, a negative exponential detection function, $g(r) = \exp(-r/\lambda)$, and k randomly placed points. Basic theory then gives $E(n) = 2\pi k D\lambda^2$, and the pdf of detection distance r is

$$f(r) = \frac{r \cdot \exp\left\{-\frac{r}{\lambda}\right\}}{\lambda^2} \qquad 0 < r,\ 0 < \lambda$$

The full likelihood is

$$\mathcal{L}(D, \lambda) = \frac{\exp\{-2\pi k D\lambda^2\} \cdot \{2\pi k D\lambda^2\}^n}{n!} \cdot \prod_{i=1}^{n}\left[\frac{r_i \cdot \exp\left\{-\frac{r_i}{\lambda}\right\}}{\lambda^2}\right]$$

which simplifies to

$$\mathcal{L}(D,\lambda) = \exp\left\{-\left(2\pi k D\lambda^2 + \frac{T}{\lambda}\right)\right\} \cdot D^n$$

where $T = \sum_{i=1}^{n} r_i$

The log likelihood is thus

$$\log_e \mathcal{L}(D, \lambda) = -2\pi k D\lambda^2 - \frac{T}{\lambda} + n \cdot \log_e(D)$$

Standard likelihood theory now leads to

$$\hat{\lambda} = \frac{T}{2n} = \frac{\bar{r}}{2} \qquad \text{var}(\hat{\lambda}) = \frac{\lambda^2}{2E(n)}$$

$$\hat{D} = \frac{n}{2\pi k\hat{\lambda}^2} = \frac{n \cdot \hat{h}(0)}{2\pi k} \qquad \text{var}(\hat{D}) = D^2\left[\frac{3}{E(n)}\right]$$

In order to find the profile likelihood, we solve

$$\frac{\partial \log_e \mathcal{L}(D, \lambda)}{\partial \lambda} = -4\pi k D\lambda + \frac{T}{\lambda^2} = 0$$

getting

$$\hat{\lambda}(D) = \left[\frac{T}{4\pi k D}\right]^{1/3}$$

Using the above to form $\log_e \mathcal{L}(D, \hat{\lambda}(D))$ allows us to find the expression for $\log_e \mathcal{L}(\hat{D}, \hat{\lambda})$, from which we construct a simple representation of $\phi(D)$:

$$\phi(D) = 6n \cdot \left[(D/\hat{D})^{1/3} - 1 - \log_e \{(D/\hat{D})^{1/3}\}\right] \tag{6.41}$$

(b) *Half-normal $g(r)$, Poisson n* Instead of a negative exponential $g(r)$, let us assume $g(r)$ is half-normal; other assumptions are as in the above case. Now basic theory gives $E(n) = 2\pi k D\sigma^2$, and the pdf of detection distances is

$$f(r) = \frac{r \cdot \exp\left\{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right\}}{\sigma^2} \qquad 0 < r,\ 0 < \sigma$$

The full likelihood is

$$\mathcal{L}(D, \sigma) = \frac{\exp\{-2\pi k D\sigma^2\} \cdot \{2\pi k D\sigma^2\}^n}{n!} \cdot \prod_{i=1}^{n}\left[\frac{r_i \cdot \exp\left\{-\frac{1}{2}\left(\frac{r_i}{\sigma}\right)^2\right\}}{\sigma^2}\right]$$

which simplifies to

$$\mathcal{L}(D, \sigma) = \exp\left\{-\left(2\pi k D\sigma^2 + \frac{T}{2\sigma^2}\right)\right\} \cdot D^n$$

for T defined as the total, $T = \sum_{i=1}^{n} r_i^2$

Standard likelihood theory now leads to

$$\hat{\sigma}^2 = \frac{T}{2n} \qquad \text{var}(\hat{\sigma}^2) = \frac{\sigma^4}{E(n)}$$

$$\hat{D} = \frac{n}{2\pi k\hat{\sigma}^2} = \frac{n \cdot \hat{h}(0)}{2\pi k} \qquad \text{var}(\hat{D}) = D^2 \cdot \left[\frac{2}{E(n)}\right]$$

To find the profile likelihood, we solve

$$\frac{\partial \log_e \mathcal{L}(D, \sigma)}{\partial \sigma} = -4\pi k D\sigma + \frac{T}{\sigma^3} = 0$$

getting

$$\hat{\sigma}(D) = \left[\frac{T}{4\pi k D}\right]^{1/4}$$

Carrying through the algebra and simplifications, we have

$$\phi(D) = 4n \cdot [(D/\hat{D})^{1/2} - 1 - \log_e \{(D/\hat{D})^{1/2}\}] \qquad (6.42)$$

Notice that Equation 6.42 is identical to 6.37; this we expected, because there is a duality in the mathematics between the case of line transects with a negative exponential detection function and point transects with a half-normal detection function, both with n distributed as Poisson and with $w = \infty$.

6.8.4 Some numerical confidence interval comparisons

We used the above results on $\phi(D)$ and $\text{var}(\hat{D})$ to compute a few illustrative numerical examples of profile, log-based and standard confidence intervals (nominal 95% coverage). To facilitate comparisons, what is presented are the ratios, (interval bound)/$\hat{D}$. Thus, the standard method yields relative bounds as $1 \pm 1.96\,\text{cv}(\hat{D})$, and the log-based relative bounds are $1/C$ and C, where

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

Some of our results are based on sample sizes that are smaller than would be justified for real data; our intent is to compare the three methods, and the differences are biggest at small n. The actual coverage of the intervals is not known to us; we take the profile likelihood intervals as the standard for comparison. Results are shown in Tables 6.2, 6.3, 6.4 and 6.5. One reason for the comparisons is to provide evidence that the log-based intervals are generally closer to the profile intervals.

Table 6.2 Some relative 95% confidence intervals, $\hat{D}_{lower}/\hat{D}$ and $\hat{D}_{upper}/\hat{D}$, for the profile, log-based and standard method, for line transects with a half-normal detection function, $w = \infty$, and Poisson distributed sample size n. Equation 6.36 is the basis of the profile interval; results are invariant to the true D and σ

n	Profile interval		log-based interval		Standard interval	
5	0.296	2.610	0.366	2.729	− 0.074	2.074
10	0.437	2.014	0.481	2.081	0.241	1.759
20	0.565	1.660	0.590	1.694	0.462	1.537
40	0.673	1.439	0.687	1.457	0.621	1.380
70	0.744	1.321	0.752	1.330	0.713	1.287
100	0.781	1.263	0.787	1.270	0.760	1.240

Table 6.3 Some relative 95% confidence intervals, $\hat{D}_{lower}/\hat{D}$ and $\hat{D}_{upper}/\hat{D}$, for the profile, log-based and standard method, for line transects with a negative exponential detection function, $w = \infty$, and Poisson distributed sample size n. Equation 6.37 is the basis of the profile interval; results are invariant to the true D and λ

n	Profile interval		log-based interval		Standard interval	
5	0.251	3.076	0.321	3.117	− 0.240	2.240
10	0.389	2.264	0.433	2.309	0.124	1.877
20	0.520	1.803	0.546	1.831	0.380	1.620
40	0.635	1.526	0.649	1.542	0.562	1.438
70	0.711	1.380	0.720	1.390	0.669	1.331
100	0.753	1.311	0.759	1.318	0.723	1.277

Table 6.4 Some relative 95% confidence intervals, $\hat{D}_{lower}/\hat{D}$ and $\hat{D}_{upper}/\hat{D}$, for the profile, log-based and standard method, for point transects with a negative exponential detection function, $w = \infty$, and Poisson distributed sample size n. Equation 6.41 is the basis of the profile interval; results are invariant to the true D and λ

n	Profile interval		log-based interval		Standard interval	
5	0.191	4.056	0.261	3.833	− 0.518	2.518
10	0.319	2.754	0.366	2.729	− 0.074	2.074
20	0.453	2.072	0.481	2.081	0.241	1.759
40	0.575	1.684	0.590	1.694	0.463	1.537
70	0.660	1.487	0.669	1.494	0.594	1.406
100	0.708	1.395	0.714	1.401	0.661	1.339

Table 6.5 Some relative 95% confidence intervals, $\hat{D}_{lower}/\hat{D}$ and $\hat{D}_{upper}/\hat{D}$, for the profile, log-based and standard method, for line transects with a negative exponential detection function (parameter λ), $w = \infty$, Poisson distributed sample size n, and cluster size as Poisson, mean κ. Equation 6.40 is the basis of the profile interval; results are invariant to true D and λ, but depend weakly on true κ; $\hat{\kappa} = 3.0$ was used for these results

n	Profile interval		log-based interval		Standard interval	
5	0.230	3.440	0.261	3.833	− 0.518	2.518
10	0.365	2.440	0.366	2.729	− 0.074	2.074
20	0.497	1.900	0.481	2.081	0.241	1.759
40	0.614	1.583	0.590	1.694	0.463	1.537
70	0.693	1.419	0.669	1.494	0.594	1.406
100	0.737	1.341	0.714	1.401	0.661	1.339

Table 6.2 corresponds to the line transect case in which $g(x)$ is half-normal and objects have a Poisson distribution. The relative interval endpoints in that table depend only upon sample size n, so these results are quite general under the assumed model. The invariance property of the ratios $\hat{D}_{lower}/\hat{D}$ and $\hat{D}_{upper}/\hat{D}$ applies also to Table 6.3 (line transect, negative exponential $g(x)$ and Poisson n) and Table 6.4 (point transect, negative exponential $g(x)$ and Poisson n). The log-based interval is slightly to be preferred to the standard method in Table 6.2, and more strongly preferred for the cases of Tables 6.3 and 6.4. The choice in Table 6.5 (line transect, negative exponential $g(x)$, Poisson n and Poisson s) is unclear. Note that the results in Table 6.3 for line transects with a negative exponential $g(x)$ are identical to results for the same values of n for point transects with a half-normal $g(r)$.

These sort of results on confidence intervals would be interesting to compute for other scenarios. We present the above specific formula for $\phi(D)$ to illustrate the ideas; in particular, the negative exponential $g(y)$ is used only because it is very easy to work with.

Table 6.5 reflects a case where the population of objects is clustered. The relative confidence intervals are for density of individuals. This is an interesting case because the log-based and standard relative confidence intervals do not depend upon D, λ or κ (because the relative intervals do not depend upon the specific values of $\bar{x}$ or $\bar{s}$). The relative profile intervals do not depend upon D or λ (thus the results in Table 6.5 are independent of the choice of D and λ), but they do depend weakly upon κ because the specific value of $\bar{s}$ (three in this example) affects even the relative profile intervals. Heuristically, this seems to be because the sample size of number of individual animals detected increases as $\bar{s}$ increases and the likelihood function uses this information. To illustrate this point, we give below relative profile interval endpoints (based on Equation 6.40 and 95% nominal coverage) for a few values of $\bar{s}$ at $n = 20$:

n	$\bar{s}$	Profile interval	
20	1.25	0.465	2.023
20	3.00	0.497	1.900
20	30.00	0.518	1.813
20	300.00	0.520	1.804

There is quite a noticeable effect here of average group size and this is an effect that is not found in either standard or log-based methods. We speculate that in realistic likelihood models, the profile interval would

be generally more sensitive to information in the data than simpler confidence interval methods.

(a) *One more example* Consider line transect sampling, in which $g(x)$ is half-normal, clusters have a homogeneous Poisson distribution, and cluster size is a geometric random variable. Further assume that $g_0 < 1$, but that it can be estimated by an independent source of information, from which it is known that, of m clusters 'on' the line, z are detected. We assume that z is distributed as binomial (m, g_0). The counts, n, will be Poisson with mean $E(n) = 2LDg_0/\{\kappa f(0)\}$ where $\kappa = E(s)$ and D is density of individuals. The geometric distribution is used here in the form $\pi(s) = p^{s-1}(1-p)$, hence $\kappa = 1/(1-p)$. Also, $f(0) = \frac{1}{\sigma} \cdot \sqrt{(2/\pi)}$, so that we have

$$E(n) = \sqrt{2\pi}\,\sigma LDg_0(1-p)$$

Maximum likelihood estimators are

$$\hat{\sigma} = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{T}{n}$$

$$\hat{D} = \frac{n \cdot \hat{f}(0) \cdot \bar{s}}{2L\hat{g}_0}$$

$$\hat{p} = \frac{\bar{s} - 1}{\bar{s}}$$

and

$$\hat{g}_0 = \frac{z}{m}$$

and the asymptotic estimated $\text{var}(\hat{D})$ is

$$\widehat{\text{var}}(\hat{D}) = \hat{D}^2 \cdot \left[\frac{1}{n} + \frac{1}{2n} + \frac{\hat{p}}{n} + \frac{1}{m} \cdot \left(\frac{1}{\hat{g}_0} - 1\right)\right]$$

The full likelihood of the data entering into $\hat{D}$ is given by the products of the likelihoods of the independent data components:

$$\mathcal{L}(D,\sigma,p,g_0) = \frac{\exp\{-[\sqrt{2\pi}\sigma LDg_0(1-p)]\} \cdot \{\sqrt{2\pi}\sigma LDg_0(1-p)\}^n}{n!} \times$$

$$\prod_{i=1}^{n}\left[\left(\sqrt{\frac{2}{\pi\sigma^2}}\right)\cdot \exp\left\{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2\right\}\right] \times \prod_{i=1}^{n}[p^{(s_i-1)}\cdot(1-p)] \times \left[\binom{m}{z}(g_0)^z(1-g_0)^{m-z}\right]$$

Dropping constants and otherwise simplifying this likelihood gives

$$\mathcal{L}(D,\sigma,p,g_0) = \left[\exp\left\{-\left(\sqrt{2\pi}\sigma LDg_0(1-p) + \frac{T}{2\sigma^2}\right)\right\}\right] \times$$
$$[D^n \cdot p^{n(\bar{s}-1)}(1-p)^{2n}(g_0)^z(1-g_0)^{m-z}] \qquad (6.43)$$

Closed form expressions for $\hat{\sigma}(D)$, $\hat{p}(D)$ and $\hat{g}_0(D)$ do not seem possible, but $\phi(D)$ can be computed using numerical optimization. A slight 'trick' simplifies the process of getting $\phi(D)$.

By setting the partial derivatives of $\mathcal{L}$ with respect to σ, p and g_0 to zero, and with D arbitrary, we derived the following results:

$$p(\sigma) = \frac{\bar{s}-1}{\bar{s}+1-\dfrac{T}{n\sigma^2}}$$

$$g_0(\sigma) = \frac{n+z-\dfrac{T}{n\sigma^2}}{n+m-\dfrac{T}{n\sigma^2}}$$

and

$$D = \frac{T}{\sqrt{2\pi}\sigma^3 L g_0(\sigma)\cdot(1-p(\sigma))}$$

While we cannot easily select D and compute $\phi(D)$, we can specify values of σ and compute the unique associated $p(\sigma)$ and $g_0(\sigma)$ that apply for D, which is then computed. These are then the values of $\hat{\sigma}(D)$, $\hat{p}(D)$ and $\hat{g}_0(D)$ to use in computing $\phi(D)$ for that computed value of D. All we need do is select a range of σ which generates a range of D, and then we treat $\mathcal{L}(D,\sigma,p,g_0)$ as a function of D, not of σ. The MLEs are known, so the absolute maximum of $\mathcal{L}$ is known, thus normalizing $\mathcal{L}$ to ϕ is easy.

Table 6.6 Some relative 95% confidence intervals, $\hat{D}_{lower}/\hat{D}$ and $\hat{D}_{upper}/\hat{D}$, for the profile, log-based and standard method, for line transects for the likelihood in Equation 6.43. Sample size n is Poisson, x is half-normal, s is geometric, and g_0 is estimated from z ~binomial(m, g_0)

n	Profile interval		log-based interval		Standard interval	
5	0.238	4.009	0.289	3.457	− 0.376	2.376
10	0.365	2.688	0.395	2.535	0.015	1.985
20	0.488	2.054	0.501	1.996	0.287	1.713
40	0.593	1.721	0.595	1.680	0.472	1.528
70	0.663	1.568	0.658	1.521	0.576	1.424
100	0.700	1.503	0.690	1.449	0.626	1.374

Table 6.6 gives a few numerical results for the model considered here. Sample sizes n and m are the dominant factors influencing the Table 6.6 results. In fact, these results do not depend on true D, L or σ. However, they do depend on $E(s)$ and g_0 too strongly to draw broad conclusions here. Inputs to the likelihood used for Table 6.6 were $T/n = 1$ (so MLE $\hat{\sigma} = 1$), $\bar{s} = 10$ ($\hat{p} = 0.9$) and $z = 16$, $m = 20(\hat{g}_0 = 0.8)$. The log-based intervals are closer to the profile intervals than are the standard intervals.

It is also worth noting that if n (i.e. the line transect sampling effort) is increased while m is fixed, the estimate of g_0 is the weak link in the data. Studies that estimate g_0 need to balance the effort for the two data types. It would be best to collect data on g_0 during the actual distance sampling study to achieve both such balance of effort (with respect to n and m) and relevance of $\hat{g}_0$ to the particular study. If in this example for $n = 100$ we also put $z = 80$ and $m = 100$, then the three relative intervals are more similar, especially profile and log-based:

Profile interval		Log-based interval		Standard interval	
0.726	1.377	0.728	1.373	0.681	1.319

In practice, it is unlikely that such a high proportion of detections (80%) could be considered as 'on' the line, necessitating the use of methods that utilize detections off the line (Section 6.4).

(b) *A general comment on precision* The relative confidence intervals in Tables 6.2–6.6 have been computed in a variety of cases: line and point transects, some with clustered populations, different detection functions, and one case with an adjustment for $g(0) < 1$. A general conclusion is that sample size has the overwhelming effect on relative precision of

$\hat{D}$. Relative confidence intervals are quite wide at $n = 40$, being roughly $\pm$ 45% of $\hat{D}$. At $n = 70$ and 100, the intervals are roughly $\pm$ 35% and $\pm$ 25%, respectively. This level of precision is under very idealized conditions that will not hold in practice for real data. With comparable sample sizes, we expect that relative interval widths will exceed the tabulated values. These results and our experience in distance sampling suggest strongly that reliable, precise abundance estimates from distance data require minimum sample sizes around 100. Coefficients of variation of around 20% (i.e. intervals of $\pm$ 40%) are often adequate for management purposes; the results presented here indicate minimum sample sizes of 40–70 in this circumstance.

6.8.5 Discussion

Reliable analysis of distance sampling data is possible without a full likelihood approach. We recommend a robust approach of empirical estimation of var(n), a semiparametric, likelihood-based estimation of $f(0)$ or $h(0)$ from the marginal distance data, and finally, estimation of $E(s)$ conditional on the observed distances y_i, using a regression approach. Other strategies are possible and use of a bivariate model for $g(y, s)$ is closer to a full likelihood approach. The difficult modelling aspect is to specify general probability models for n and s and it is those steps we bypass.

There are, however, reasons to develop a full likelihood approach: (1) intellectual curiosity, (2) efficiency of estimators and tests if the assumed model is correct, (3) availability of well-developed likelihood based theory for profile likelihood intervals and for model selection such as AIC, (4) the convenience of further developing such models by having parameters as functions of covariates (effort, environmental, spatial factors), and (5) as a necessary part of a Bayesian approach to distance sampling. We consider some of these points and difficulties of the approach.

Models for the marginal function $g(y)$ are abundant and choices also exist for bivariate versions, $g(y, s)$ (e.g. Quinn and Gallucci 1980; Drummer 1985, 1990; Drummer and McDonald 1987; Thompson and Ramsey 1987; Otto and Pollock 1990), and for the distribution $\pi(s)$ of cluster size in the sampled population. Any probability model of a discrete random variable on $s = 1, 2, \ldots$ is a candidate for $\pi(s)$, and if s can take on hundreds of values (such as for dolphin schools), continuous models could be used (such as a log-normal distribution for s). Good probability models for n are more problematic.

The Poisson distribution for n is not reasonable. The negative binomial model might be tenable, but in general, a reasonable model for

$\Pr(n)$ may need more than two parameters. The negative binomial is given by

$$\Pr(n) = \frac{\Gamma(\theta + n)}{\Gamma(\theta) \cdot \Gamma(n+1)} \cdot (1-\tau)^n \cdot \tau^\theta, \qquad 0 < \theta,\ 0 < \tau < 1,\ 0 \leqslant n$$

which has

$$E(n) = \theta \cdot \frac{1-\tau}{\tau}$$

$$\text{var}(n) = \frac{E(n)}{\tau}$$

$\Pr(n)$ is not used as parameterized, but rather the relationship $E(n) = 2LD/f(0)$ (line transects), or $E(n) = 2\pi k D/h(0)$, must be imposed on the parameters in the distribution. With a multiparameter $\Pr(n)$, such as the negative binomial, there is no obvious unique way to reparameterize $\Pr(n)$. We suggest it will instead be necessary to optimize the log-likelihood function subject, for example, to the constraint $E(n) = 2LD/f(0)$, where $f(0)$ is replaced by its form as a function of the parameters in the detection function $g(x)$. In some cases, it might be meaningful to associate one parameter in $\Pr(n)$ with $E(n)$, such as having τ a free parameter, and setting

$$\theta = \frac{2LD\tau}{f(0) \cdot (1-\tau)}$$

There are other generalized distributions possible for n, see for example Johnson and Kotz (1969).

Constructing the full likelihood in the general case is complicated, but not fundamentally difficult, if strong assumptions of independence are made. These independence assumptions are often not reasonable, but robust variances can be found by appropriate quasi-likelihood or bootstrap methods. Under independence and k replicate lines, the probability model for the data (from which the likelihood is derived) is symbolically

$$\Pr(n_i,\ x_{ij},\ s_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, k)$$
$$= \prod_{i=1}^{k} \left[\Pr(n_i | l_i) \cdot \left\{ \prod_{j=1}^{n_i} f(x_{ij} | s_{ij}) \cdot \pi^*(s_{ij}) \right\} \right] \qquad (6.44)$$

where

$$\pi^*(s) = \frac{f(0)}{f(0|s)} \cdot \pi(s)$$

is the distribution of cluster sizes given the cluster is detected and $\pi(s)$ is the probability distribution in the entire population. It is $\pi(s)$ we suggest modelling. For point transects, we have k points and

$$\Pr(n_i, r_{ij}, s_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, k) = \prod_{i=1}^{k} \left[\Pr(n_i | \text{point } i) \cdot \left\{ \prod_{j=1}^{n_i} f(r_{ij}|s_{ij}) \cdot \pi^*(s_{ij}) \right\} \right] \tag{6.45}$$

where

$$\pi^*(s) = \frac{h(0)}{h(0|s)} \cdot \pi(s)$$

More on the theory of $f(x|s)$, $f(r|s)$ and $\pi^*(s)$ is given in Section 6.7 along with the explanations of $f(0)$, $h(0)$, $f(0|s)$ and $h(0|s)$ and their relationship to $g(\cdot|s)$.

Theory development in capture–recapture is in some ways more advanced than in distance sampling; capture–recapture is also in some ways a simpler statistical problem. The state of the statistical art in capture–recapture for survival estimation is represented by Lebreton *et al.* (1992), in which inference is based on (full) likelihood models, and model selection is based on Akaike's Information Criterion (AIC; Akaike 1985). We have made use of AIC in model selection for the marginal detection function, but only for that model component. Drummer (1991) uses AIC in a bivariate detection function, $g(x, s)$. The full likelihood approach to capture–recapture is very powerful. Also, using such explicit parametric models allows meaningful modelling of embedded parameters as functions of auxiliary information. These approaches could be similarly useful in distance sample and deserve to be explored.

Survival analysis in capture–recapture deals with only two classes of parameters: survival rates and capture rates. A fully parametric approach to distance sampling would deal with D, the parameters in $\Pr(n)$ (say $\underline{\theta}$), in $g(\cdot)$ (say $\underline{\beta}$), and in $\pi(s)$ (say $\underline{\gamma}$). Moreover, if the locational information in spatial coordinates of detected objects is used, then D is in effect expanded into a fourth class of parameters. Thus in its most general form, distance sampling deals with more classes of parameters than capture–recapture and is in that sense a harder problem.

If Bayesian methods are to be fully developed for distance sampling, they will require likelihoods as in Equations 6.44 and 6.45 to augment to priors on the parameters. In either case of a full likelihood or a Bayesian approach, there will be a need for numerical optimization and integration methods, possibly on objective functions with a dozen, or many more, parameters. Even 50 or 100 parameters are not too many for numerical optimization methods (MLEs and profile likelihoods), so this scope of problem is numerically feasible now.

6.9 Distance sampling in three dimensions

Conceptually, line transects can be considered as one-dimensional distance sampling, because only distances perpendicular to the line of travel are used, even though objects are distributed in two dimensions. Point transects sample distances in those two dimensions because radial detection distances are taken at any angle in what could be represented as an x–y coordinate system. In principle, distance sampling can be conducted in three dimensions, such as underwater for fish, or in space for asteroids, where objects can be located anywhere in three dimensions relative to the observer. The observer might traverse a 'line', and record detection distance in two dimensions perpendicular to the line of travel, or remain at a point, recording data in three-dimensions within the sphere centred at the point. Given the assumption of random line or point placement with respect to the three dimensional distribution of objects, the mathematical theory is easily developed for the three-dimensional case. In practice, the third dimension may pose a problem: there may only be a thin layer in three dimensions, and in the vertical dimension, objects may exhibit strong density gradients (e.g. birds in a forest canopy, or fish near the sea surface). Operational problems could be difficult; we do not claim this extension to three dimensions has application, but it is interesting to consider.

Assume that we follow a line randomly placed in three dimensions. Now we sample volume, not area, so D = objects/unit volume; line length is still L. Assume we record distances r for all objects detected out to perpendicular distance w. Counting takes place in a cylinder of volume $v = \pi w^2 L$, rather than a strip of area $a = 2wL$. The statistical theory at a fixed slice through the cylinder perpendicular to the line of travel is just like point transect theory. This sort of sampling (i.e. a 'tube transect') is like 'pushing' the point transect sampling point a distance L in three dimensions.

Aside from volume v replacing area a, we need little new notation: n is the sample size of objects detected in the sampled cylinder of radius

w; p_v is the average probability of detecting an object in the cylinder of volume v; $g(r)$ = probability of detecting an object that is at perpendicular distance r, $0 \leqslant r \leqslant w$; D is the true density of objects in the study space.

A basic starting place to develop theory is the formula

$$E(n) = D \cdot v \cdot p_v = D \cdot \pi w^2 \cdot L \cdot p_v$$

The unconditional detection probability is easily written down because objects are, by assumption, uniformly distributed in space within the cylinder. Therefore, the pdf of the radial distance r for a randomly specified object (before the detection process) is

$$u(r) = \frac{2\pi r}{\pi w^2}$$

The unconditional detection probability is $p_v = E[g(r)]$, where expectation is with respect to pdf $u(r)$. This is a weighted average of $g(r)$:

$$p_v = \int_0^w u(r) \cdot g(r)\, dr = \frac{1}{\pi w^2} \int_0^w 2\pi r \cdot g(r)\, dr$$

Notice that this p_v is identical to the unconditional detection probability in point transects.

A direct approach can be used to derive $E(n)$. Let v_ε be a small volume in the cylinder centred at distance r and position l along the line $(0 \leqslant l \leqslant L)$. Thus $D \cdot v_\varepsilon$ = the expected number of objects in volume $v_\varepsilon = 2\pi r\, dr\, dl$, and the expected count of these objects is then $g(r) \cdot D \cdot v_\varepsilon$. $E(n)$ can now be expressed as

$$E(n) = \int_0^L \int_0^w g(r) \cdot D \cdot 2\pi r\, dr\, dl = L \cdot D \cdot \int_0^w 2\pi r \cdot g(r)\, dr = L \cdot D \cdot \pi w^2 \cdot p_v$$

An estimator of D is

$$\hat{D} = \frac{n}{\pi w^2 \cdot L \cdot \hat{p}_v} \qquad \text{or} \qquad \hat{D} = \frac{n}{L \cdot \hat{\mu}_w}$$

where

$$\mu_w = \pi w^2 \cdot p_v = \int_0^w 2\pi r \cdot g(r)\, dr$$

The sample of distances to detected objects is $r_1, \ldots, r_n$. The pdf of distance r to detected objects is

$$f(r) = \frac{2\pi r \cdot g(r)}{\mu_w} = \frac{r \cdot g(r)}{\int_0^w r \cdot g(r)\, dr}$$

This result is identical to that for point transects and can be proven using the same theory. In fact, slight modifications of point transect theory suffice as a complete theory for line transect sampling in three dimensions. In particular,

$$f'(r) = \frac{2\pi \cdot g(r)}{\mu_w} + \frac{2\pi r \cdot g'(r)}{\mu_w}$$

so if $g'(0)$ is finite and $g(0) = 1$, then $f'(0) = 2\pi/\mu_w$. For consistency with point transects, we use

$$h(0) = f'(0) = \frac{2\pi}{\mu_w}$$

and hence we have

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

Compare this with the point transect estimator,

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

The only difference is that L replaces k.

In fact, all the theory for point transects applies to line transect sampling in three dimensions if we replace k by L. Thus, estimation of $h(0)$ or p_v could be done using program DISTANCE and treating the detection distances, r_i, as point transect data. The case of objects as clusters poses no additional problems, giving our general formulation:

$$\hat{D} = \frac{n \cdot \hat{h}(0)}{2\pi L c \hat{g}_0} \cdot \hat{E}(s)$$

where c is the fraction of the circle, around the line, in which detections are recorded ($c = \phi/2\pi$ for some sector angle ϕ in 0 to 2π). For the clustered case, all the theory in Section 3.6.6 for point transects applies to line transects in three dimensions with k replaced by L. We do not know of any data for three-dimensional line transects as described here; however, if any such studies are ever done, we note that a complete theory for their analysis already exists.

Point transect sampling in two dimensions can be extended to three dimensions. (To people who use the term variable circular plots, such extension becomes a variable spherical plot.) Now the detection distances r are embedded in a three-dimensional coordinate system. There is no existing theory for this type of distance sampling, although theory derivation methods used for line and point transects are easily adapted to this new problem, and we present some results here.

In this case, the observer would be at a random point and record detections in a full (or partial) sphere around that point. For a sphere of radius w, the volume enclosed about the point is

$$v = \frac{4}{3}\pi w^3$$

Given truncation of the data collection process at distance w, the expected sample size of detections at k random points is

$$E(n) = k \cdot D \cdot v \cdot p_v$$

To derive p_v, we note that the pdf of radial distance for a randomly selected object in the sphere is

$$u(r) = \frac{4\pi r^2}{4\pi w^3/3}$$

and $p_v = E[g(r)]$ with respect to $u(r)$

$$p_v = \int_0^w u(r) \cdot g(r)\, dr = \frac{1}{4\pi w^3/3} \int_0^w 4\pi r^2 \cdot g(r)\, dr$$

so that

$$E(n) = k \cdot D \cdot \frac{4}{3}\pi w^3 \cdot p_v = k' \cdot D \cdot \int_0^w 4\pi r^2 \cdot g(r)\, dr$$

An alternative derivation is to consider that the volume, v_ε, of space in the shell at distances r to $r + dr$ is $4\pi r^2 dr$ (to a first order approximation, which is all we need as we let $dr \to 0$). Thus,

$$E(n) = k \cdot D \cdot \int_0^w g(r) \cdot v_\varepsilon\, dr = k \cdot D \cdot \int_0^w 4\pi r^2 \cdot g(r)\, dr$$

Now define μ_w as

$$\mu_w = \int_0^w 4\pi r^2 \cdot g(r)\, dr$$

so that

$$\hat{D} = \frac{n}{k \cdot \hat{\mu}_w} \equiv \frac{n}{k \cdot (4\pi w^3/3) \cdot \hat{p}_v}$$

The pdf of detection distance r is

$$f(r) = \frac{4\pi r^2 \cdot g(r)}{\mu_w}, \qquad 0 < r < w$$

Taking second derivatives, we get

$$f''(r) = \frac{8\pi \cdot g(r)}{\mu_w} + \frac{16\pi r \cdot g'(r)}{\mu_w} + \frac{4\pi r^2 \cdot g''(r)}{\mu_w}$$

Hence, if $g(0) = 1$ and both $g'(0)$ and $g''(0)$ are finite (preferably zero as then the estimators have better properties), then

$$f''(0) = \frac{8\pi}{\mu_w}$$

For simplicity of notation, we define $d(0) = f''(0)$, so that

$$\hat{D} = \frac{n \cdot \hat{d}(0)}{8\pi k}$$

The estimation problem reduces to fitting a pdf $f(r)$, as given above, to the detection distances $r_1, \ldots, r_n$ based on some model for the detection function, $g(r)$. This will lead to $\hat{d}(0)$ and $\widehat{\text{var}}\{\hat{d}(0)\}$ by any of a variety of statistical methods. Because the variance of $\hat{d}(0)$ is conditional on n,

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

As with point transects in two dimensions, the theory for three dimensions can be transformed to look like line transect theory in one dimension. The transform is from radial distance r to the volume sampled, $\eta = \frac{4}{3}\pi r^3$, giving the pdf of η as

$$f(\eta) = \frac{g(\xi)}{\mu_w}, \qquad 0 < \eta < v = \frac{4}{3}\pi w^3$$

where

$$\xi = \left(\frac{3v}{4\pi}\right)^{1/3}$$

Then, if $g(0) = 1$, $f(0) = 1/\mu_w$. As for two-dimensional point transects, we do not recommend that analysis be based on such a transformation (c.f. Buckland 1987a).

The case of objects as clusters with size-biased detection can be developed for this three-dimensional point transect sampling using the methods of Section 3.6.6. First, we would have a conditional detection function, $g(r|s)$, and a distribution of cluster sizes in the entire population, $\pi(s)$. The following result holds for each cluster size:

$$D(s) = \frac{E[n(s)] \cdot d(0|s)}{8\pi k} \qquad 1 \leqslant s$$

The density of clusters irrespective of size is

$$D = \frac{E(n) \cdot d(0)}{8\pi k}$$

Thus, 'dividing' the first by the second of these two formulae, we get

$$\frac{D(s)}{D} = \pi(s) = \frac{E[n(s)]}{E(n)} \cdot \frac{d(0|s)}{d(0)} = \pi^*(s) \cdot \frac{d(0|s)}{d(0)}$$

where $\pi^*(s)$ is the distribution of detected cluster sizes. Summing both sides of the above leads to

$$d(0) = \sum \pi^*(s) \cdot d(0|s)$$

whereas rearranging the formula and summing produces

$$d(0) = \frac{1}{\sum \dfrac{\pi(s)}{d(0|s)}}$$

where all summations are over $s = 1, 2, 3, \ldots$

Thus

$$\pi(s) = \frac{\pi^*(s) \cdot d(0|s)}{\sum \pi^*(s) \cdot d(0|s)}$$

and

$$E(s) = \frac{\sum s \cdot \pi^*(s) \cdot d(0|s)}{\sum \pi^*(s) \cdot d(0|s)}$$

from which expressions, estimators of $\pi(s)$ and $E(s)$ are evident.

Straightforward expressions for $d(0)$ and $d(0|s)$ are

$$d(0) = \frac{2}{\int_0^w r^2 \cdot g(r)\, dr}$$

$$d(0|s) = \frac{2}{\int_0^w r^2 \cdot g(r|s)\, dr}$$

Two more formulae are just stated here:

$$g(r) = \sum g(r|s) \cdot \pi(s)$$

$$f(r|s) = \frac{r^2 \cdot g(r|s)}{\int_0^w r^2 \cdot g(r|s)\, dr}$$

Also of interest are conditional distributions of cluster size given detection distance, r. These distributions are useful for exploring $E(s|r)$, where now the s is from the size-biased detected sample. The result is

$$\pi^*(s|r) = \frac{g(r|s) \cdot \pi(s)}{\sum g(r|s) \cdot \pi(s)} \equiv \frac{g(r|s) \cdot \pi(s)}{g(r)}$$

This is exactly the same as for either line or point transect results given in Section 3.6.6.

Perhaps some day three-dimensional point transect data will be taken in deep space or oceans.

6.10 Cue counting

Cue counting (Hiby 1982, 1985; Hiby and Hammond 1989) is a method developed for estimating whale numbers that has very similar design considerations as line transect sampling – and in fact is sometimes carried out simultaneously with line transect sampling – yet theoretically is much more closely related to point transects. An observer scans a sector ahead of the viewing platform – usually an airplane or the bow of a ship – and records the distance to each detected cue. The cue is usually defined to be a whale blow. Cues are recorded irrespective of whether the whale was previously detected, and it is not necessary to estimate school (cluster) size. The method yields estimates of cue density, which can only be converted into whale density by estimating the cue or blow rate, ρ, from separate surveys. Cue density is estimated much as bird density is estimated from point transect data. The observer records only radial distances. Perpendicular distances are not needed, and angles only determine whether a cue is within or outside the observation sector. To estimate cue rate ρ, individual whales are followed, and the observed rate is used as an estimate of the cue rate for the whole population. This is the main weakness of the approach, as relatively few whales can be monitored for sufficiently long periods to obtain reasonable cue rate estimates. Further, these whales may not exhibit typical cue rates; for example whales with high cue rates are less likely to be 'lost' before an estimate can be obtained, and whales monitored over a long time period may change their cue rate in response to the vessel.

Suppose cues are recorded out to a distance w. For cue counting, as for point transect sampling, the area of a ring of incremental width δr at distance r from the observer is proportional to r. It follows that $f(r) = 2\pi r g(r)/\nu$, where $\nu = 2\pi \int_0^w r g(r)\, dr$. Given that a cue occurs in the sector of area $c \cdot a$, where ϕ is the sector angle and $a = \pi w^2$, so that

$c = \phi/2\pi$, let the probability that it is seen be P_a. Then this probability is $\nu/(\pi w^2)$. Thus $a \cdot P_a = \nu$, which holds as $w \to \infty$. Assuming all cues very close to the observer are seen ($g_0 = 1$), Equation 3.1, with $E(s) = 1$, yields the following estimate of cue 'density' per unit time (i.e. the number of cues per unit area per unit time):

$$\hat{D}_c = \frac{2\pi n}{\phi \hat{\nu} T}$$

where n is the number of cues recorded in time T. The constant T is the total time that the observer is searching (i.e. 'on effort'), and corresponds to the line transect parameter, L. If the cue rate is estimated as $\hat{\rho}$ cues per unit time per animal, then estimated whale density is

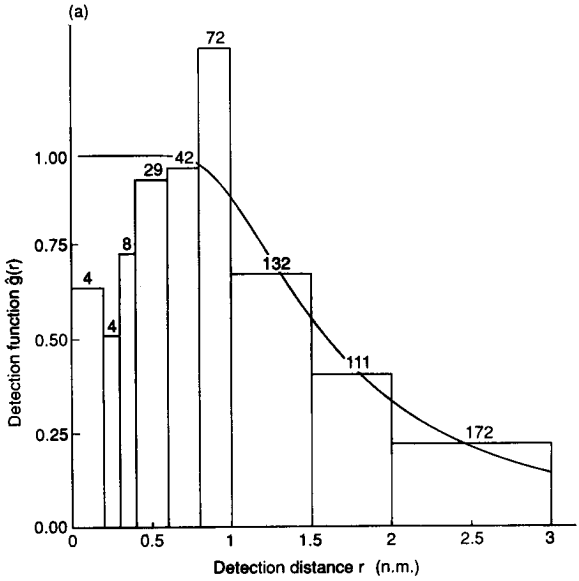
$$\hat{D} = \frac{2\pi n}{\phi \hat{\nu} T \hat{\rho}}$$

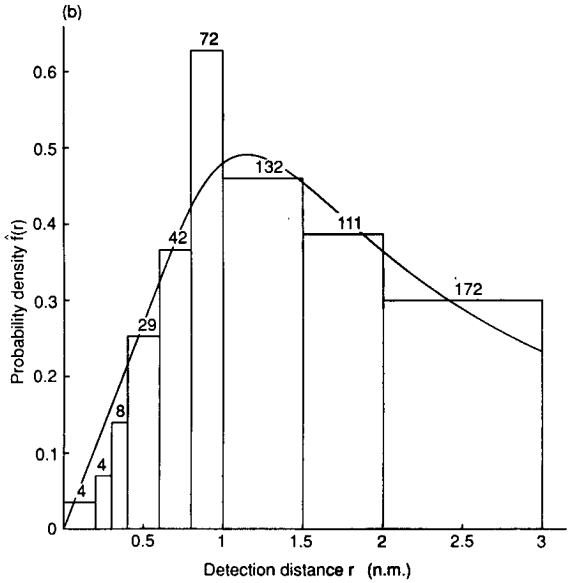
As for point transects, $\hat{\nu} = 2\pi/\hat{h}(0)$, where $h(0) = \lim_{r \to 0} f(r)/r$, so that

$$\hat{D} = \frac{n \cdot \hat{h}(0)}{\phi T \hat{\rho}}$$

The value of $\hat{h}(0)$ may be obtained by modelling the recorded distances to cues, as if they were distances from a point transect survey. DISTANCE has a cue count option to carry out the above analysis (below). Because successive cues from the same whale, or cues from more than one whale in a pod, may be counted, the distances are not independent observations. This does not invalidate the method, but analytic variances should not be used. The bootstrap, applied by taking say cruise legs as the sampling unit, provides valid variance estimation.

Line transect sampling of whale populations is beset with problems of how to estimate g_0, especially for aerial surveys, where a whale may be below the surface while it is in range of the observer, and for species such as sperm whales, which typically dive for around 40 minutes at a time. Cue counting does not require that all whales on the centreline are detected. Instead, it assumes that all cues occurring immediately ahead of the observer are seen. Thus, of those on the centreline, only whales that are at the surface when the vessel passes are assumed to be detected with certainty. In practice, whales may show vessel avoidance, so that the recorded number of cues very close to the vessel is depressed. Because the area surveyed close to the vessel is small, the effect of vessel avoidance might be expected to be small, unless avoidance occurs at


(a)
72
1.00
42
29
0.75
8
132
4
4
0.50
111
0.25
172
0.00
0
0.5
1
1.5
2
2.5
3
Detection function ĝ(r)
Detection distance r (n.m.)


(b)
72
0.6
0.5
132
0.4
111
42
0.3
172
29
0.2
8
0.1
4
4
0
0.5
1
1.5
2
2.5
3
Probability density f̂(r)
Detection distance r (n.m.)

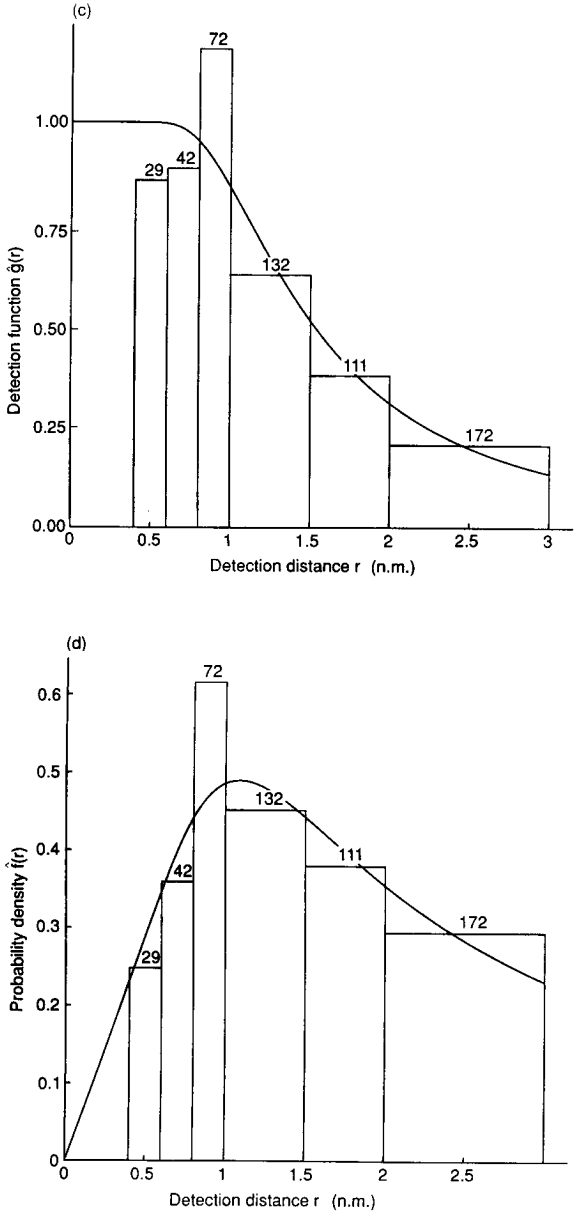


Fig. 6.7. Histograms of the cue count data. Also shown are the fits of the hazard-rate model to the data without left-truncation (a and b) and with left-truncation (c and d). The fitted detection functions are shown in (a) and (c) and the corresponding density functions in (b) and (d).

relatively large distances. If avoidance is suspected, the distance data may be left-truncated. This solution should prove satisfactory provided the effects of vessel avoidance only occur well within the maximum distance for which the probability of detecting a cue is close to unity.

If cues immediately ahead of the vessel might be missed, double-counting methods similar to the independent observer analyses given in Section 6.4 (with $d = 0$) may be used. This has the advantage over those analyses in that it is easier to identify whether a single cue is seen from both platforms, for example by recording exact times of cues, than to identify whether a single animal or animal cluster is seen by both platforms, since the two platforms may see different cues from the same animal.

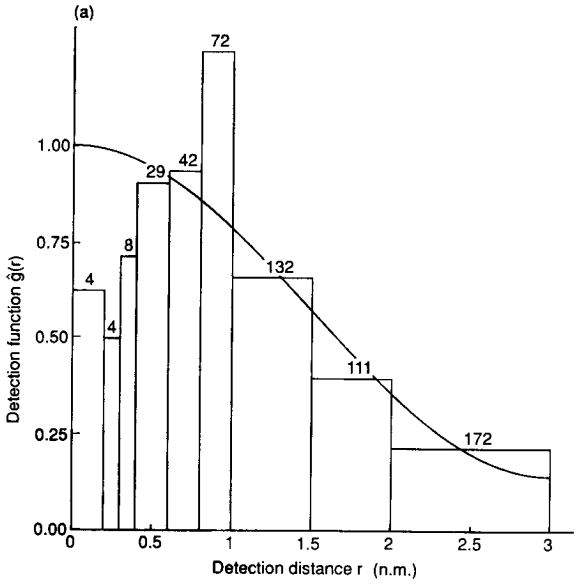
Cue counting has been used in aerial surveys to estimate fin whale densities near Iceland (Hiby *et al.* 1984) and in shipboard surveys to estimate whale densities in the North Atlantic (Hiby *et al.* 1989) and minke whale densities in the Antarctic (Hiby and Ward 1986a, b; Ward and Hiby 1987). We use here data from Hiby and Ward (1986a) to illustrate the method. Annual surveys of Southern Hemisphere minke whales have been carried out since the 1978–79 season. The first attempt to use cue counting during shipboard surveys occurred on the 1984–85 cruise. Hiby and Ward considered that cues close to the vessel were under-represented, possibly because whales showed vessel avoidance behaviour or because blows close to the vessel were under-recorded by observers. We therefore analysed the data both with no left-truncation and with left-truncation at 0.4 n.m. (nautical mile). The data were right-truncated at 3 n.m. Under the hazard-rate model, frequencies at distances less than 0.4 n.m. are not significantly below expected frequencies, and truncation makes little difference; the only anomaly is the relatively high frequency at 0.8–1.0 n.m. (Fig. 6.7), which may be chance fluctuation, or, more likely, preferential rounding to that distance interval. Hiby and Ward (1986a) appear to have interpreted these data too pessimistically, suggesting that detections close to the vessel are too few because (1) blows are less visible at short distances, (2) whales show vessel avoidance behaviour, or (3) observers did not appreciate the need to record all cues at short distances. Because successive cues are not independent, goodness of fit tests are likely to give spurious significant results. If they are carried out regardless for the hazard-rate model, they are not significant at the 5% level, so Hiby and Ward's conclusion that the data cannot be analysed seems pessimistic. Data sets collected more recently suggest that the method performs adequately.

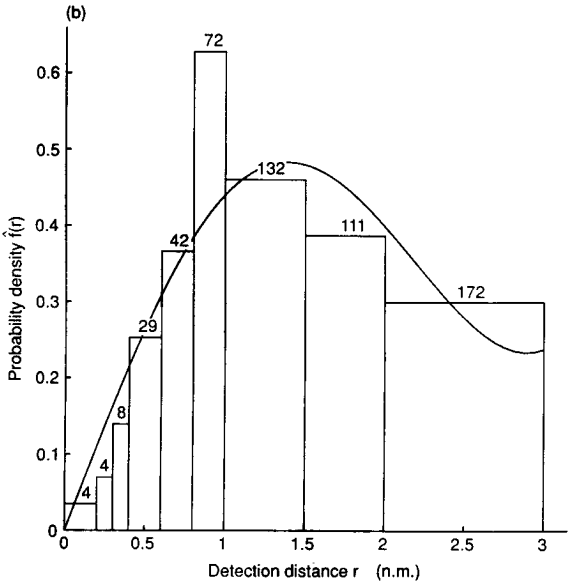
The fits of the hazard-rate model to the data both with and without left-truncation are shown in Fig. 6.7. In these trials, both blows and sightings of the body of the whale were counted as cues. Hiby and Ward (1986a) estimated the cue rate at 34.98 cues per whale per hour

($\widehat{\text{se}} = 4.74$). Supplying this estimate to DISTANCE, together with an estimate of time on effort of 35.8 hours (430 n.m. divided by an average speed of around 12 knots), yields an estimated density of 0.24 whales/n.m.^{2} from untruncated data and 0.26 whales/n.m.^{2} from the truncated data. The goodness of fit statistics are $\chi^2_6 = 11.7$ and $\chi^2_3 = 7.3$ respectively. The p-values for the goodness of fit tests are invalidated by the lack of independence between successive cues from the same animal or animal cluster. Similarly, the analytic estimates of variance are invalid. Without the raw data, it is not possible to apply either the bootstrap or the empirical method to obtain valid variance estimates, because cue counts are not given by cruise leg in Hiby and Ward. In Fig. 6.8, the fits of the Fourier series model to these data, with and without left-truncation, are shown. It yields an estimated density of 0.24 whales/n.m.^{2} without truncation and 0.31 whales/n.m.^{2} with truncation, with respective goodness of fit statistics of $\chi^2_7 = 18.9$ and $\chi^2_3 = 9.1$, indicating a worse fit than the hazard-rate model. Again the p-values corresponding to these statistics are invalid, and we do not present them. The flatter shoulder of the hazard-rate model enables it to fit the counts at short distances more closely. The estimate of density from a line transect survey carried out at the same time as the cue rate trial was 0.37 whales/n.m.^{2}.

6.11 Trapping webs

The estimation of population size (N) from capture data is usually formulated as a capture–recapture problem (e.g. White *et al.* 1982). There, traps are positioned, often at intersections of a rectangular grid, and animals are captured, marked, and released for possible recapture on a subsequent trapping occasion. If the trapping grid is enclosed or the trapped area samples the entire area of interest, then density = number/area can be estimated. However, the usual case is that an area surrounding the trapping grid contains animals that are subject to being captured and thus the effective area being sampled is larger than the area of the grid. One might naïvely estimate density as $\hat{N}/A_g$, where A_g = the area covered by the trapping grid. However, density is then overestimated because grid area is smaller than the area actually sampled by the traps. This problem has been well known for over half a century (Dice 1938). The use of a trapping web (Anderson *et al.* 1983) is an attempt to reformulate the density estimation problem into a distance sampling framework, where density is estimated directly, rather than separately estimating population size and effective area (but see Wilson and Anderson 1985a for an alternative).


(a)
72
1.00
29
42
0.75
8
132
4
4
0.50
111
0.25
172
0.00
0
0.5
1
1.5
2
2.5
3
Detection function ĝ(r)
Detection distance r (n.m.)


(b)
72
0.6
0.5
132
0.4
111
42
0.3
172
29
0.2
8
0.1
4
4
0
0
0.5
1
1.5
2
2.5
3
Probability density f̂(r)
Detection distance r (n.m.)

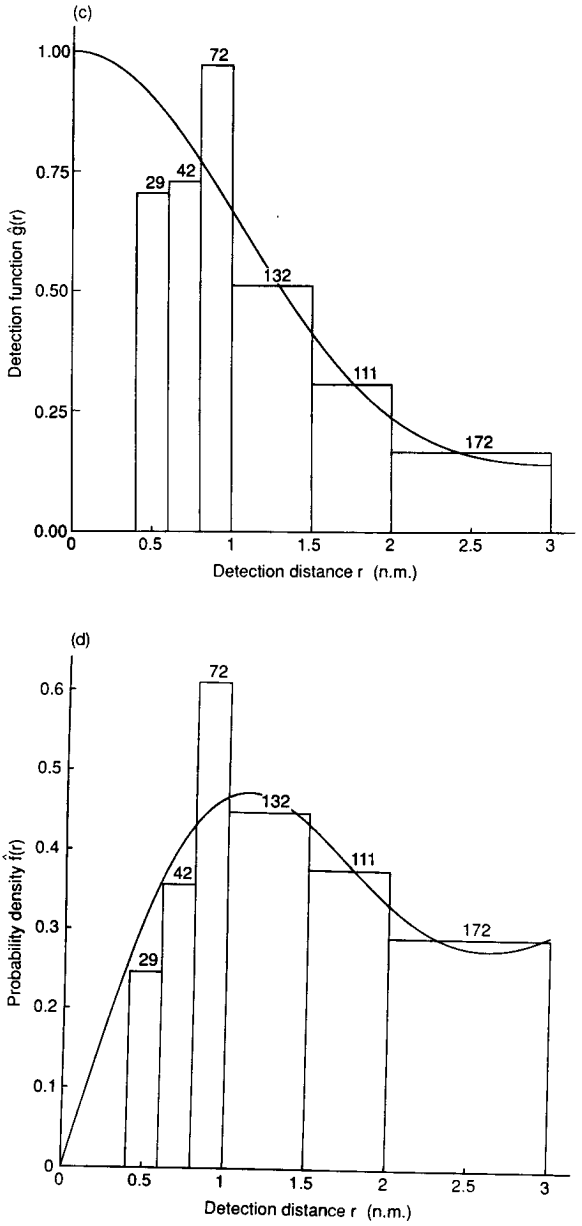


Fig. 6.8. Histograms of the cue count data. Also shown is the one-term Fourier series fit to the data with no left-truncation (a and b) and the two-term fit to the left-truncated data (c and d). The fitted detection functions are shown in (a) and (c) and the corresponding density functions in (b) and (d).

Trapping webs are a special case of point transect theory useful in estimating density of animal populations where 'detection' is accomplished by trapping. Animals are trapped in live, snap-trap or pitfall traps. Mist nets or other devices can be employed. Such trap devices are placed in a 'web' such that the density of traps is highest near the centre (Fig. 6.9), thus attempting to assure that $g(0) = 1$.

The web design consists of m lines of equal length, α_T, and each of T traps, radiating from randomly chosen points. A useful rule of thumb is to ensure that $m \times T \geq 200$. The traps are located along each of the m lines, usually (but not necessarily) at some fixed distance interval θ, starting at distance $\alpha_1 = \theta/2$. Points b_i are defined along each line, halfway between traps, for $i = 1, 2, \ldots, T$, with $b_0 = 0$ representing the web centre, and b_T the boundary of the web beyond the last trap. The traps are then at distances $\alpha_i = \theta(i - 0.5)$ for $i = 1, 2, \ldots, T$, and the b_i are at distances $i\theta$ for $i = 0, 1, 2, \ldots, T$.

Thus, traps are placed in rings of increasing radius from the web centre at equal distances along the m lines (Fig. 6.9). All captures in the ith ring of traps are considered to be detections of objects at distance α_i from the centre of the web. The distance data are analysed as grouped data. That is, the total number of captures arising from the ring of traps at distance α_i are treated as grouped data over the interval from distance b_{i-1} to b_i. The total area of the web out to interval i is $c_i = \pi b_i^2$ and the area trapped by the ith ring of traps is then $\Delta_i = c_i - c_{i-1}$. Generally, only first captures (removal data) are recorded and used in the estimation of density. This procedure reduces the impact on estimation of heterogeneity in trap response due to trap-happy or trap-shy animals.

Traps can be placed easily by trained technicians using a stake driven in the ground at the web centre and a rope with knots tied to indicate trap spacing (the α_i). Disturbance of the site should be minimized while traps are being placed in the sampled area. Several trapping webs would be required to sample an area of interest adequately. If only initial captures are of interest, then captured animals can be given a batch mark, to indicate that they have been 'removed', and released back into the population. Sampling is carried out on t occasions (often consecutive days or nights), where typically t is between three and eight.

6.11.1 Assumptions

Analytic theory for the trapping web is an application of point transect sampling theory and the general assumptions apply. The three major assumptions of distance sampling are slightly restated here for the trapping web:

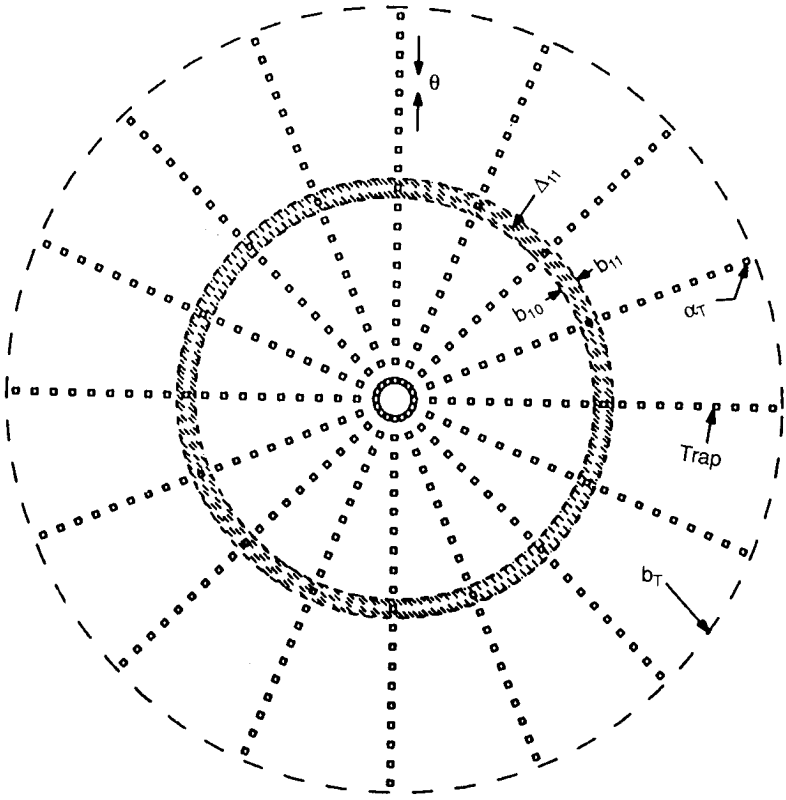


Fig. 6.9. Trapping web with 16 lines ($m = 16$), each of equal length α_T and 20 traps per line ($T = 20$), giving in total 320 traps. The traps are at distances $\alpha_1, \ldots, \alpha_T$ from the centre of the web. The points along each line, halfway between traps, are denoted by b_i, $i = 0, \ldots, T$, where $b_0 = 0$ is the centre and b_T is the boundary of the web, just beyond the last trap. Captures in the eleventh ring of traps are assigned to the shaded ring Δ_{11}, which has area $c_{11} - c_{10}$, where $c_i = \pi \cdot b_i^2$.

1. All animals at the centre of the web are captured at least once during the t occasions. That is, trapping continues until evidence exists that no new animals are being caught near the centre of the web.
2. During the trapping period, animals move over distances that are small relative to the size of each web. Thus, migration through the web is not allowed. Trap spacing is an important consideration and is species-dependent, taking account of the size of home ranges or 'territories'.
3. Distances from the centre of the web to each trap are measured accurately. This assumption is trivial if the trap spacing has been carefully laid out.

Assumption 1 is critical but can be monitored by examining the number of new individuals trapped near the web centre over trapping occasions. Animals near the centre should be captured with probability one. However, if substantial movement occurs over the t occasions (assumption 2), animals that are initially away from the centre of a web may move, eventually to be caught where the trap density is highest. This situation is analogous to point transect sampling where the observation period is long and birds move around the study area. Such movement causes detections near the point to increase and leads to positive bias in the estimator. Bias is worse if animals are attracted to the point or web centre.

6.11.2 Estimation of density

The basic data are the number of first captures in traps in ring j of web i on trapping occasion l, n_{ijl}, where $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, T$ and $l = 1, 2, \ldots, t$. Pooling the data over t occasions, the data can be summarized as n_{ij}, where

$$n_{ij} = n_{ij1} + n_{ij2} + \ldots + n_{ijt}$$

Hence n_{ij} is the number of animals trapped in the jth ring of the ith web. Let the total sample size be $n = \sum_i \sum_j n_{ij}$. Then density can be estimated by

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

where the estimate $\hat{h}(0)$ is obtained through standard point transect methods (Chapter 5). The estimator of the sampling variance is

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

If the population is distributed randomly (i.e. Poisson), then Wilson and Anderson recommend $[\text{cv}(n)]^2 = 1/n$. Generally, some degree of spatial aggregation can be expected, and $[\text{cv}(n)]^2 = 2/n$ or $3/n$ might then be more appropriate. If the number of replicate webs, k, is sufficient, it is preferable to estimate the sampling variance of n empirically (Section 3.7.2).

Data analysis is similar to the general theory for point transects, including model selection and inference issues. The challenge with the trapping web is to collect trapping data that mimic the assumptions of

point transect sampling and analysis theory. In particular, trap spacing must be related to average home range size or average distance moved and there are presently few guidelines for this decision.

Excessive animal movement near the web centre is problematic. The density of traps near the web centre is high relative to that near the edge of the web. Thus, even if animal movement is random, there is a tendency to trap animals near the web centre, regardless of their original location. If the trap spacing is too small, the problem is made worse and overestimation will likely result. If the animals tend to move in home ranges that are small relative to the size of the web and the trap spacing, then the trapping web may perform well. Alternatively, if animals move somewhat randomly over wide areas in relation to the size of the web and the trap interval chosen, then overestimation may be substantial (see the darkling beetle example, below).

6.11.3 Monte Carlo simulations

Wilson and Anderson (1985b) performed a Monte Carlo study to investigate the robustness of density estimation from trapping web data. Their simulations mimicked small mammal populations whose members were allowed to move in defined home ranges. Home range was simulated from bivariate normal, bivariate uniform and bivariate U-shaped distributions, and from a 'random excursion' model. More details are given by Wilson and Anderson (1985b). A 4 ha area was simulated, 320 traps were positioned in a two-dimensional plane, and animal density was set at two levels, 100/ha and 25/ha. Home range centres were allowed to be spatially random (Poisson), or clumped at three levels of aggregation. Three average probabilities of first capture were simulated at 0.09, 0.16 and 0.24, and these probabilities were allowed to vary by time (trapping occasion), behaviour (trap-shy or trap-happy) and heterogeneity (individual variability); this is model M_{tbh} in Otis *et al.* (1978: 43). Trapping was simulated for six, eight and ten occasions.

The Monte Carlo results indicated that the combination of a trapping web design and a point transect estimator of density was quite robust. The procedure had typically low bias under a wide variety of realistic situations. Confidence interval coverage was lower than the nominal level, due in part to the use of the Fourier series estimator (Buckland 1987a). The method was recommended in cases where the capture probability was > 0.16 and the number of trapping occasions was at least six. In some extreme situations (e.g. the random excursion model, with low capture probabilities and a clumped spatial distribution), the bias was in the 20–30% range, which might still be substantially less than traditional capture–recapture estimators.

The trapping web does not make any assumptions about geographic closure and is easy to implement in the field. No unique marks or tags are required, and several different types of trap can be used. The results of Wilson and Anderson (1985b) indicated that the trapping web was very promising as an alternative to standard capture–recapture methods. The work of Parmenter *et al.* (1989), summarized in Section 6.11.5, was carried out as a field test of the method where the true density was known.

6.11.4 A simple example

Anderson *et al.* (1983) presented an example of trapping web data from a 4.8 km area south of Los Alamos, New Mexico, where *Peromyscus* spp. were trapped for $t = 4$ nights on a web very similar to that of Fig. 6.9, with trap separation of $\theta = 3$m. The mice were captured in baited live traps and marked using a monel metal tag placed in one ear. Only initial captures were used in the analysis; animals were thus 'removed by marking'. No unmarked mice were caught in the inner area (out to ring 7) on the fourth night and only two new captures were made in this area on the third night. This was taken as evidence that the probability of capture near the web centre was one. A plot of the histogram indicated that mice from beyond the web were being attracted to the baited traps in the web, as the number of captures in rings 19 and 20 (i.e. n_{19} and n_{20}) was somewhat higher than expected. Thus, the distance data were truncated to exclude the two outer rings. This left 76 'detections' in 18 distance groups for analysis; frequencies were 1, 1, 0, 6, 2, 2, 3, 2, 4, 7, 4, 5, 8, 6, 7, 6, 7 and 5, respectively (Anderson *et al.* 1983). Note the lower frequencies in the inner rings, where the area sampled is small relative to that in the outer rings.

Two models were fit to these data: half-normal and hazard-rate, each with cosine adjustment parameters. No adjustment parameters were required and the AIC values were similar for the two models (424.18 and 425.76, respectively). Both models fit the data well as judged by the χ^2 goodness of fit tests ($\chi^2 = 13.2$ with 16 df and 13.6 with 15 df, respectively). Density is estimated at 97.8 mice/ha ($\widehat{\text{se}} = 21.3$) under the half-normal model and 86.1 mice/ha ($\widehat{\text{se}} = 12.7$) under the hazard-rate model. These estimates compare with 76 animals trapped on the web of area 0.97 ha, suggesting that most animals were caught. Only one web was sampled, hence no inference to a larger area is justified in this simple example.

6.11.5 Darkling beetle surveys

Populations of two species of ground-dwelling darkling beetles were studied in a shrub-steppe ecosystem in southwestern Wyoming to field

test the validity of the trapping web on a series of known populations (Parmenter *et al.* 1989). These beetles (10–30 mm body length) attain natural densities so great (> 2000 beetles/ha) that relatively small plots could be surveyed and still have test populations of reasonable size. These beetles are wingless and could be contained by low metal fences. They are easily marked on their elytra with coloured enamel paint, and are relatively long-lived, allowing longer periods of trapping and increased capture success. Pitfall traps were made from small metal cans (80–110 mm), and the web was surrounded by a metal enclosure wall. Traps were placed along 12 lines, each 11 m long, with 1 m trap spacing along the lines. Nine additional traps were placed at the centre of the web, giving 141 in total.

Beetles were captured, marked with enamel and released. These marked beetles constituted the population of known size that was subsequently sampled using the trapping web design. Surveys were done in two different years and different colours were used to denote the year. Several subpopulations, each of known size, were released, allowing analyses to be carried out both separately and in combination, to test the method on a wide range of densities. Additional details were given in Parmenter *et al.* (1989).

Overall the method performed quite well, yielding a correlation coefficient between $\hat{D}$ and D of at least 0.97 for each of four models for $g(r)$. The negative exponential model performed better than the Fourier series, exponential power series and half-normal models. The data exhibited a spike near the web centre, almost certainly caused by considerable movement of beetles and by trap spacing that was too small. All models were fitted after transforming the distance data to areas, a procedure that is no longer recommended. In summary, the results of these field tests were certainly encouraging.

Reanalysis of the darkling beetle data using current theory and program DISTANCE provided a less optimistic impression in that density was substantially overestimated, leading to important insights. The first is that traps were too closely spaced along the lines; trap spacing should have been greater to compensate for the wide area over which beetles of this species move. Second, the beetles had no 'home range' and thus tended to wander widely in relation to the size of the web, which is a function of trap spacing. The trapping web design was envisioned for use with animals that have some form of home range or 'territory'. Third, it is clear that some random movement may result in too many animals being trapped near the web centre. The problem can also arise in bird surveys where some random movement results in the detection of too many birds near the point. This condition leads to a spiked distribution and overestimation of density. Clearly, the additional

nine traps placed near the web centre aggravated this problem. Thus, we do not recommend that traps be concentrated at the centre. Research is needed to understand the relationship between spacing and density of traps and the nature of the movement.

If the data are spiked, one might analyse the distance data using some left truncation to eliminate the high numbers trapped near the web centre. Alternatively, one could constrain $\hat{g}(r)$ to be a low order, slowly decreasing function that does not track the spiked nature of the data, but this solution is rather arbitrary and may be ineffective. More experience is needed with sampling from populations of known size to understand better trap spacing and appropriate modelling. Still, this application of point transect sampling has many advantages over capture–recapture. Further studies on populations of known size, using live, snap or pitfall traps or mist nets, could lead to additional insights.

6.12 Migration counts

The main theme of this book is estimation of population abundance by modelling distance data. Counts from migration watch points may be converted into estimates of population size using similar techniques, but by modelling time instead of distance. Typically there will be regular, perhaps daily, counts of numbers of animals passing a watch point. If the animals pass in clusters, then the sampling unit will be the cluster. The basic data are start and end times of watch periods and number of animals or clusters passing during each watch period. Thus the data are in frequency form, being grouped by watch period. There will be gaps between watch periods, corresponding to night or to poor weather. For the basic method, animals are assumed to migrate at the same rate during unwatched periods as during watches. If no migration occurs at night, then time should be defined to end at dusk and start again at dawn. If migration occurs at night, but possibly at a different rate, the rate should be estimated by another method, for example by sonar (active or passive) or radar, or by radio-tagging animals. To model migration time, as distinct from distances in line transect sampling, the following changes are needed to the methodology. First, in line transect sampling the density function is assumed to be symmetric about the line, so only even functions (cosines for the Fourier series model and even powers for polynomial models) are used. For migration counts, odd functions are also needed. Related to this, the key function requires both a location and a scale parameter, whereas only a scale parameter is necessary for line transect sampling, because if sightings to the left of the line are recorded as negative, and distances to the right as

positive, the expected distance from the transect is zero under the assumption that the density function is symmetric about the line. Third, allowance must be made for a large number of counts, equal to the number of separate watch periods, whereas in a grouped analysis of line transect data, the number of groups for perpendicular distances seldom exceeds a dozen or so. Finally, having fitted the density to migration times, abundance is estimated by taking the ratio of the area under the entire density to the combined area corresponding to watch periods alone, and multiplying this ratio by the total number of animals counted during watches. Thus, different software is needed to obtain the abundance estimate and its standard error.

We use here as an example the California grey whale census data collected at Monterey, California. The California stock of grey whales migrates from feeding grounds in the Bering and Chukchi Seas to calving areas in Mexican waters every winter, returning north in spring. Aerial and ship surveys confirm that almost the entire population passes close inshore at several points. Counts at coastal migration watch points can therefore be used to estimate population size. Counts at Monterey were annual from 1967–68 through to 1979–80, and further surveys were carried out in 1984–85, 1985–86 and 1987–88. Reilly *et al.* (1980, 1983) gave more information on these surveys, and Buckland and Breiwick (in press) provided abundance estimates corresponding to all surveys. We use analyses of the grey whale count data for 1987–88, extracted from Breiwick *et al.* (unpublished) and Buckland *et al.* (in press), to illustrate analysis of migration count data. In that year, counts were made from two stations (north and south) a few yards apart, to allow estimation of numbers of pods missed during watch periods. The data analysed were numbers of pods passing within each count period, so that the data are grouped, the group endpoints being the start and end of each watch period. Information on duplicate detections was used to reduce the data sets from both stations to a single set of counts of the number of pods detected by at least one station in each watch period. Pods detected travelling north were excluded from the analyses.

The key function selected for fitting the counts was, apart from a scaling factor, the normal density:

$$\alpha(y) = \exp\left\{-\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$

where y corresponds to time, measured in days from a predetermined date. Adjustments to the fit of the key were made by adding Hermite polynomial terms sequentially, adjusting the fit first for skewness, then

kurtosis, and so on. Four adjustment terms were fitted to the data sets, and likelihood ratios were used to determine which fit was 'best'. If a one-term fit was found to offer no significant improvement over no terms, but a two-term fit gave a significant improvement over no terms at the 5% level, the two-term fit was favoured over both the one-term and the zero-term fits. A three-term (i.e. five-parameter) fit was selected, and this fit is shown in Fig. 6.10. To convert the fitted density to an estimate of population size, it is necessary to evaluate the proportion of the entire untruncated density that corresponds to watch periods. To ensure that the Hermite polynomial fits were sensible in the tails of the migration, zero counts were added for 1 December 1987, before the migration started, and 29 February 1988, after it ended. This had little effect for 1987–88, when counts took place throughout the main migration period (Fig. 6.10), but for some earlier surveys, many pods had passed before the first or after the last count of the season, making the addition of zero counts necessary (Buckland and Breiwick, in press).

In total, and excluding pods travelling north, $n = 3593$ pods were seen from at least one station. The χ^2 goodness of fit statistic corresponding to Fig. 6.10 was $\chi^2_{125} = 334.55$. This value is more indicative of overdispersion of counts than of intrinsic lack of fit of the Hermite polynomial model; in other words, counts in successive watches show greater than

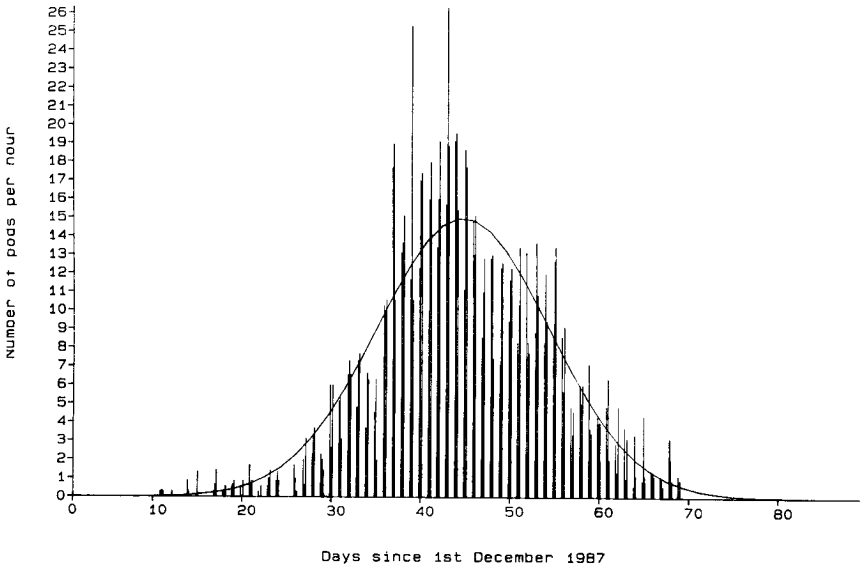


Fig. 6.10. Histogram of number of California grey whale pods sighted, adjusted for watch length, by date, 1987–88 survey. Also shown is the Hermite polynomial fit to the histogram.

Poisson variation. The overdispersion was compensated for by multiplying the Poisson variance on the total count by the dispersion parameter, estimated as the χ^2 statistic divided by its degrees of freedom; this multiplicative correction is sometimes termed a variance inflation factor (Cox and Snell 1989). Thus the dispersion parameter estimate is 334.55/125 = 2.676, giving $\widehat{\text{se}}(n) = \sqrt{(3593 \cdot 2.676)} = 98.1$ pods. The fit of the Hermite polynomial model to the counts yields a multiplicative correction for animals passing outside watch periods of $\hat{f}_t = 2.4178$ with standard error 0.0068.

Swartz *et al.* (1987) reported on experiments in which whales were radio-tagged in 1985 and 1986. Of these, 15 were recorded both at night and in daylight. An unpaired *t*-test on the difference in log day and night speeds revealed no significant difference between Monterey and the Channel Islands ($t_{11} = -1.495$; $p > 0.1$). After pooling the data from both locations, a paired *t*-test revealed a significant difference in log speeds between day and night ($t_{14} = 2.284$; $p < 0.05$). A back-transformation with bias correction gave a multiplicative correction factor for hours of darkness of 1.100 ($\widehat{\text{se}} = 0.045$); thus it is estimated that rate of passage is 10% higher at night, and thus night counts, if they were feasible, would generate counts 10% higher. Counts were carried out for ten hours each day. On average, it is reasonable to add an hour to each end of the day, giving roughly 12 hours of daylight (including twilight) per 24 hours. Thus the multiplicative correction applies approximately to one half of the total number of whales estimated, giving a multiplicative correction factor of $\hat{f}_n = 1.050$ ($\widehat{\text{se}} = 0.023$). Swartz (personal communication) notes that the behaviour of the animals off the Channel Islands is very different from when they pass Monterey. If a correction factor is calculated as above from the nine radio-tagged whales off Monterey that were recorded both during the day and at night, we obtain $\hat{f}_n = 1.020$ ($\widehat{\text{se}} = 0.023$). Although this does not differ significantly from one, we apply it, so that the variance of the abundance estimate correctly reflects the uncertainty in information on this potentially important parameter.

During the 1987–88 season, counts were carried out independently by observers in identical sheds, 5 m apart. Buckland *et al.* (in press) analysed these double count data using the approach of Huggins (1989, 1991) and Alho (1990), which incorporates covariates to allow for heterogeneity in mark–recapture experiments. We summarize the method here. The procedures for matching detections from the two stations are described by Breiwick *et al.* (unpublished). We assume the matches are made without error.

Assuming that the probability of detection of a pod from one station is independent of whether it is detected from the other, and independent

of whether other pods are detected by either station, the full likelihood for all pods passing Monterey during watch periods is

$$L^* = K \cdot \prod_{i=1}^{M} \prod_{j=1}^{2} p_{ij}^{\delta_{ij}} \cdot (1 - p_{ij})^{1-\delta_{ij}}$$

where M = total number of pods passing during count periods,
p_{ij} = probability that pod i is detected from station j, $i = 1, \ldots, M$, $j = 1, 2$,

$$\delta_{ij} = \begin{cases} 1, & \text{pod } i \text{ is detected from station } j, \\ 0 & \text{otherwise,} \end{cases}$$

and K depends on M, but not on the parameters that define p_{ij}.

Huggins (1989) shows that inference can be based on the conditional likelihood,

$$L = \prod_{i=1}^{n} \prod_{j=1}^{2} \pi_{ij}^{\delta_{ij}} \cdot (1 - \pi_{ij})^{1-\delta_{ij}}$$

where n = number of pods detected from at least one station,

and $\pi_{ij} = \dfrac{p_{ij}}{p_i}$

with $p_i = 1 - \prod_{j=1}^{2} (1 - p_{ij})$ = probability that pod i is detected from at least one station.

Thus π_{ij} is the probability that pod i is detected from station j given that it is detected from at least one station.

Both Huggins (1989, 1991) and Alho (1990) model the p_{ij} using logistic regression. Algebra yields:

$$\log_e \frac{\pi_{ij}}{1 - \pi_{ij}} = \log_e \frac{p_{ij}}{1 - p_{ij}} - \log_e p_{ij'}, \quad \text{where} \quad j' = 3 - j$$

Hence logistic regression for the p_{ij} can be obtained simply by carrying out logistic regression for the conditional probabilities π_{ij}, and setting an offset variable, equal to $-\log_e \hat{p}_{ij'}$, for each observation. In the first iteration, the offset variable is set to zero (corresponding to normal logistic regression for π_{ij}). Estimates $\hat{p}_{ij}$ are then calculated from the $\hat{\pi}_{ij}$, from which the offset variable is estimated. The model is refitted, and the process is repeated until convergence is achieved. Model fitting was carried out using Genstat (Genstat 5 Committee, 1987).

Potential covariates were date, Beaufort, components of wind direction parallel and perpendicular to the coast, visibility code, distance offshore, pod size and rate of passage (pods per hour); observer, station and watch period were entered as factors. Estimates $\hat{p}_{ij}$ and $\hat{p}_i$ were calculated from the final iteration, from which M was estimated as

$$\hat{M} = \sum_{i=1}^{n} \frac{1}{\hat{p}_i}$$

with

$$\widehat{\text{var}}(\hat{M}) = \sum_{i=1}^{n} \frac{1}{\hat{p}_i^2} \cdot (1 - \hat{p}_i)$$

Thus a correction factor for pods missed by both stations is given by

$$\hat{f}_m = \frac{\hat{M}}{n}$$

with

$$\widehat{\text{se}}(\hat{f}_m | n) = \frac{\sqrt{\widehat{\text{var}}(\hat{M})}}{n}$$

Probability of detection of a pod was adequately modelled as a function of five covariates: pod size; rate of passage; migration date; visibility code; and the component of wind direction parallel to the coast. None of the factors (observer, watch period, station) explained a significant amount of variation. Probability of detection increased with pod size ($p < 0.001$), with rate of passage ($p < 0.001$) and with migration date ($p < 0.05$), and decreased with visibility code ($p < 0.05$). It was also greater when the wind was parallel to the coast from 330° (slightly west of north), and smaller when from 150° (east of south). The correction factor f_m was estimated by $\hat{f}_m = 1.0632$, with standard error 0.00447.

The number of whales passing Monterey is equal to the number of pods multiplied by the average pod size, which was estimated by the average size of pods detected (excluding those moving north). This gave $\bar{s} = 1.959$ ($\widehat{\text{se}} = 0.020$). A correction factor for mean pod size was calculated using data from Reilly *et al.* (1980), comparing recorded pod sizes with actual pod sizes, determined by observers in an aircraft. For pods of size one, an additive correction of 0.350, with standard error $0.6812/\sqrt{225} = 0.0454$, was used. The correction for pods of size two was

0.178 ($\widehat{\text{se}} = 0.9316/\sqrt{101} = 0.0927$), for pods of size three, 0.035 ($\widehat{\text{se}} = 1.290/\sqrt{28} = 0.244$), and for pods of size four or greater, the correction was 0.333 ($\widehat{\text{se}} = 0.7825/\sqrt{27} = 0.151$). A multiplicative correction factor for mean pod size was then found as:

$$\hat{f}_s = 1 + \frac{0.350n_1 + 0.178n_2 + 0.035n_3 + 0.333n_{4+}}{n \cdot \bar{s}} = 1.131$$

with

$$\widehat{\text{se}}(\hat{f}_s | n) \simeq \sqrt{[(0.0454n_1)^2 + (0.0927n_2)^2 + (0.2438n_3)^2 + (0.1506n_{4+})^2 + 0.6812^2 n_1 + 0.9316^2 n_2 + 1.290^2 n_3 + 0.7825^2 n_{4+}]/(n \cdot \bar{s})} = 0.026$$

where n = total number of pods recorded,
n_i = number of pods of size i, $i = 1, 2, 3$,
and n_{4+} = number of pods of size four or more.

The revised abundance estimate was thus found as follows. Counts of numbers of pods by watch period were combined across the two stations, so that each pod detected by at least one station contributed a frequency of one. The Hermite polynomial model was applied to these counts, to obtain a multiplicative correction factor $\hat{f}_t$ to the number of pods detected for whales passing at night or during poor weather. The correction for different rate of passage at night $\hat{f}_n$ was then made. Next, the multiplicative correction $\hat{f}_m$ was applied, to allow for pods passing undetected during watch periods. The estimated number of pods was then multiplied by the mean pod size, and by the correction factor $\hat{f}_s$ for underestimation of pod size, to obtain the estimate of the number

Table 6.7 Estimates of abundance and of intermediate parameters, California grey whales, 1987–88

Parameter	Estimate	Std error	% contribution to $\widehat{\text{var}}(\hat{N})$	95% confidence interval
E(Number of pods seen by at least one station) $= E(n)$	3593	98	39	(3406, 3790)
Correction for pods passing outside watch periods, f_t	2.418	0.007	0	(2.405, 2.431)
Correction for night passage rate, f_n	1.020	0.023	27	(0.976, 1.066)
Correction for pods missed during watch periods, f_m	1.063	0.004	1	(1.054, 1.072)
Total number of pods passing Monterey	9419	337		(8781, 10 104)
Mean recorded pod size	1.959	0.020	5	(1.920, 1.999)
Correction for bias in recorded pod size, f_s	1.131	0.026	28	(1.081, 1.183)
Total number of whales passing Monterey	20 869	913		(19 156, 22 736)

Table 6.8 Estimated number of pods, pod size and number of whales by year. (Standard errors in parentheses.) For any given fit, the number of parameters is two greater than the number of terms, corresponding to the two parameters of the normal key

Year	No. of terms	χ^2 [df]	Sample size (pods)	Estimated no. of pods	Estimated average pod size	Relative abundance estimate	Absolute abundance estimate
1967–68	4	83.0	903	4051	2.438	9878	12921
		[45]		(253)	(0.063)	(667)	(964)
1968–69	0	70.6	1079	4321	2.135	9227	12070
		[61]		(134)	(0.046)	(348)	(594)
1969–70	1	104.5	1245	4526	2.128	9630	12597
		[67]		(155)	(0.043)	(383)	(640)
1970–71	2	116.2	1458	4051	2.021	8185	10707
		[90]		(115)	(0.033)	(267)	(487)
1971–72	0	71.3	857	3403	2.193	7461	9760
		[56]		(127)	(0.048)	(323)	(524)
1972–73	4	91.5	1539	5279	2.187	11543	15099
		[71]		(152)	(0.034)	(378)	(688)
1973–74	4	133.7	1496	5356	2.098	11235	14696
		[66]		(186)	(0.034)	(431)	(731)
1974–75	0	159.2	1508	4868	2.034	9904	12955
		[74]		(174)	(0.035)	(394)	(659)
1975–76	2	101.1	1187	5354	2.073	11100	14520
		[47]		(218)	(0.039)	(497)	(796)
1976–77	0	139.7	1991	5701	2.052	11700	15304
		[87]		(153)	(0.028)	(353)	(669)
1977–78	0	50.2	657	7001	1.843	12904	16879
		[31]		(356)	(0.046)	(731)	(1095)
1978–79	4	152.9	1730	4970	2.016	10018	13104
		[84]		(159)	(0.034)	(361)	(629)
1979–80	4	109.3	1451	6051	2.068	12510	16364
		[55]		(220)	(0.033)	(498)	(832)
1984–85	3	105.2	1756	7159	2.290	16393	21443
		[49]		(301)	(0.038)	(740)	(1182)
1985–86	1	141.4	1796	6873	2.237	15376	20113
		[104]		(191)	(0.042)	(515)	(927)
1987–88N	3	205.9	2426	7756	2.040	15825	
		[92]		(221)	(0.027)	(497)	
1987–88S	3	152.8	2404	7642	2.104	16082	
		[91]		(194)	(0.029)	(464)	
1987–88 (average)						15954 (481)	20869 (913)

of whales passing Monterey during the 1987–88 migration. Thus the abundance estimate for 1987–88 is given by

$$\hat{N} = n \cdot \hat{f}_t \cdot \hat{f}_n \cdot \hat{f}_m \cdot \bar{s} \cdot \hat{f}_s$$

with

$$\text{cv}(\hat{N}) \simeq \sqrt{\{[\text{cv}(n)]^2 + [\text{cv}(\hat{f}_l)]^2 + [\text{cv}(\hat{f}_n)]^2 + [\text{cv}(\hat{f}_m)]^2 + [\text{cv}(\bar{s})]^2 + [\text{cv}(\hat{f}_s)]^2\}}$$

Table 6.7 shows the different components to the estimate $\hat{N}$. Combining them, estimated abundance is 20 869 whales, with $\text{cv}(\hat{N}) = 0.0437$ and approximate 95% confidence interval (19 200, 22 700).

Buckland and Breiwick (in press) scaled their relative abundance estimates for the period 1967–68 to 1987–88 to pass through an absolute abundance estimate for 1987–88. Rescaling them to pass through the revised estimate above yields the estimates of Table 6.8. Figure 6.11 plots the absolute abundance estimates and shows the estimated increase in abundance assuming an exponential model with non-zero asymptote. The estimated mean annual rate of increase is 3.3% per annum ($\widehat{\text{se}} = 0.4\%$).

6.13 Point-to-object and nearest neighbour methods

The term 'distance sampling' has been used by botanists in particular to describe methods in which a random point or object is selected, and distances from it to the nearest object(s) are measured. A discussion of these methods was given by Diggle (1983: 42–4). In the simplest case, the distance y to the nearest object is measured; y is a random variable with a pdf, say $f(y)$. However, there is no detection function; the nearest object will be detected with probability one. This is very different from the distance sampling from which this book takes its title, for which there is also a sample of distances y with pdf $f(y)$. The two pdf's can be very similar mathematically, but they are conceptually very different. It is the concept of a detection function that distinguishes the distance sampling of this book. Hence we do not describe point-to-object and nearest neighbour methods in detail here.

For point-to-object and nearest neighbour methods, if the distribution of objects is random, then object density is estimated by

$$\hat{D} = \frac{kn}{\pi \sum_{j=1}^{k} \sum_{i=1}^{n} r_{ij}^2}$$

where k = number of random points or objects,

n = number of point/object-to-object distances measured at each point or object,

r_{ij} = distance of ith nearest object to the jth random point or object, $i = 1, \ldots, n$; $j = 1, \ldots, k$.

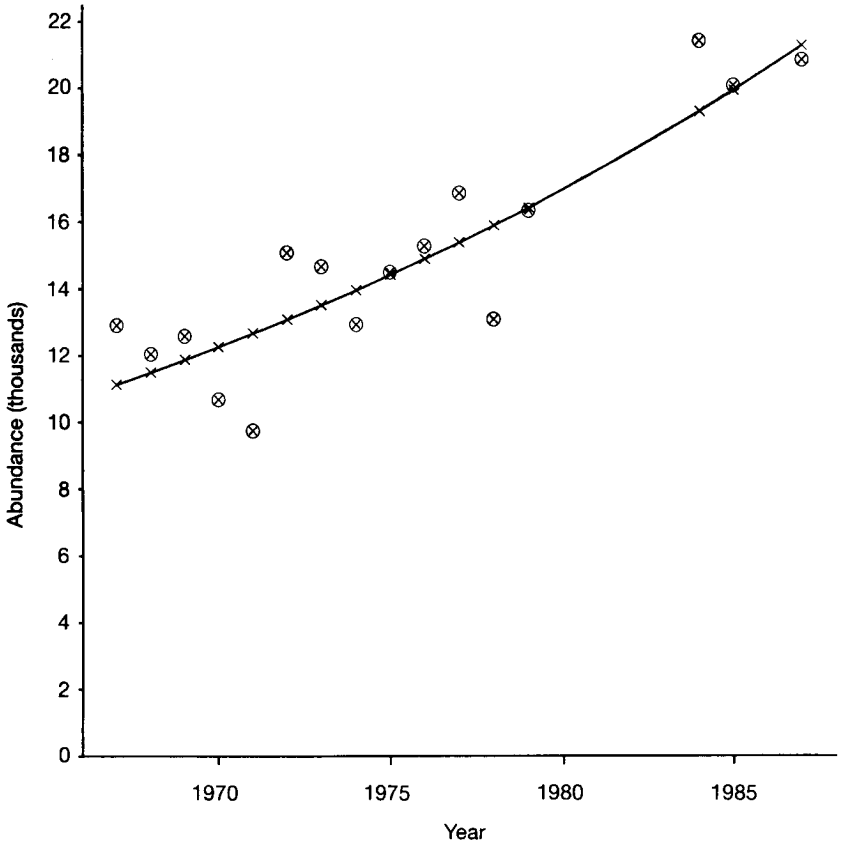


Fig. 6.11. Estimates of abundance by year of California grey whales, and predicted abundance from a weighted exponential regression of abundance estimates on year. Year 1967 signifies winter 1967–68, *etc.*

When the distribution of objects is overdispersed (i.e. aggregated), density is underestimated if distances are measured from a random point, and overestimated if distances are measured from a random object. An average of the two therefore tends to have lower bias than either on its own. Diggle (1983) listed three *ad hoc* estimators of this type.

Some authors have used point-to-object distances only, together with a correction factor for non-Poisson distribution (Batcheler 1975; Cox 1976; Warren and Batcheler 1979), although Byth (1982) showed by simulation that the approach can perform poorly.

Nearest neighbour and point-to-object methods have been used primarily to measure spatial aggregation of objects, and to test the assumption that the spatial distribution is Poisson. Their sensitivity to

departures from the Poisson distribution is useful in this context, but renders the methods bias-prone when estimating object density. Except in special cases, such as estimating the density of forest stands (Cox 1976), we do not recommend these methods for density estimation. Their disadvantages are:

1. All objects out to the nth nearest to the selected point or object must be detected. In areas of low density, this may require considerable search effort.
2. It can be time-consuming to identify which are the n nearest objects, and at lower densities it may prove impractical or impossible to determine them.
3. The effective area surveyed cannot be easily predicted in advance, and is highly correlated with object density; a greater area is covered in regions of low object density. By contrast, good design practice in line and point transect surveys ensures that area covered is independent of object density within strata, leading to more robust estimation of average object density.

Point transects and point-to-object methods may both be considered as generalizations of quadrat counts. In both cases, the quadrat may be viewed as circular. For point transects, the area searched, $a = k\pi w^2$, is fixed (and possibly infinite), but the observer is not required to detect all objects in that area. For point-to-object methods, the number of objects n to be detected from each point is fixed, but the radius about each point is variable; all objects within that radius must be detected.