Line Transect Abundance Estimation with Uncertain Detection on the Trackline

D.L. Borchers

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematical Statistics

UNIVERSITY OF CAPE TOWN

February, 1996

0.1 Abstract

After critically reviewing developments in line transect estimation theory to date, general likelihood functions are derived for the case in which detection probabilities are modelled as functions of any number of explanatory variables and detection of animals on the trackline (i.e. directly in the observer's path) is not certain.

Existing models are shown to correspond to special cases of the general models. Maximum likelihood estimators are derived for some special cases of the general model and some existing line transect estimators are shown to correspond to maximum likelihood estimators for other special cases. The likelihoods are shown to be extensions of existing mark-recapture likelihoods as well as being generalizations of existing line transect likelihoods.

Two new abundance estimators are developed. The first is a Horvitz–Thompson–like estimator which utilizes the fact that for point estimation of abundance the density of perpendicular distances in the population can be treated as known in appropriately designed line transect surveys. The second is based on modelling the probability density function of detection probabilities in the population. Existing line transect estimators are shown to correspond to special cases of the new Horvitz–Thompson–like estimator, so that this estimator, together with the general likelihoods, provides a unifying framework for estimating abundance from line transect surveys.

The performance of the new estimators and an existing mark-recapture Horvitz– Thompson like estimator are compared in a simulation study. The new Horvitz– Thompson like estimator is shown to perform best on statistical criteria. It also has a number of other advantages over the other estimators, not least of which is the fact that it is readily adapted to provide an estimator which is able to correct for random or responsive animal movement.

Unresolved issues in line transect theory are mentioned and suggestions for future reserarch are presented.

Acknowledgements

Firstly to Carol for her support through the long process of completing this thesis, despite the weekends and evenings of my time it has taken from her and Alice.

To Walter Zucchini for the encouragement and inspiration he provided while he was at the University of Cape Town. To Doug Butterworth for reading and commenting on the thesis draft. To Steve Buckland for making me feel I was doing something worthwhile.

Finally, to my mother for her support, financial and otherwise.

Contents

0.2 Summary

One of the key assumptions of line transect (LT) theory is that all animals in the observer's path (on the "trackline") in a LT survey are detected. In many marine LT surveys in particular, this assumption is violated. This is essentially because marine animals are underwater and hence unobservable for a considerable portion of the time. LT theory and methods for the conventional case, in which all animals on the trackline are detected, is well established, and to a large degree the estimation problem for this scenario has been solved. (Note that "conventional LT models" is used to mean "LT models in which all animals directly in the observer's path are detected".) Although some progress has been made in obtaining estimators of abundance when detection of animals on the trackline is not certain, no general theory comparable to that for the conventional LT case exists for this case. A general theory and estimators for the general case are developed in this thesis. The properties of the estimators are investigated by simulation.

The development of conventional LT methods is critically reviewed. This includes methods based on radial distance alone, perpendicular distance alone and bivariate methods which include perpendicular distance and group size as explanatory variables. The review provides an overview of the development of LT theory which is not readily available in existing LT literature. This is followed by a critical review of univariate (based on perpendicular distance only) methods for estimating abundance when detection on the trackline is not certain. One notable omission from the review and the thesis as a whole is methods based on modelling animal availability as a discrete process. These methods are described only in very rough outline. In addition to giving an overview of existing methods and illustrating the relationship between some of the existing estimators, the review highlights a fundamental difference between the conventional case and the case in which detection of animals on the trackline is uncertain. This is that pooling data over explanatory variables results in biased estimation in the latter case. As a consequence of this fact, the conventional LT theory concept of "pooling robust" estimation does not apply when detection of animals on the trackline is not certain. In this case, unbiased inference depends on the availability of adequate models for probability of detection as a function of all explanatory variables which affect detectability substantially. The development of such models has received very little attention in the LT literature. A description is given of a generalization of existing generalized linear model link functions which seems able to provide suitable multivariate models for the detection function when detection of animals on the trackline is uncertain, although this not investigated in any detail in the thesis.

Likelihood functions are developed for the most general case, in which detection probabilities are modelled as functions of any number of explanatory variables and detection on the trackline is not certain. Existing LT models are shown to be special cases of these general models. Maximum likelihood estimators are derived for some special cases, and some existing estimators are shown to correspond to maximum likelihood estimators for other special cases. The general likelihoods represent a marriage of mark-recapture (MR) and LT likelihoods. They are shown to be both extensions of existing MR likelihoods

and generalizations of existing LT likelihoods. The development of the general models highlights the special role that perpendicular distance plays in LT estimation theory. This is a consequence of the fact that, in principle at least, LT surveys can often be designed so that the perpendicular distance from the trackline of any particular animal in the survey region is a uniform random variable.

The models provide a general likelihood framework within which all current LT models can be placed as special cases. To give an overview of the types of models covered, it is useful to classify LT models according to some criteria. The criteria I have used are as follows.

(1) **Univariate or Multivariate?** A LT model is "univariate" if detection probability depends on perpendicular distance (x) only. If it depends on x and other covariates ($\boldsymbol{z}$) it is "multivariate". Note that the use of "multivariate" in this sense is contrary to the conventional use in statistical literature, where it usually implies multiple responses rather than multiple explanatory variables.

(2) **Conventional or MR?** "MR" (for "mark-recapture") indicates a model in which detection of animals which are on the trackline is not assumed to be certain and two independent observers are used. If a LT model is not an MRLT model, it is a "Conventional" LT model.

(3) **Constant Shape or Variable Shape?** A model has a "constant- shape" if variables other than x affect only the scale parameter of the detection function, if they have any effect at all. Obviously all univariate models are constant-shape models.

(4) **Constant-G** or **Variable-G ?** A model is a "constant-G" model if the probability of detecting animals which are on the trackline is not affected by any explanatory variables. All univariate LT models are by definition constant-G models, as are all Conventional LT models.

This classification of LT models is illustrated in Figure 0.1. Detection probability is denoted by $p(\ )$, and "$p(0) = 1$" denotes certain detection on the trackline. The bold boxes correspond to models for which substantial LT theory already exists. Models written in bold type are covered in this thesis (in varying degrees of detail). The shaded box is the most general model, which is the central focus of this thesis.

Two new abundance estimators are developed for the most general case. The first is a Horvitz–Thomompson-like estimator, which utilizes the fact that the density of perpendicular distances in the population of interest can often be treated as known for the purposes of point estimation of abundance in appropriately designed LT surveys. The second new estimator is based on modelling the probability density function of detection probabilities in the population. Existing LT estimators are shown to correspond to special cases of the new Horvitz–Thomompson–like estimator, so that this estimator, together with the general likelihoods, provides a unifying framework for estimating abundance from any LT survey.

The performance of the new estimators, as well as that of an existing MR Horvitz–Thomompson–like estimator, are compared in a simulation study. The study shows that when the perpendicular distances of animals from the trackline can be treated as known for point estimation of abundance, the performance

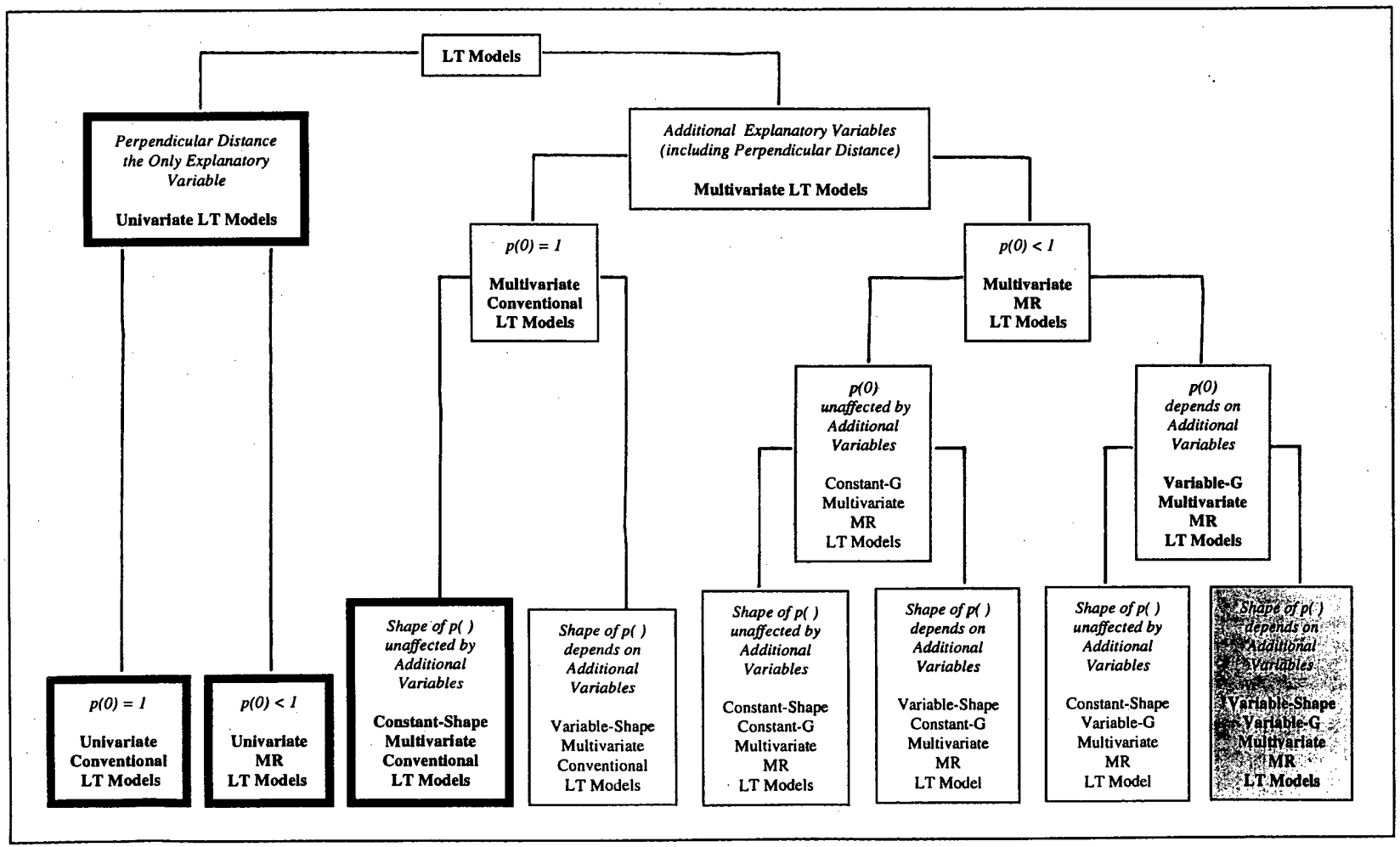


Figure 0.1: A Classification of Line Transect Models

of Horvitz–Thompson–like estimators can be improved substantially if this information is used in the estimation process. The new Horvitz–Thompson–like estimator is shown to perform best overall on selected statistical criteria. The performance of all of the estimators is poor when mean detection probability is low unless sample size is very large. The estimator which models the probability density function of detection probabilities is substantially less biased and less variable than the Horvitz–Thompson–like estimators in such cases, but the estimator which models the probability density function of detection probabilities then has the undesirable property of being an inconsistent estimator for N. The new Horvitz–Thompson–like estimator has a number of advantages over the other estimators, not least of which is the fact that it can be readily adapted to provide an estimator which is able to correct for random or responsive animal movement.

Unresolved issues in LT theory are discussed and suggestions for future research are presented. Primary among these is the development of estimation methods which incorporate uncertainty arising in the identification of duplicates (animals detected by both observers on surveys in which two observers search the same region simultaneously and independently of one another).

Chapter 1

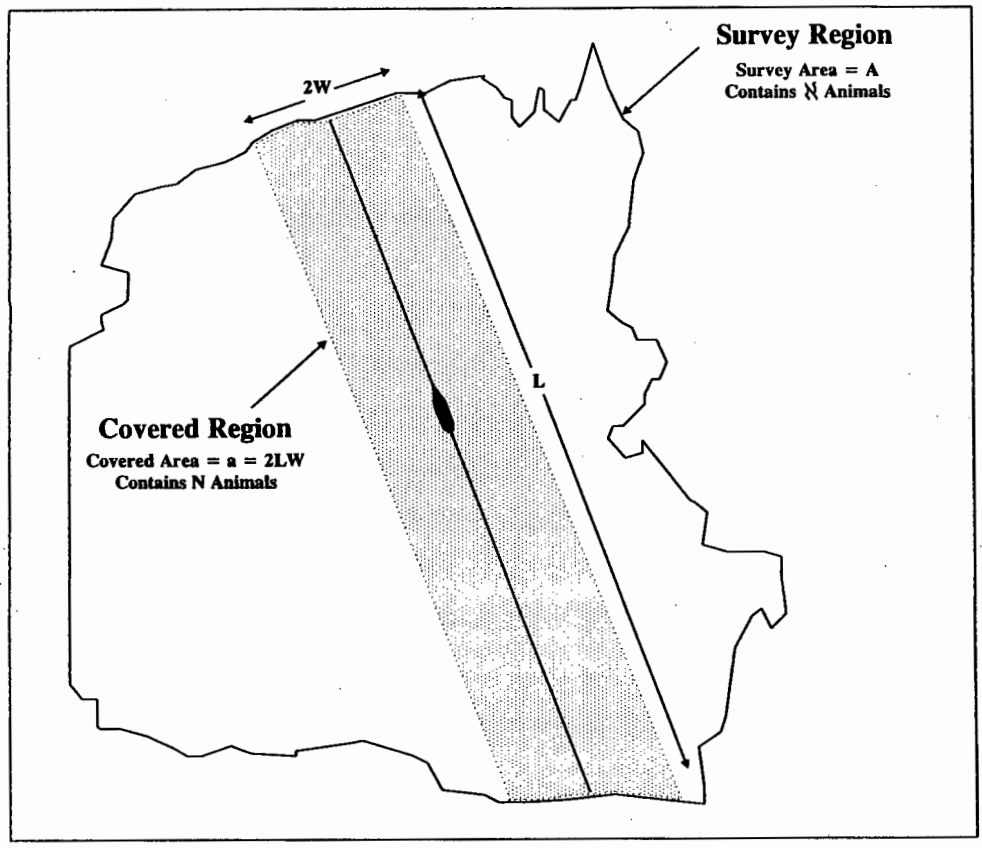
Introduction

Line transect (LT) surveys are used to estimate abundances of a wide variety of fauna, flora and inanimate objects. For simplicity I refer to the objects as "animals" in all cases. My experience with LT surveys is primarily in a marine context and this fact occasionally shows (in the first two figures, for example), but the content of the thesis is quite general. When a group of animals is the unit of detection the group is referred to as the "animal" (see below). I refer to the surveyor(s) at one location in the singular, as the "observer", even though this "observer" may consist of more than one person. In order to refer to observers in the third person I need to nominally assume a gender for observers. I choose to refer to them as female throughout, for a change.

A LT survey consists of an observer traversing a set of "transects" or "tracklines" laid down according to some design in the survey region, searching for animals and recording at least the position at which each animal is detected. Define W to be a maximum perpendicular distance from the trackline beyond which either no animals are detected, or detected animals are discarded for the purposes of abundance estimation. Before going any further, I need to define a few more terms and symbols. Some of these are also illustrated in Figure 1.1.

- **Survey region** refers to the region within which the abundance of animals is to be estimated.
- **Survey area** refers to the surface area of the survey region. A is used to denote survey area.
- **Effort** is the total trackline length. L is used to denote this survey effort in the survey region.
- **Covered region** refers to the region within perpendicular distance W of the trackline. This region is "covered" (albeit incompletely in general) in the sense that it is the region which is searched to generate the data used in abundance estimation.
- **Covered area** refers to the surface area of the covered region. The symbol a is used to denote the covered area. Note that $a = 2LW$.
- **Survey design** refers to the method of choosing the tracklines (or equivalently, conditional on W, the method of choosing the covered region). For example, a common survey design is one which

Figure 1.1: Schematic representation of a LT survey

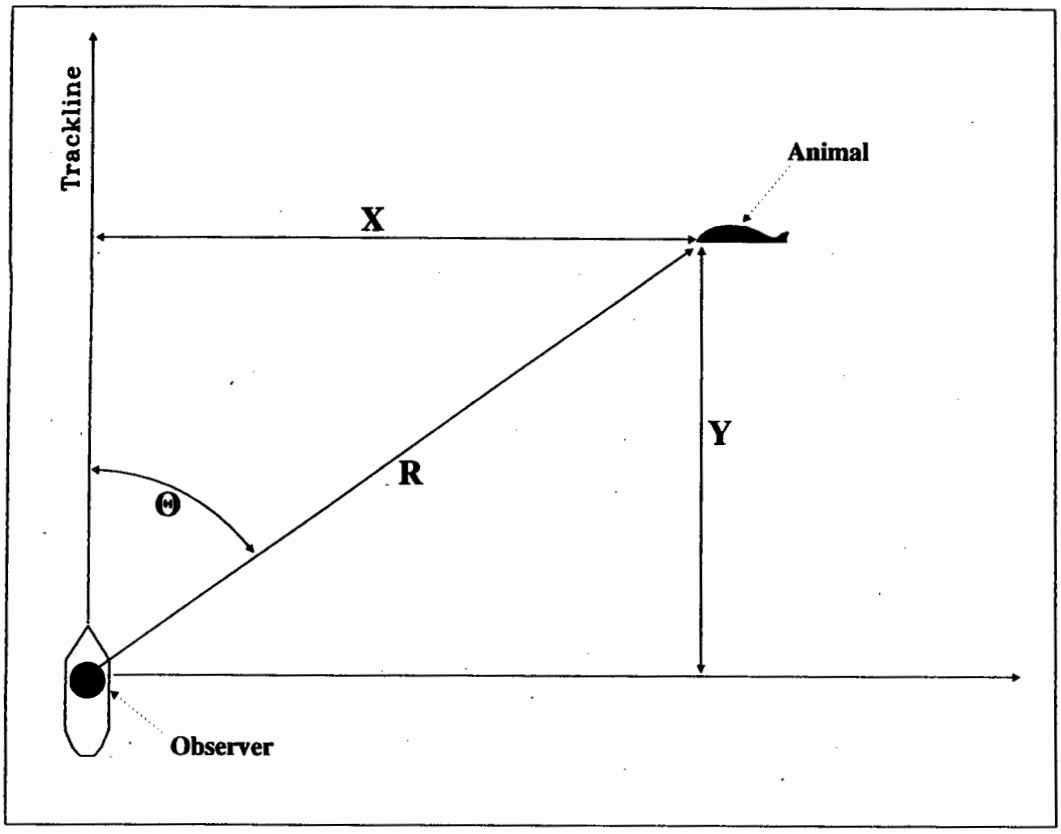


assigns each point in the survey region an equal probability of falling in the covered region. This would be called a design with uniform "coverage probability".

- $\aleph$ is used to denote the number of animals in the survey region.
- N is used to denote the number of animals in the covered region.
- n is used to denote the number of animals detected in the survey region (which is the number of animals detected in the covered region). Various subscripts will be attached to n in what follows. The meaning of the subscripts will be made clear at the point where they are first used.
- X, Y, R and Θ are the perpendicular distance, the forward distance, the radial distance and the angle from the trackline of detected animals (see Figure 1.2). I do not stick to the common convention of using uppercase symbols to denote random variables and lowercase symbols to denote the values which the random variables take, primarily because this adds one more complication to an already complicated notation and because it is usually clear from the context whether reference is being made to a random variable or a value of a random variable.

Most of the LT literature deals with what I call "conventional" line transect models, which are defined at the most basic level by a set of assumptions about the survey process. While different authors use different sets of assumptions to define their "conventional" LT models and to use as a basis from which

Figure 1.2: Schematic representation of Positional Variables



to develop abundance estimators, there is enough common ground between the sets of assumptions to make the concept of "conventional" LT models a useful one. The assumptions which I use to define "conventional" line transect estimators (which I will refer to without the quotation marks from now on) are set down below. They form the core assumptions for most LT models in the literature.

Conventional LT Model Assumptions:

(1) The probability that an animal is detected depends **only** on the animal's position relative to the observation platform.

(2) Animals directly in the observer's path (i.e. on the "trackline") are detected with certainty.

(3) Animals do not move prior to detection and are detected at their exact initial location.

(4) Animals in the covered region are detected independently of one another.

(5) Animals are distributed independently of one another in space.

The last assumption is one which is frequently implicit in models in the literature. It is not important for unbiased estimation of abundance, but most analytic expressions for the variance of the abundance estimators in the LT literature implicitly assume this independence. This is not the case when variance is estimated empirically, from replicate transects, for example.

The primary aim of a LT survey is estimation of $\aleph$. This is generally achieved *via* estimation of N. The survey design is the basis for making inferences about $\aleph$ from estimates of N. The assumptions above provide a basis for estimation of N. With very few exceptions, all LT estimators in the literature implicitly assume a sampling design with equal coverage probability in order to make inferences relating to $\aleph$. The final fundamental assumption defining conventional LT models is therefore as follows.

(6) All points in the survey region are equally likely to fall within the covered region. (This implies that for all animals, the probability that the animal falls within the covered region is $\frac{a}{A}$.)

In many situations these assumptions are reasonable. Possibly the least reasonable is the independence assumption (5), since fauna and flora rarely distribute themselves with complete disregard of the positions of other fauna and flora. Given a survey design with equal coverage probability (assumption (6)), however, assumption (5) is unneccesary for unbiased estimation of abundance. It does have implications for the variance of the estimator, but variance and confidence interval estimation methods are available which do not rely on this independence assumption (see Buckland *et al*, 1993a, for example). I retain the independence assumption for the moment because it is mathematically convenient, although I do discuss relaxing it at various points in this thesis.

Without going into any details here, the options for dealing with violation of assumption (5) are either to use robust methods for variance and confidence interval estimation (as indicated in the previous paragraph), or to assume some other model for the spatial distribution of animals in the survey region. The latter option has received less attention in the LT literature than the former. (Buckland *et al.*, 1993a, contains some discussion of the options and of alternative models for the spatial distribution of animals.)

I retain assumption (6) for all of this thesis, although I do discuss abundance estimation without this assumption very briefly in Chapters 7 and 10. Abundance estimation with unequal coverage probabilities has received very little attention in the LT literature, even though many survey designs (in a marine context in particular) are not designs with equal coverage probability. This is an area of LT theory which needs further development, but it is not an area which I pursue in this thesis.

Assumption (4) is violated if, for example, detecting one animal alerts the observer to the presence of other animals and makes her more likely to detect them as well. The most common cause for this in practice is probably animals clustering together in schools/pods/flocks/etc. When the observer sees one animal in the cluster, she becomes very much more likely to see others in the cluster. This is the last of the assumptions which I do not deal with in detail in this thesis. The particular case that I do deal with to some extent, is the case in which detection of one animal in the group results in detection of all other animals in the group with probability one. This is dealt with by redefining "animal" to be the group, not the individual, and treating group size as an attribute of the "animal". This is not one of the main points of focus of this thesis but I do discuss it briefly it in Chapters 4 and 7 in particular.

The central focus of the thesis is the development of abundance estimation methods for situations in

which assumptions (1) and (2) are simultaneously violated, and to lesser extent, situations in which assumptions (1), (2) and (3) are simultaneously violated. It turns out that estimators developed for situations in which assumptions (1) and (2) are simultaneously violated can relatively easily be combined with an estimator developed by Buckland and Turnock (1992) for the situation in which assumptions (2) and (3) are violated, to provide estimators for the situation in which all three assumptions are violated.

1.1 Thesis Outline

In Chapters 2 and 3 I review the conventional LT models in the literature, focussing on the methods used to model the detection process. I attempt to give an overview of the development of conventional LT models, highlighting common themes which may not have been apparent at the time the estimators were developed. Chapter 2 deals with models based on both the radial and perpendicular distances of detected animals, while Chapter 3 deals with models based on perpendicular distances alone.

Chapter 4 covers perpendicular distance LT models in the literature which deal with violation of assumption (1) only, i.e. models which allow detection probability on the trackline to vary as a function of variables other than perpendicular distance while retaining assumption (2), that detection on the trackline is certain.

Chapter 5 deals with violation of assumption (2) only. In particular, it reviews methods of estimating the probability that animals on the trackline are detected, under the assumptions that not all are detected and that the probability of detection depends only on perpendicular distance.

Chapters 6 and 7 form the core of the thesis. In Chapter 6 I develop a general likelihood framework for LT models in which assumptions (1) and (2) are simultaneously violated. These models are shown to be:

(i) natural generalizations of conventional LT models, and

(ii) particular types of mark-recapture (MR) models which allow capture probabilty to vary as a function of observable explanatory variables, and which incorporate a probability density for these explanatory variables.

In Chapter 7 I show that some existing LT estimators can be derived as maximum likelihood estimators from special cases of the general likelihoods developed in Chapter 6. I also derive new estimators for other special cases. General estimators of abundance for the models of Chapter 6 are then developed, borrowing ideas from both the LT literature and the MR literature. Three forms of abundance estimator are developed for the general case in which both assumptions (1) and (2) are violated. An estimation method for the case in which assumptions (1), (2) and (3) are simultaneously violated is also developed. Some asymptotic properties of the estimators are discussed in this Chapter, but the small-sample properties of the estimators turn out to be analytically intractable.

Chapter 8 describes the simulation model used to investigate some of the small-sample properties of the

three estimators. In Chapter 9 the results of a limited simulation study of these properties are presented and discussed.

Finally, Chapter 10 contains a review of the main results of the thesis and a discussion of directions for future work.

or, transforming from (X, R) to (S, R), where $S = X/R$, they can be estimated from

$$L_{f(r,s)} = \kappa \prod_{j=1}^{n} f_{R,S}(r_j, s_j) \tag{2.14}$$

(see below).

Note that the x's are sufficient for estimation of N when a model for $g(x)$ or $f(x)$ is specified, since equation 2.13 can be written as $\kappa \prod_j f_X(x_j) f_{X|R}(x_j|r_j)$ and $f_{X|R}(x_j|r_j)$ does not involve the parameters of $g(x)$ or $f(x)$. In order to estimate P^{-1} or $f(0)$, and hence N, from perpendicular distance data alone, one needs to model detection probability as a function of perpendicular distance alone. This could be obtained by modelling either $h_Y(y|x)$ or $h_R(r|x)$ and using this to obtain a model for $g(x)$. It could also be obtained by modelling $g(x)$ (or $f(x)$) directly, without reference to $h_Y(y|x)$ or $h_R(r|x)$. Many early LT estimators were based on the observation of radial distances. It seems intuitively obvious that the probability of detecting an animal must be a function of the radial distance of the animal from the observer (because of weakening signal size and strength with distance). This, together with the fact that it may seem *a priori* easier to construct reasonable models for the probability of detection as a function of radial distance, than as a function of perpendicular distance, no doubt contributed to the early prevalence of approaches based on radial distances.

This is not to say that all the early LT estimators were based on radial distances. Aside from early *ad hoc* estimators (which were not based on sound statistical models), there are two main perspectives from which the LT estimation problem was approached. One is the radial distance approach. The other has its roots in quadrat sampling or strip transect sampling, in which it is assumed that all objects in the covered region are detected with certainty, i.e. that $g(x) = 1$ in the covered region. The extension from strip transects to line transects is made by allowing $g(x)$ to decline as x increases. This perspective leads more naturally to models based on perpendicular distances alone. Radial distance models dominated the early days of statistical LT models, and I deal with them first here.

From the radial distance perspective $f(0)$ can be written as

$$f(0) = \int_0^\infty f_{X|R}(0|r) f_R(r) dr = E_R[f_{X|R}(0|r)] \tag{2.15}$$

where $f_{X|R}(0|r)$ is the conditional distribution of observed perpendicular distances given observed radial distance r, and evaluated at $x = 0$, while $f_R(r)$ is the marginal density of observed radial distances.

If $f_{X|R}(0|r)$ were known, an estimator of $f(0)$ based on radial distances alone could be obtained from the radial distance data alone, since $f(0)$ would be a function of radial distances alone. In this case the following is an unbiased estimator of $f(0)$.

$$\hat{f}(0) = \frac{1}{n}\sum_{i=1}^{n} f_{X|R}(0|r_i) \tag{2.16}$$

An additional assumption needed for estimators based on radial distances (over and above those needed for estimation based on perpendicular distances), is contained in the specification of a form for $f_{X|R}(0|r)$. (This specification may or may not be made *via* explicit assumptions about the radial hazard function $h_R(r|x)$; but if not, the specification will contain implicit assumptions about $h_R(r|x)$.)

2.2 Independent Radial Distances and Angles

The estimation problem is simplified if it can be assumed that the observed radial distances (r_j, $j = 1 \ldots n$) and the observed detection angles (θ_j, $j = 1 \ldots n$) are independent, since in this case estimation of the parameters of the pdf of observed radial distances, $f_R(r)$, and of the pdf of observed angles, $f_\Theta(\theta)$, can be performed separately. Not surprisingly therefore, most radial distance based estimators for LT surveys rest on this assumption. Some general results which are useful in this case are given below, before specific models and estimators are considered.

Using the fact that the random variables R and Θ are independent (and therefore that R and $S = \sin(\Theta)$ are independent), and changing variables from (R, S) to (R, X) one has

$$\begin{aligned} f_{R,S}(r,s) &= f_R(r)\, f_S(s) \\ \Rightarrow f_{R,X}(r,x) &= f_R(r)\frac{1}{r} f_S\left(\frac{x}{r}\right) \\ \Rightarrow f_{X|R}(x|r) &= \frac{1}{r} f_S\left(\frac{x}{r}\right) \\ \Rightarrow f_{X|R}(0|r) &= \frac{1}{r} f_S(0) \end{aligned} \tag{2.17}$$

Therefore

$$\begin{aligned} f_X(0) = f(0) &= E_R[f_{X|R}(0|r)] \\ &= E_R\left[\frac{1}{r}\right] f_S(0) \end{aligned} \tag{2.18}$$

With independent R and S, $E_R[r^{-1}]$ can be estimated from the observed radial distances alone (call the estimator $\hat{r}_{inv}$), and $f_S(0)$ can be estimated from the observed angles alone (call the estimator $\hat{s}_0$). Given $\hat{r}_{inv}$ and $\hat{s}_0$, $f(0)$ can be estimated by

$$\hat{f}(0) = \hat{r}_{inv}\, \hat{s}_0 \tag{2.19}$$

The variance of $\hat{f}(0)$ is usually estimated as follows (see Seber, 1982, for example).

$$\hat{Var}[\hat{f}(0)] = \hat{f}(0)^2 \{cv^2[\hat{r}_{inv}] + cv^2[\hat{s}_0] + cv^2[\hat{r}_{inv}]cv^2[\hat{s}_0]\} \tag{2.20}$$

Burnham, *et al.* (1980) note that for reasonable data (with a $cv^2[\hat{r}_{inv}]$ and $cv^2[\hat{s}_0]$ both less than 0.5) negligible inaccuracy is introduced into the variance estimator by neglecting $cv^2[\hat{r}_{inv}]cv^2[\hat{s}_0]$. They argue that the following is a sufficiently accurate estimator of $Var[\hat{f}(0)]$.

$$\hat{Var}[\hat{f}(0)] = \hat{f}(0)^2 \{cv^2[\hat{r}_{inv}] - cv^2[\hat{s}_0]\} \tag{2.21}$$

2.2.1 Hayne-like Models

Within the class of conventional line transect models, what I call Hayne-like models are uniquely characterized by the fact that they can be represented as continuous hazard-rate processes in which **the radial hazard function is a function of radial distance alone**. When this is the case, the observed radial distances and detection angles are independent and $s_0 = 1$ (Theorem 1 of Appendix 2.6.1). This simplifies the estimation problem substantially, as the radial distance data on their own are then sufficient to estimate $f(0)$.

Hayne-like models come in two varieties: those which assume a specific from for the probability of detection as a function of radial distance, and those that do not. The latter, relying on weaker assumptions about the detection process, are more robust. In contrast, the former may be more efficient when they correspond to the true detection process, but run an associated risk of model misspecification.

Hayne's Nonparametric Estimator

Hayne (1949) proposed the original robust form of this class of estimator. This was the first LT estimator which was based on a sound statistical foundation. (Before 1949, a number of *ad hoc* radial distance estimators had been proposed and used.) The estimator has been derived in a number of ways. Hayne (1949) derived it from the assumption that each animal has a detection circle of fixed radius associated with it, such that animal will be detected immediately the observer enters the detection circle, but not before. Eberhardt (1978) derived the same estimator by assuming the existence of a radial hazard function which is a function of the radial distance of the observer from the animal only.

The two approaches are indistinguishable from each other using the LT survey data, and both approaches yield $f_S(0) = 1$. From the above it follows that

$$f(0) = E_R\left[\frac{1}{r}\right] \tag{2.22}$$

provided that this expectation exists. Hayne's nonparametric estimator is

$$\hat{f}_H(0) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{r_j} = \frac{1}{\bar{r}_h} \tag{2.23}$$

(The subscript "H" is for "Hayne" and $\bar{r}_h$ is the harmonic mean of the observed r's.) It is easy to show that the estimator is unbiased under the model assumptions. Assuming independence, an unbiased estimator of $Var[\hat{f}_H(0)]$ is

$$\hat{Var}[\hat{f}_H(0)] = \frac{1}{n(n-1)}\left\{\sum_{j=1}^{n}\left(\frac{1}{r_j}-\frac{1}{\bar{r}_h}\right)^2\right\} \tag{2.24}$$

as each $\frac{1}{r_i}$ is an unbiased estimator of $f(0)$.

Parametric Hayne-like Estimators

Hayne's estimator involves no assumptions about the form of the radial hazard function, other than the implicit assumption that it depends only on radial distance. A number of authors after Hayne derived estimators which involve some additional assumptions about the form of the radial hazard function, generally expressed in terms of the distribution of observed perpendicular or radial distances, rather than in terms of the radial hazard function itself.

(1) Gates' Model

After observing that radial distances for given perpendicular distances from line transect surveys of ruffed grouse appeared to follow a negative exponential distribution, Gates (1969) proposed the following parametric model with negative exponential form for both perpendicular distances and radial distances conditional on perpendicular distances.

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} \\ f_{R|X}(r|x) &= \lambda e^{-\lambda(r-x)} \end{aligned} \tag{2.25}$$

This leads to the marginal distributions for R and S given below, from which it is apparent that R and S are independent.

$$f_R(r) = \lambda^2 r e^{-\lambda r} \tag{2.26}$$

$$\begin{aligned} f_S(s) &= \int_0^\infty \lambda^2 r e^{-\lambda r} dr \\ &= 1 \end{aligned} \tag{2.27}$$

Since, in addition to R and S being independent, $f_S(s) = 1$, the radial hazard function must depend only on R (Theorem 1 of Appendix 2.6.1) and the model is seen to be a particular case of a Hayne-like model, one in which the distribution of detected radial distances has the specific function form given in equation 2.26 above. With a parametric form for $f(0)$, estimation can be based on maximising the appropriate likelihood. The likelihood of observed radial distances is as follows.

$$L_{f(r,s)} = \kappa \prod_{j=1}^{n} \lambda^2 r_j e^{-\lambda r_j} \tag{2.28}$$

The mle for λ is $\frac{2}{\bar{r}}$, where $\bar{r}$ is the arithmetic mean of the observed radial distances. $E[\frac{1}{r}]$ is easily seen to be λ, so that $\hat{\lambda}_{mle} = \frac{2}{\bar{r}}$ is the mle for $f(0)$. The mle is biased and Gates proposed the unbiased estimator $\hat{\lambda}_G = (2n-1)\hat{\lambda}_{mle}/(2n)$. (The subscript "$G$" is for "Gates".)

(2) **An attempt to generalize Gates' Model**

Since the negative exponential distribution is a special case of the Gamma distribution, Sen, Tourigny and Smith (1974) tried to generalize Gates' model using the Gamma distribution. Specifically, they assumed the following two functional forms for $g(x)$ and $f_{R|X}(r|x)$.

$$\begin{aligned} g(x) &= 1 - \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} r^{\alpha-1} e^{-\beta r} dr \\ &= 1 - G_R(x) \end{aligned} \tag{2.29}$$

$$f_{R|X}(r|x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (r-x)^{\alpha-1} e^{-\beta(r-x)} \tag{2.30}$$

($G_R(x)$ is the cumulative density function (cdf) of a Gamma random variable with parameters α and β, evaluated at x.) Note that when $\alpha = 1$ this reduces to Gates' negative exponential model.

However, in deriving the marginal distribution of radial distances, Sen *et al.* (1974) erroneously assumed that $f(x) = G'_R(x)$ (where $G'_R(x)$ is the derivative of $G_R(x)$). This is only true if $G_R(x)$ is the negative exponential cdf - a fact pointed out by Burnham and Anderson (1976). Their estimator is therefore based on contradictory assumptions, and as such is really an *ad hoc* estimator.

(3) **Other Hayne-like Models**

Hayes and Buckland (1983) considered another form for the radial hazard function as a function of r only - namely $h_R(r|x) = ar^{-\beta}$. They put this forward as an illustration of their hazard rate model rather than as a form to be used in practice, and did not develop a radial distance estimator for this case. Butterworth (1982b) and Schweder (1974, 1977) also proposed models in which the radial hazard function depends only on radial distance, but neither developed conventional estimators based on radial distance for their models. (Schweder later went on to develop estimators based on radial distance for the situation in which animals are not continuously available for detection. These are mentioned briefly in Chapter 5.)

2.2.2 Generalizations of Hayne-like Models

Some attempts were made to relax the rather strong assumption of Hayne-like models that the radial hazard function is a function of radial distance alone, and so generalize Hayne-like estimators while retaining the statistically convenient assumption that observed radial distances and angles are independent of each other. This was done by proposing different distributions (other than uniform) for S.

The "Modified" Hayne Model

This model was developed by Burnham and Anderson (1976). They argued that a uniform distribution of S corresponded to a situation in which animals are detected when they flush in response to the approach of the observer (as in surveys of grouse, for example). A uniform distribution for Θ was supposed to correspond to a situation in which observers search uniformly through 90^o either side of the trackline for inanimate objects. Any particular survey, it was argued, can be considered to be a mixture of these two extreme cases. The distribution of S is therefore assumed to be a mixture of the uniform and the distribution corresponding to uniform Θ, as follows.

$$f_S(s) = (1-\gamma)f_1(s) + \gamma f_2(s) \tag{2.31}$$

where

$$f_1(s) = 1 \tag{2.32}$$

$$f_2(s) = \frac{1}{90^o\sqrt{1-s^2}} \tag{2.33}$$

and γ is an unknown mixing parameter. Because R and Θ are assumed to be independent, γ can be estimated from the observed angles alone. The mle for γ is not in closed form and Burnham and Anderson (1976) proposed a moment estimator, and an *ad hoc* estimator of γ which results in $f_S(s) = f_1(s)$ when the mean detection angle ($\bar{\theta}$) is 32.7^o (which is the mean detection angle when S is uniform) and in $f_S(s) = f_2(s)$ when it is 45^o (the mean detection angle when Θ is uniform).

2.3 Other Radial Distance Models

2.3.1 Elliptic Hayne Models

Burnham's Model

Burnham (1979) proposed another generalization of the distribution of S. It is based on a generalization of the idea that animals have circular detection areas (originally "flushing" areas for birds which flush when approached by an observer) associated with them, such that they are detected immediately an

observer enters the area, but not before. The generalization was to allow for elliptic detection areas. The model allows detection ellipses of unknown shape and size, subject to the constraints that the shape of all the ellipses is the same and their major axis are all oriented in the same direction - either parallel, or perpendicular to the trackline. The idea was to allow animals direct approach by the observer to have a different probability of detection at a given radial distance than that for animals located far from the trackline at the same radial distance. The generalization introduces a single extra parameter into the pdf of S, namely a parameter c for the shape of the ellipses - defined as the ratio of the length of the axis parallel to the trackline (V) to that perpendicular to the trackline (Z).

Using the assumption that animals are distributed uniformly with respect to the trackline, Burnham (1979) derived the following pdf for S (where $c > 0$).

$$f_S(s) = \frac{c}{(1-(c^2-1)s^2)^{\frac{3}{2}}} \tag{2.34}$$

It is $f_S(0)$ that we want to estimate (see equation 2.18), so the parameter of interest is c because $f_S(0) = c$. The mle is not in closed form. Burnham initially proposed a computationally simpler moment estimator however, after simulation testing of the moment estimator and the mle, he recommended use of the latter.

It turns out, however, that the assumtion of independence of S and R is not consistent with the existence of detection ellipses with fixed shape. This can be seen from the following argument (which is, I think, easier to follow than that of Otten and de Vries (1984), who originally exposed the contradiction). Since all the ellipses are the same shape and are oriented in the same direction, ellipse size (Z, say, where Z is the length of the ellipse axis perpendicular to the trackline) is the only ellipse characteristic that varies between animals. The model implicitly assumes the existence of a pdf for ellipse sizes ($f_Z(z)$) in the population.

To see that R and $S = \sin(\Theta)$ are not independent, consider the conditional distribution of observed radial distance R, given the observed detection angle Θ. Because of the constant shape of the ellipses, the pdf of animals seen at angle θ which were seen at radial distance r, is just equal to the pdf of animals with detection ellipses of size

$$z(r,s) = \sqrt{r^2(c^2+(1-c^2)s^2)} \tag{2.35}$$

namely $f_Z(z(r,s))$. The fact that this is a function of s unless $c = 1$ (the circular detection area case) implies that R and S are **not independent** unless $c = 1$ (in which case the ellipses are circles). The model is not, therefore, a Hayne-like model.

de Vries's Model

de Vries (1979) used Burnham's idea of elliptic detection areas - but without the assumption that R and S are independent - to derive an estimator of abundance based on non-contradictory assumptions about the detection process. He replaced the assumption that they are independent, with an assumption that S (or Θ) and Z are independent and then used the $f_S(s)$ for detection ellipses derived by Burnham (1979) to derive an estimator of abundance as follows. If animals are distributed uniformly about the trackline then $f_{X|Z}(x|z) = \frac{1}{z}$ so that $f_X(0) = E[\frac{1}{z}]$. However, Z is not observable; it is a function of the unknown ellipse shape parameter, c, as follows.

$$z = \frac{r}{c}\sqrt{1+(c^2-1)s^2} \tag{2.36}$$

The first step of de Vries' estimation procedure involves estimating c as per Burnham (1979). Estimates of the z_j's are then obtained using this estimate of c ($\hat{c}$) in place of c in equation 2.36 and an estimate of $f_X(0)$ is obtained conditional on these estimated $\hat{z}$'s as follows.

$$\hat{f}_X(0) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\hat{z}_j} \tag{2.37}$$

2.4 A Horvitz–Thompson Perspective

Before moving on, I want to highlight a property of two of the radial distance based estimators of this section. I draw attention to it primarily because it has parallels with developments in later Chapters, particularly Chapter 7.

Assume either the circular detection area model or the elliptic detection area model applies. Then from simple geometric arguments (and assuming a sampling design with equal coverage probability), the respective probabilities of including in the sample the jth animal with detection radius r_j in the circular case, and the jth animal with ellipse width (perpendicular to the trackline) z_j in the elliptic case are as follows.

$$p_j = \frac{r_j}{W} \qquad \text{(circular)} \tag{2.38}$$

$$p_j = \frac{z_j}{W} \qquad \text{(elliptic)} \tag{2.39}$$

Given p_j, $(j = 1 \ldots n)$ from the LT survey, a Horvitz–Thompson estimator (Horvitz and Thompson, 1952) of N is simply

$$\hat{N} = \sum_{j=1}^{n}\frac{1}{p_j} \tag{2.40}$$

In the circular detection area model, the p_j's are observed because they depend only on the observed r_j's and Hayne's nonparametric estimator is the Horvitz–Thompson estimator.

$$\begin{aligned} \hat{N}_H &= W n \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{r_j} \right) \\ &= \sum_{j=1}^{n} \frac{W}{r_j} = \sum_{j=1}^{n} \frac{1}{p_j} \end{aligned} \tag{2.41}$$

Recognising this does not lead to any dramatic practical advantage. Under strong independence assumptions, and within a design-based inference paradigm, nbiased estimators of the variance of $\hat{N}$ and $\hat{f}(0)$ are as follows (see Thompson, 1993, p49, for example).

$$\widehat{Var}[\hat{N}_H] = \sum_{j=1}^{n} \frac{1 - p_j}{p_j^2} \tag{2.42}$$

$$\widehat{Var}[\hat{f}(0)] = \frac{\widehat{Var}[\hat{N}_H]}{n^2 W^2} \tag{2.43}$$

In the case of elliptic detection areas, the p_j's are not observable because the z_j's are not observable. The distribution of detection angles is used to estimate the ellipse shape parameter, c, and hence to estimate the detection probabilities which are functions of c, r and $s = \sin(\theta)$.

$$\hat{p}_j = \frac{\hat{z}_j}{W} = \frac{1}{W} \left(\frac{r_j}{\hat{c}} \sqrt{1 + (\hat{c}^2 - 1) s_j^2} \right) \tag{2.44}$$

where $\hat{c}$ is Burnham's (1979) estimator of c. Conditional on the $\hat{p}_j$'s, de Vries's (1979) estimator is a Horvitz–Thompson estimator of abundance.

$$\begin{aligned} \hat{N} &= W n \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\hat{z}_j} \right) \\ &= \sum_{j=1}^{n} \frac{W}{\hat{z}_j} = \sum_{j=1}^{n} \frac{1}{\hat{p}_j} \end{aligned} \tag{2.45}$$

In this case the properties of the Horvitz–Thompson estimator cannot be applied directly because p_j is estimated. In particular the estimator is not necessarily unbiased and the Horvitz–Thompson estimator of the variance of $\hat{N}$ (equation 4.20) will underestimate the variance of $\hat{N}$ because the component of variance due to estimation of p_j is neglected.

2.5 The Failure of Models Based on Radial Distance

Hayne-Like Models

The fundamental assumptions of the models on which Hayne-like estimators are based (namely that R and S are independent, and that $f_S(0) = 1$; or equivalently that the radial hazard function is a function of R alone) are statistically convenient but are not based on substantive information about the detection process. The assumptions can be tested, but the power of tests to detect deviations from them is typically low in practice (Burnham *et al.*, 1980).

Because Hayne's estimator is unbiased under a variety of models of the detection process (namely any model which has the radial hazard function depending only on radial distance), Eberhardt (1978) called it model robust. This is misleading; the estimator is certainly model robust within the class of models having their radial hazard functions depending only on radial distance, but this is a very restrictive class of models. The prospects for model-robustness for the parametric Hayne-like estimators are worse than those for Hayne's estimator, because they involve more restrictive assumptions.

Although under the assumptions on which they are based, both Hayne's and Gates' estimators are unbiased, studies by Robinette *et al.* (1974), Burnham *et al.* (1980) and others show both to be often substantially biased in practice. Robinette *et al.* (1974) assessed the performance of Hayne's estimator and Gates' estimator (among others) on 20 datesets from populations of known density. They found that the mean detection angle was usually significantly greater than 32.7^o - its expected value under the assumptions of the model on which the two estimators are based. (This is not to say that the assumptions never hold - Burnham *et al.* (1980) quote cases in which the mean detection angle is close to 32.7^o.) Robinette *et al.* (1974) also found Hayne's estimator to be positively biased by an average of 47%, and Gates' estimator to be positively biased by an average of 186%, for the datasets they investigated. Burnham *et al.* (1980) give examples of studies in which the mean detection angle was greater than 32.7^o, for which Hayne's estimator was biased by as much as 106%.

Comparisons by Sen *et al.* (1974) of the goodness of fit of the negative exponential and the Gamma distributions to three real radial distance datasets, indicated that a Gamma distribution fitted these data as well or better than the negative exponential. They took this to be a recommendation for their estimator and model - which it is not, because of their mistake in deriving the marginal distribution of R.

Hayes and Buckland (1983) exposed a feature of all Hayne-like models which does not fit well, conceptually, with many line transect surveys. Using a nonspecific hazard-rate model for the detection process, they pointed out that the assumptions underlying Hayne-like models imply that animals closer to the trackline run a higher risk of being detected, per unit time, than do animals farther from the trackline at the same R. They explain clearly how this comes about and I therefore won't repeat the full explanation here. Briefly, it is a result of the facts that (1) the radial hazard function is a function of R alone, and

(2) R changes more slowly the further the animal is from the trackline. For many types of target object (animal or otherwise) one would expect that if the observer is searching uniformly over a range of angles, all target objects in this range at the same radial distance would be equally at risk of detection. This is not the case with Hayne's model. If it were more the case, the mean detection angle would be greater than 32.7^o - which it is frequently observed to be in practice.

Both conceptually and in practice, Hayne-like models are not appropriate for many line transect surveys. If mean detection angle from a survey is not significantly different from 32.7^o and a test of independence of R and S (or R and Θ) is not significant, a Hayne-like estimator may be appropriate. If not, the estimator is likely to yield a substantially positively biased estimate of abundance. Even when the tests are not significant, there is substantial risk of the estimate being biased as the tests typically have low power to detect departures from the assumptions of the model. Using parameteric estimators like that of Gates (1969), or the Gamma-based estimator developed here instead of Hayne's estimator, will only increase this risk since the estimators are based on still more restrictive assumptions than Hayne's estimator.

Generalizations of Hayne-like Estimators

Relatively little investigation of either the theoretical or practical properties of the generalized estimators has been carried out. Both the "Modified" Hayne estimator and Burnham's (1979) elliptic detection area estimator are somewhat *ad hoc* (the latter because of contradictory assumptions made in its derivation). And while Otten and de Vries (1984) showed that de Vries' (1979) estimator is asymptotically unbiased, Hayes and Buckland (1983) questioned the validity of the assumptions on which the generalizations of Hayne's estimator are based.

Because a circular detection area model (i.e. one with uniform observed S and with S independent of R) corresponds to a model which is often conceptually inappropriate, basing a mixture model (the "Modified" Hayne estimator) on this model seems unwise. The other component distribution of the mixture is open to similar criticism; while it seems reasonable to expect observed detection angles to be uniformly distributed when a stationary observer scans uniformly through a range of angles for inanimate target objects, it does not follow that this should be the case when the observer is moving. In short, there is little conceptual justification behind assuming a uniform distribution for either Θ or S - and there is no reason to expect that a mixture of two possibly inappropriate distributions will produce a more appropriate one.

Hayes and Buckland (1983) raised similar conceptual problems in relation to models with elliptic detection areas. It seems more likely that objects close to the trackline will be detected at larger radial distances than objects far from the trackline, than that the reverse would be true. This would correspond to detection ellipses with their major axis *parallel* to the trackline - which corresponds to mean detection angles *less than* 32.7^o. In fact, however, most LT surveys have a mean detection angle greater than 32.7^o. This corresponds to a situation in which the major axis of the ellipses is perpendicular to the trackline - a situation which causes some conceptual difficulty.

The generalizations of Hayne–like estimators have not been widely used, and there is not much evidence on which to judge them in this regard. Judged in terms of the nature of the detection processes which the models imply, none of them do well. The assumptions of the models underlying these estimators are of doubtful validity, and their utility is at best uncertain as a result.

2.5.1 Summary

While the development of Hayne's estimator was a landmark in conventional line transect history because it was the first estimator to be based on a coherent statistical model, the assumptions of this model have proved to be too restrictive for general application to LT surveys. In particular, any model which implies (as does Hayne's) that the radial hazard function is a function of the radial distance of the animal from the observer alone, appears to be inappropriate for all but a very narrow class of LT surveys. Models which have attempted to generalize Hayne's model have, with the exception of one (that of de Vries, 1979), done so while retaining the convenient but restrictive assumption that observed radial distances and detection angles are independent. They have simply involved convenient but not necessarily reasonable forms for the distribution of S (which under Hayne's model is assumed to be uniform). This manoeuvre results in estimators of $f_X(0)$ and abundance which are appealingly simple extensions of Hayne's estimator; the new estimators are simply Hayne's estimator multiplied by an estimator of $f_S(0)$.

None of the resulting estimators are model robust and all are based on doubtful, albeit convenient assumptions about the nature of the detection process. While the single estimator which does not assume independence of R and Θ (de Vries' estimator) has not been tested for model robustness, it is based on the idea of detection ellipses - an idea which does not fit very well with mean detection angles greater than 32.7^o (as frequently observed).

In any particular study, it may happen that one of these proposed estimators may be appropriate. But tests of the assumptions on which the estimators are based typically have low power, so that there is a real risk of model misspecification going undetected. The central problem with radial distance estimators is one of model robustness. Given that the associated tests have little power, it is important to use estimators which are relatively insensitive to violation of these assumptions. Conventional LT estimators based on radial distances do not enjoy this property.

2.6 Appendices

2.6.1 Some results from a Hazard-rate representation of the Detection Process

Let $g_{R|X}(r|x)$ be the probability that an animal at perpendicular distance x is detected by the time it is at a radial distance r from the observer, and let $g'_{R|X}(r|x)$ be the derivative of this function with respect to R.

Using a standard survival analysis result, $h_R(r|x)$ can be expressed as follows (see Cox and Oakes, 1984, for example).

$$h_R(r|x) = \frac{g'_{R|X}(r|x)}{g_{R|X}(r|x)} \tag{2.46}$$

Note that the function $g_{R|X}(x|x)$ (i.e. $g_{R|X}(r|x)$ evaluated at $r = x$) is the perpendicular distance detection function, $g(x)$. The cumulative distribution function for observed radial distances, conditional on an animal being at perpendicular distance x is as follows.

$$\begin{aligned} F_{R|X}(r|x) &= \frac{Pr\{\text{detect animal by radial distance } r \,|\, x\}}{Pr\{\text{detect animal} \,|\, x\}} \\ &= \frac{Pr\{\text{detect animal by radial distance } r \,|\, x\}}{Pr\{\text{detect animal by radial distance } x \,|\, x\}} \\ &= \frac{g_{R|X}(r|x)}{g(x)} \end{aligned} \tag{2.47}$$

Theorem 1 *$h_R(r|x)$ depends only on radial distance iff R and S are independent and $f_S(s) = 1$.*

Proof:

($\Rightarrow$) Assume that $h_R(r|x)$ is independent of x. It follows that $g_{R|X}(r|x)$ does not depend on x, except insofar as x determines the minimum r below which $g_{R|X}(r|x) = 0$ (because animals are assumed to be detectable with probability 1 if r gets to the given value of x). For brevity $g_{R|X}(r|x)$ is abbreviated to $g_R(r)$. Now

$$\begin{aligned} F_{R|X}(r|x) &= \frac{g_{R|X}(r|x)}{g(x)} \\ &= \frac{g_R(r)}{g(x)} \end{aligned} \tag{2.48}$$

Let $g'_R(r) = \frac{d}{dr} g_R(r)$. Then the conditional pdf of R, given x is $f_{R|X}(r|x) = g'_R(r)/g(x)$. Let $\omega = \int_0^\infty g(x)\,dx$. Then the joint density of observed R and X, the joint density of observed R and S, and the marginal density of observed R, are as follows when $h_R(r|x)$ is a function of r alone.

$$f_{R,X}(r,x) = f_{R|X}(r|x) f_X(x) = \frac{g'_R(r)}{g(x)} \frac{g(x)}{\omega} \qquad (x \le r < \infty)$$

$$= \frac{g'_R(r)}{\omega} \tag{2.49}$$

$$f_{R,S}(r,s) = \frac{g'_R(r)}{\omega} r \tag{2.50}$$

$$f_R(r) = \int_0^1 \frac{g'_R(r)}{\omega} r\,ds$$

$$= \frac{g'_R(r)}{\omega} r \tag{2.51}$$

Hence $f_{R,S}(r,s) = f_R(r)$, and R and S are independent. It also follows that $f_S(s) = 1$.

($\Leftarrow$) Now assume that R and S are independent and that $f_S(s) = 1$. Then $f_{R,S}(r,s) = f_R(r)$ and, changing variables from (R,S) to (R,X):

$$f_{R,X}(r,x) = f_R(r)\frac{1}{r} \qquad (0 \le x \le r) \tag{2.52}$$

$$f_{R|X}(r|x) = \frac{f_R(r)}{r} \frac{1}{f_X(x)} \tag{2.53}$$

From equation 2.47, $g_{R|X}(r|x) = g(x) F_{R|X}(r|x)$. Further, using equation 2.52 and the fact that $g(x) = \mu f(x)$ (where $\mu = \int g(x)\,dx$),

$$g_{R|X}(r|x) = \mu f(x) \frac{1}{f(x)} \int_r^\infty \frac{f_R(r^*)}{r^*} dr$$

$$= \mu \int_r^\infty \frac{f_R(r^*)}{r^*} dr^* \tag{2.54}$$

Thus $g_{R|X}(r|x)$ is independent of x, and $g'_{R|X}(r|x)$ is independent of x. Substituting into equation 2.46, it follows that the radial hazard funciton $h_R(r|x)$ is independent of x, and this completes the proof.

Chapter 3

Conventional LT Models Based on Perpendicular Distance

3.1 Introduction

Observed perpendicular distance data from LT experiments with $g(0) = 1$ are sufficient for the estimation of N. A rationale for basing estimation on radial distances as well as perpendicular distances is that the joint distribution of radial and perpendicular distances can in principle provide additional useful information on the shape of the detection function. But the indication from the application of models based on radial distance was that in practice the detection process varies sufficiently from application to application that it could not be modelled robustly using simple convenient assumptions about the detection process. As a result, conventional LT estimation methods switched increasingly away from detailed models of the detection process which took account of radial distances, and towards the development of robust empirical models and estimators based on perpendicular distances alone.

In the LT context, "robust" estimators are estimators which are relatively insensitive to the form of the true underlying detection process, i.e. estimators which provide nearly unbiased estimates over a wide range of true detection processes. This property is called "**model robustness**" in the LT literature. (The term was first used in the context of LT estimation by Burnham *et al.*, 1979.) Model robustness was the primary criterion on which conventional radial distance estimators failed and it was one of the primary motivations for developing estimators based on perpendicular distance.

This Chapter contains an overview of the development of models and estimators based on perpendicular distance. I do not attempt to cover every model in the literature (there are too many), nor do I go into much detail regarding methods of inference; this has already been done in other texts (Buckland *et al*, 1993a, and Burnham *et al.*, 1980). Instead I try to give an overview of the main themes in the development of the methods, using particular models and estimators for illustration.

3.2 Assumptions about the Detection Function, $g(x)$

Although conceptually the distribution of observed perpendicular distances, $f(x)$, is generated by appropriate integration of a hazard function $h_Y(y|x)$ over the forward direction, perpendicular distance models avoid having to make strong assumptions about $h_Y(y|x)$ by modelling $f(x)$ rather than $h_Y(y|x)$, $h_R(r|x)$, or $f_{X,Y}(x,y)$. To do this, some assumptions need to be made about the shape of $f(x)$. In the interests of robust estimation, the assumptions should be valid for a wide variety of detection processes.

With uniform distribution of animals in the covered area ($\pi(x) = W^{-1}$), $f(x)$ is proportional to the detection function $g(x)$ so the shapes of the two functions are identical. Assumptions about the shape of $f(x)$ are made by considering what restrictions on the shape of $g(x)$ might be reasonable in general. Burnham and Anderson (1976) proposed the following two central constraints on $g(x)$ which have become standard in the LT literature.

(1) **A Monotonicity Criterion:** $g(x)$ must be monotonically decreasing with x.

(2) **A Shape Criterion:** $\frac{d}{dx}g(x) = 0$ at $x = 0$, that is $g(x)$ must have a "shoulder" at the origin.

At this point, I should note that LT models invariably assume symmetry about the trackline, and analysis of the data is performed after "folding" data from the left hand side of the trackline over onto the right hand side (so x is really $|x|$). The integral of the detection function from $x = 0$ to $x = W$ ($\int_0^W g(x)\,dx$) is referred to as the "effective strip width" (esw) although "effective strip half-width" would be more accurate.

For almost all applications, the monotonicity criterion is reasonable. There are isolated cases in which it is not reasonable (usually because the trackline is obscured), but I do not consider these here. One example is given by Alldredge and Gates (1985), who developed a model and estimator for the case in which the trackline is obscured. Quang and Lanctot (1991) give another example, for which they developed an estimator which does not rely on the monotonicity criterion, using a truncated Beta distribution to model $g(x)$.

The shape criterion is less self-evident. Burnham and Anderson (1976) and Burnham *et al.* (1980) motivate it on the grounds that the physical setting of terrestrial LT surveys results in the probability of detecting animals within some small distance of the trackline being the same as that on the trackline and is therefore unity. Buckland (1985) argues that usually $g(x)$ could reasonably be expected to be smooth in the region of the origin as well as elsewhere and if $\frac{d}{dx}g(x) \neq 0$ this would not be the case. (Recall that the modelled $g(x)$ is only the positive half of a detection function, which is symmetric about $x = 0$.) Whatever the cause, perpendicular distance data which do not have a shoulder present difficulties for robust estimation (Buckland *et al.*, 1993a).

Despite the fact that the shape criterion above has become an almost standard part of estimation of $f(x)$, Johnson and Routledge (1985) successfully developed an estimator based on the following different

shape criterion.

(2*) **An Alternative Shape Criterion:** $g(x)$ must have a concave (i.e. $\frac{d^2}{dx^2}g(x) < 0$) shoulder, followed by a convex tail, with a single point of inflection.

3.3 The Effect of Ignoring Other Explanatory Variables

In most LT surveys the probability of detecting an animal depends on other variables in addition to perpendicular distance. (It might depend on animal/group size, environmental conditions, etc.) Nevertheless, the detection function is modelled as a function of perpendicular distance alone. How does this affect inferences about the detection function? Patil *et al.* (1993) showed that when detection of animals on the trackline is certain, then as long as the correct forms are used for the detection functions, nothing is lost by pooling data over all explanatory variables. I paraphrase their argument here for completeness of the section.

Let z be a vector random variable composed of all things other than perpendicular distance which affect the detectability of animals. Let $g(x,z)$ be the probability that an animal "at" (x,z) is detected, and let $g(0,z) = 1$. Define the following pdf's: $f(x,z)$ is the pdf of (x,z) for detected animals; $f(x)$ is the pdf of x for detected animals; $\pi(x)$ is the pdf of x in the population (i.e. of both detected and undetected animals); $\pi(z|x)$ is the conditional pdf of z in the population, given x. Finally, let $p_.$ be the mean detectability of animals in the population ($p_. = \int\int g(x,z)\,\pi(z|x)\pi(x)\,dxdz$). Then $f(x,z)$ and $f(x)$ can be written as follows.

$$f(x,z) = \frac{g(x,z)\,\pi(z|x)\pi(x)}{p_.} \tag{3.1}$$

$$f(x) = \int f(x,z)dz = \frac{E_z\left[\,g(x,z)\,|\,x\right]\,\pi(x)}{p_.} = \frac{g(x)\,\pi(x)}{p_.} \tag{3.2}$$

Here $E_z[\,|x]$ indicates the expectation with respect to z, given x; $g(x)$ is the average value of the detection function at x (where averaging is over z). The point is that the density of the observed x's when the detection function varies as a function of variables other than x, is identical to that for the case where the detection function is constant with respect to z (and equal to $g(x)$). So providing one uses the correct form for the detection function ($g(x)$), and $g(0,z) = 1$, there is no gain in modelling detection probability as a function of other explanatory variables.

In practice one never knows the correct form for the detection function and this is why it is desirable to use robust estimators of $f(x)$ and/or to develop models for the detection probability as a function of x and z. (Note that $\pi(z|x)$ will generally not be known. As a result it is not obvious how to estimate $p_.^{-1}$, given an estimate of $g(x,z)$. This question is addressed in Chapter 4.)

Burnham *et al.* (1980) coined the term "**pooling robust**" for estimators which are robust with respect to variation in $\boldsymbol{z}$. An estimator of $f(x)$ is pooling robust if the data can be pooled over $\boldsymbol{z}$ and still yield a reliable estimate of density. Burnham *et al.* (1980) contains a fuller discussion of this property. Pooling robustness is really just a type of model robustness. Buckland *et al.* (1993a) point out that while strictly only models that are linear in their parameters are pooling robust, all models which are model robust are also approximately pooling robust.

3.4 Two Approaches to Model Robustness

With a few exceptions, the development of estimators of $f(0)$ have followed one of two separate approaches. The first involves assuming a specific functional form for $g(x)$. The second models $f(x)$ using truncated convergent series.

3.4.1 Estimation *via* a Specific Functional Form for $g(x)$

The first approach involves specifying a parametric functional form for the detection function $g(x)$ (which should preferably satisfy the monotonicity and shape criteria), and then estimating the parameters of this function, usually by maximum likelihood. Maximum likelihood estimates can be obtained using either grouped or ungrouped data (also called binned and unbinned data). Buckland *et al.* (1993a) recommend estimation based on grouped data as the default method. This has the advantage of "smearing" out some of the inevitable inaccurracies in most positional data. Smearing occurs to the extent that all data within the same perpendicular distance interval/bin contribute identically to the likelihood, irrespective of the particular perpendicular distance within the bin to which they might have been assigned. (I do not discuss methods of smearing here. See Butterworth, 1982a, Hammond, 1984, Butterworth *et al.*, 1984a, Buckland and Anganuzzi, 1988 and Buckland *et al.*, 1993a for discussions and details of smearing methods.)

Maximum likelihood estimation using binned or unbinned data can be performed by maximizing one of the following likelihoods with respect to the parameters of $g(x)$. For the binned likelihood the interval $[0; W]$ is divided into K bins with I_k $(k = 1 \ldots K)$ being the set of x-values in the kth bin, and n_k $(k = 1 \ldots K)$ being the number of detected animals falling in the kth bin. The appropriate likelihood for binned data is as follows.

$$L_{Bf(x)} = \kappa \prod_{k=1}^{K} \left(\int_{I_k} f(x)\, dx \right)^{n_k} \tag{3.3}$$

(The subscript "B" is for "Binned".) Recall that

$$f(x) = \frac{g(x)\,\pi(x)}{\int_0^W g(x)\,\pi(x)\,dx} \tag{3.4}$$

and that $\pi(x) = W^{-1}$.

The binned likelihood is a component of the following multinomial likelihood of Burnham *et al.* (1980).

$$L_{Bx} = \left(\frac{N!}{n!\,(N-n)!}\right)(1-p_.)^{N-n}p_.^{n}\prod_{k=1}^{K}\left(\int_{I_k} f(x)\,dx\right)^{n_k} \tag{3.5}$$

(Here $n = \sum_k n_k$, and recall that $p_. = \int_0^W g(x)\,\pi(x)\,dx$.)

Maximum likelihood estimation from unbinned data is achieved by maximizing the following likelihood with respect to the parameters of $f(x)$ (or, equivalently, the parameters of the detection function).

$$L_{Uf(x)} = \prod_{j=1}^{n} f(x_j) \tag{3.6}$$

(The subscript "U" is for "Unbinned".) This is the second component of the Seber's (1982) likelihood given in equation 2.6 of Chapter 2.

The list of functions which have been proposed for $g(x)$ is long, and not all these have been used much in practice. Any function with an intercept of 1 which satisfies the monotonicity criterion and the shape criterion could be used as the basis for a general model for $g(x)$. In fact, some models which have been used quite widely in practice (the negative exponential, for example) do not even satisfy the shape criterion.

Experience with LT datasets and simulation studies suggest that single–parameter models for $g(x)$ are insufficiently flexible to be model robust (see Burnham *et al.*, 1980, and Buckland *et al.*, 1993a) A simulation study which serves to illustrate the point was conducted by Ramsay (1979). He used the exponential power series proposed independently by Quinn (1977) and Pollock (1978) as the most general form for $g(x)$:

$$g(x) = \exp\left\{-\left[\Gamma\left(1+\frac{1}{\gamma}\right)\frac{x}{\omega}\right]^{\gamma}\right\} \qquad (\gamma > 0) \tag{3.7}$$

($\Gamma(\,)$ denotes the gamma function and ω is the effective strip width.) When $\gamma = 1$, this reduces to the negative exponential form, while when $\gamma = 2$ it reduces to the half-normal of Quinn and Gallucci (1980).

Ramsay's (1979) results are not general in that they address only questions about the choice of particular members of the exponential power series family, but they do serve to illustrate the point. Both the negative exponential and the half–normal models are single parameter models. Ramsay (1979) found that they performed well when the assumed $g(x)$ (i.e. the assumed γ) was the same as the simulated $g(x)$ (i.e. the simulated γ). However, when the assumed $g(x)$ was not the same as the simulated $g(x)$, the performance of the generalized exponential model (for which a value for γ is estimated rather than

assumed) was better in terms of mean squared error than either of the models which assumed a fixed value for γ.

Other two-parameter forms for $g(x)$ in the literature include the exponential quadratic model of Burnham *et al.* (1980) and the hazard rate model of Hayes and Buckland (1983). The exponential quadratic models $g(x)$ is as follows.

$$g(x) = \exp\{-\beta_1 x - \beta_2 x^2\} \tag{3.8}$$

The negative exponential ($\beta_2 = 0$) and the half-normal models ($\beta_1 = 0$) are special cases of this model. Constraints need to be placed on the parameters for the function to be monotonically nonincreasing in x; both β_1 and β_2 must be greater than or equal to zero. In this case, all curves fall between the negative exponential and half-normal. The curve does not satisfy the shape criterion unless $\beta_1 = 0$, since $\frac{df}{dx} = -\beta_1$, but this criterion may often be almost met in that $\frac{df}{dx}$ may be close to zero.

The hazard rate form proposed by Hayes and Buckland (1983) is

$$g(x) = 1 - \exp\left\{-\left(\frac{x}{\sigma}\right)^b\right\} \tag{3.9}$$

Here σ is a scale parameter and b is a shape parameter. The form has an appealing property which is unique among models for $g(x)$. This is that it was derived by modelling the hazard function, and as such is not an entirely empirical model in the way that the other models are. Hayes and Buckland (1983) derived it initially by assuming specific forms for the hazard function. They noted that the form of $g(x)$ given above results from a variety of hazard functions, and suggested that it may be model robust form for $g(x)$ as a result.

Hayes and Buckland (1983) were not the first to derive forms for $g(x)$ from models for the hazard function, but they took the approach further than anyone else. Schweder (1977) developed a discrete hazard rate model (in which animals are not continuously available for detection) although he did not develop an estimator. Butterworth (1982b) derived more than one $g(x)$ on the basis of assumptions about the hazard function, but did not pursue the approach after he found that the $g(x)$'s he derived provided a relatively poor fit to the minke whale survey data he was analysing.

3.4.2 Estimation *via* Series Representations of $f(x)$

The second approach models $f(x)$ as a truncated convergent series. This approach uses the data more explicitly in determining what an appropriate form for $f(x)$ (or $g(x)$) might be. (The data are used to determine which terms of the series to include.) Three types of series for $f(x)$ have been proposed:

(1) **Simple polynomials** (Anderson and Pospahala, 1970, Anderson *et al.*, 1980, Gates and Smith, 1980)

$$f(x) = \frac{1}{w} + \sum_{m=1}^{M} \alpha_m x^{2m} \tag{3.10}$$

(2) **Fourier series** (Crain *et al.*, 1978)

$$f(x) = \frac{1}{w} + \sum_{m=1}^{M} \alpha_m \cos(\frac{m\pi x}{w}) \tag{3.11}$$

(3) **Hermite polynomials** (Buckland, 1985)

$$f(x) = \frac{\exp\left\{-\frac{x_*^2}{2}\right\}}{\sqrt{2\pi}} \left[1 + \sum_{m=1}^{M} \alpha_m H_{2m+2}(x_*)\right] \tag{3.12}$$

$H_{2m+2}(x_*)$ is the $(2m+2)$th Hermite polynomial, $x_* = \frac{x}{\sigma}$, a standardised form of perpendicular distance (see Buckland, 1985, for details).

Assumed symmetry of the detection function about zero means only the even polynomials appear in (1), only cosine terms appear in the Fourier series, and only the even Hermite polynomial terms appear in (3).

Estimation of $f(0)$ when $f(x)$ has a series representation is qualitatively different from the case in which a functional form for $g(x)$ is specified, in that it involves two steps, as follows.

(1) A **model selection step** to decide which terms of the series to include. This presupposes some rule for deciding which of two contending models (which contain different sets of terms) is the better model. If the estimation step (below) is performed by maximum likelihood and the two models are nested, a likelihood ratio test can be used. An alternative is the Akaike information criterion (AIC) of Akaike (1973). And as with any model selection procedure the issue of the order in which comparisons should be made arises (sequential/stepwise/etc.). See Buckland *et al.* (1993a) for a discussion of the options.

(2) An **estimation step** to estimate the parameters (the α_m's), given the set of m's which are to be included. Estimation can be based on either grouped or ungrouped data. The model parameters can be estimated by maximum likelihood, although other methods have also been used. If the monotonicity criterion and the shape criterion are to be satisfied, constrained maximum likelihood estimation may have to be implemented. (A truncated series representation of $f(x)$ can quite easily violate the monotonicity criterion in particular.)

In contrast, the method based on a chosen functional form for $g(x)$, involves a more subjective model selection step and the data are used only in an estimation step, which is conditional on the selected model.

Estimation of the parameters of the models can be accomplished using either grouped or ungrouped data. Crain *et al.* (1978) obtained closed form expressions for the estimators of the parameters of the Fourier series model for ungrouped data, as well as for the variance and covariances of these estimators. Buckland (1985) considered only estimation from grouped data for the Hermite polynomial model.

3.5 The Performance of the Estimators

With the variety of available models and estimators for $f(0)$, the question of how best to estimate abundance from a LT survey is to a large degree a question of model selection. One particular model for $f(x)$ may provide a better model for a particular dataset than another form for $f(x)$, but may result in model misspecification when applied to a new dataset. There are many examples of this in the literature. One is contained in the analysis of a shipboard LT survey of porpoises performed by Quinn and Gallucci (1980). The porpoise survey data exhibit a dramatic spike at the origin, which the negative exponential form is able to fit quite well. For most LT surveys, however, the negative exponential is a poor model which can result in substantial positive bias in $f(0)$ and hence in abundance estimates. The Fourier series model, which does have a shoulder, has been shown to perform well in a variety of situations (Buckland *et al.*, 1993a) and is a better candidate than the negative exponential for most LT datasets.

Clearly the more that is known about the true underlying shape of $f(x)$, the easier it is to choose a suitable functional form to use to estimate $f(0)$. However, the detection process generating observations in LT surveys is generally not very well understood, and appears to vary sufficiently from application to application to make it seldom possible to specify the most suitable model for $f(x)$ *a priori* (before seeing the data). An exception is the case in which a series of similar surveys are conducted for the same species in the same environment. In this case, it is more reasonable to choose a suitable functional form for $g(x)$ or $f(x)$ *a priori* on the basis of previous datasets, before seeing the new datset.

In the face of substantial variation in the shape of $f(x)$, model robustness is probably the most critical attribute for estimators of $f(0)$ to possess. So how do the various estimators compare on this and other criteria? The most comprehensive single investigation to date of the performance of estimators was conducted by Buckland (1985). He evaluated the performance of seven estimators of $f(0)$ on 23 sets of real data. He compared the following estimators.

(1) The exponential power series model of Quinn (1977).

(2) The exponential quadratic model (a special case of the exponential polynomial model of Burnham *et al.*, 1980).

(3) The hazard rate model of Hayes and Buckland (1983).

(4) The Fourier series model of Crain *et al.* (1978) with two terms only.

(5) The Fourier series model of Crain *et al.* (1978) with a variable number of terms.

(6) The Hermite polynomial model of Buckland (1985) with two terms only.

(7) The Hermite polynomial model of Buckland (1985) with a variable number of terms.

Each of the estimators was used to estimate $f(0)$ for each of 23 datsets, nine of which displayed a shoulder at the origin, eight a spike at the origin, and six of which appeared to have violated one or more of the assumptions of conventional LT estimators (e.g. non-uniform distribution of animals about the trackline as a result of avoidance of the trackline - evidenced in a clear spike away from the origin.) Buckland (1985) compared the estimators on the basis of four statistics, namely the estimates of $f(0)$ and their standard errors, the goodness of fit as measured by the relative magnitudes of the likelihood for each dataset, and the goodness of fit as measured by the χ^2 statistic.

On balance, the hazard rate estimator performed best over the tested datasets, although it did not perform best in every case. It was found to be both efficient and to provide the best fit on average to datasets with a shoulder, while at the same time being able to fit spiked datsets about as well as the exponential power series and exponential quadratic models (forms which allow $\frac{d}{dx}g(x) < 0$). In addition, it is the only model which provided a reasonable fit to all datasets for which there was no obvious failure of the conventional LT assumptions. The exponential power series and exponential quadratic were found to be inefficient compared to the other estimators.

The exponential quadratic, two–term Fourier series, and two–term Hermite polynomial models provided poor fits to datsets with wide shoulders - they were only able to fit large shoulders by showing an increase over some of their range. The two–term Fourier series model proved to be unable to fit peaked datasets, consistently under-estimating $f(0)$, and doing so with a small standard error.

Of the variable-term estimators, the Hermite polynomial performed better than the Fourier series model in two respects. Firstly, the estimates of variance of the former are both less dependent on the number of terms chosen (the one–term Fourier series estimator leads to a substantially smaller estimate of variance than the two–term model in general) than those obtained from the latter (which tends to under-estimate variance). Secondly, the number of terms chosen with the Hermite polynomial model is less variable for different shaped perpendicular distance distributions than for the Fourier series model. When either of these estimators is used with four or more terms, the estimated $f(x)$ tends to show an increase over part of its range.

The results of Buckland's (1985) comparisons of the estimators are confirmed in the main by a similar set of comparisons of the negative exponential, generalized exponential, hazard rate, Fourier series and Hermite polynomial estimators when used to estimate $f(0)$ from the 1978/79 to 1984/85 IWC Antarctic minke whale survey data (Buckland, 1987a).

While the hazard rate form of Hayes and Buckland (1983) is possibly the best general-purpose model for LT data, it is not the best model in every case. It is model robust, but a model which is less so may be more appropriate for a particular dataset.

3.6 A Unified Approach

Buckland (1992a, 1992b) combined the two approaches to modelling $f(x)$ (namely choosing a specific functional form for $g(x)$, and representing $f(x)$ by a truncated series) into a single theoretical framework. The framework allows the analyst to take advantage of the strengths of both approaches, and provides an environment in which to apply objective statistical model selection techniques to a greater degree in deciding on the most appropriate model for the data at hand. In every case, the analyst has access to model robust estimators, while at the same time having the ability to adjust the model, using objective model selection criteria, to accommodate the particular dataset at hand.

With this approach $f(x)$ is modelled as follows.

$$f(x) = \frac{k(x)}{\Omega}\left[1 + \sum_{m=1}^{M} \alpha_m s_m(x_*)\right] \tag{3.13}$$

- $k(x)$ is the "key" function (which can be one of the specific robust functional forms assumed for $g(x)$ above, for example).
- $s_m(x_*)$ is a series expansion term:
  - $s_m(x_*) = x_*^{2m}$ for a simple polynomial
  - $s_m(x_*) = H_{2m+2}(x_*)$ for a Hermite polynomial
  - $s_m(x_*) = \cos(m\pi x_*)$ for Fourier series

  The function $[1 + \sum_{m=1}^{M} \alpha_m s_m(x_*)]$ is referred to as the "adjustment function" below.
- x_* is simply x scaled appropriately for the series being used, so that x_* is scale invariant (see Buckland, 1992a, for details).
- Ω is a normalizing function, which is a function of the parameters only.

Model selection based on likelihood ratio tests and/or the AIC is discussed in Buckland *et al.* (1993a). Models from either of the two approaches described in the preceding sections are special cases of this general model, as illustrated below.

- Any one of the specific (non-series) functional forms given above is obtained simply by setting the key function, $k(x)$, equal to this form and all the α_m's equal to zero.
- The Fourier series model is obtained by setting $k(x) = W^{-1}$ and $s_m(x_*) = \cos(m\pi x_*)$, with $x_* = x/W$.
- The Hermite polynomial model is obtained by setting the key function equal to the half-normal model and $s_m(x_*) = H_{2m+2}(x_*)$, with $x_* = x/\sigma$ where σ is the scale parameter of the half-normal.

New models are easily constructed by combining elements from each of the two approaches. For example, one can start with a hazard rate model key function, and adjust it to improve the fit using the Fourier series adjustment function. Buckland *et al.* (1993a) discusses the possibilities in more detail.

Parameter estimation is achieved by maximum likelihood, using either the grouped or ungrouped likelihoods (of equations 3.3 and 3.6 respectively), with $f(x)$ defined by equation 3.13. Details can be found in Buckland (1992b) or Buckland *et al* (1993a). For any given {key function, adjustment function} combination, selection of the number of adjustment terms to be included can be achieved using the likelihood ratio statistic or the AIC. Selection between models with different key and/or adjustment functions can be performed using the AIC.

3.7 An Alternative Approach

All the models of either of the two approaches outlined above contain implicit assumptions about the detection process (over and above monotonicity and shape criteria). These come in the form of limitations on the shapes which the function used to estimate $f(x)$ can take. The constraints may not be obvious. Johnson and Routledge (1985) gives an example of this: the fact that the exponential power series model for $f(x)$ can exhibit a sharp decline to a broad tail only by being spiked at the origin. They recommend the use of estimators of $f(0)$ based on more explicit assumptions about the shape of $g(x)$ and propose an estimator which is based on explicit shape restrictions only. In principle their estimator can be applied with any set of shape restrictions, but the fundamental restrictions they envisage are as follows.

(1) $f(x)$ is monotonically nonincreasing with x, and

(2) $f(x)$ has a shoulder at the origin and a single point of inflection.

Other restrictions can be added, such as the range of concavity for example, or a range for the point of inflexion.

Their estimation procedure requires that the perpendicular distance data be grouped into K intervals, with frequency n_k in the kth interval ($k = 1 \ldots K$). The estimator has K parameters, η_k, which are estimated by minimising the sum of squares $\Sigma_{k=0}^{K}(n_k - \eta_k)^2$ subject to the explicit restrictions on the shape of $f(x)$. They use a least distance programming algorithm, described by Lawson and Hanson (1974, Chapter 22) to perform this minimization. $f(0)$ is then estimated as $\hat{f}(0) = \hat{\eta}_0$. They develop a Monte Carlo method of obtaining a conservative estimating confidence interval (i.e. ostensibly an interval as wide or wider than the true confidence interval) for $f(0)$.

Johnson and Routledge (1985) proposed their estimator before the unifying theory of Buckland (1992a, 1992b). The only comparative assessment of their estimator was also conducted (by them) prior to Buckland's papers appearing. As a result, no comparison of the relative performance of their estimator and the unified approach has ever been performed. They assessed the performance of their estimator of

cover each of these here. I restrict comments to some general remarks, and outline the methods developed for the unified approach of Buckland (1992a, 1992b). More detailed discussion of the methods can be found in the texts to which reference is made.

With the unified approach, the model parameters and $f(0)$ are estimated by maximum likelihood, and the variance of the estimate (or any analytic function of the model parameters) can be estimated using an approach based on the information matrix. Buckland (1993a) gives details, including examples for special cases of the unified approach which correspond to estimators from each of the two separate approaches proposed previously. From standard likelihood theory, $\hat{f}(0)$, the mle of $f(0)$, is asymptotically normally distributed and confidence intervals for $f(0)$ can be obtained under the assumption of normality and independence using the estimate of variance based on the estimated information matrix.

Two particular situations in which this information matrix based approach is not valid can occur frequently. The first is the case in which constraints are placed on the model of equation 3.13, and constrained maximum likelihood estimation is used to estimate the model parameters and $f(0)$. The second is the case in which the observed x's are not independent, as a result of a non-independent distribution of animals in the survey area for example. In both cases, the asymptotic theory on which the approach is based breaks down.

In either of these cases, methods which use replicate transects as the sampling unit can be used to estimate variance and obtain confidence intervals (see Buckland *et al.*, 1993a). Transects lengths are typically much longer than the distance over which there is correlation between the positions of animals, so that transects can usually be reasonably assumed to be independent.

Variance and confidence interval estimation by way of the bootstrap (see Buckland *et al.*, 1993a) introduces the possibility of estimating the uncertainty due to model selection. If the model selection procedure is automated, it can be applied at each iteration of the bootstrap algorithm, allowing the selected model to vary from iteration to iteration. Bootstrap variance and confidence intervals calculated in this way include variation due to model selection, something which is usually ignored in interval estimation.

3.9 Putting It All Together

The ultimate aim of a LT survey is estimation of $\aleph$, not $f(0)$. The conventional estimator is as follows.

$$\hat{\aleph} = A\frac{n}{2L}\hat{f}(0) \tag{3.14}$$

The prevalent method in the literature for estimating the variance of $\hat{\aleph}$ is based on the "Delta method" approximation, as follows.

$$cv^2[\hat{\aleph}] = cv^2[n] + cv^2[\hat{f}(0)] \tag{3.15}$$

Buckland *et al.* (1993a, pp 53-54) discuss the rationale for this approximation, including the basis for treating n and $\hat{f}(0)$ as independent in the approximation. $Var[\hat{f}(0)]$, and hence $cv^2[\hat{f}(0)]$, can be estimated using the methods described in the preceding section. $Var[n]$ and $cv^2[n]$ can be estimated by making some distributional assumption about n (that n is Poisson, for example), but are more robustly estimated nonparametrically, using transects as sampling units (see Buckland *et al.*, 1993a).

Having estimated $cv^2[\hat{N}]$, confidence intervals for N are better estimated using the lognormal based method of Burnham *et al.* (1987, p212) than by assuming that $\hat{N}$ is normally distributed. (See the simulation study reported in Buckland *et al.*, 1993a, pp254-260.)

The variance of $\hat{N}$, and confidence intervals for N, can be estimated without using the Delta method approximation by using a bootstrap procedure over the whole estimation process. That is, by resampling transects and calculating $\hat{N}$ for each pseudo-sample. A bootstrap estimate of $Var[\hat{N}]$ and confidence intervals for N can then be calculated from the list of estimates of N obtained in this way from the pseudo-samples. This method allows variance due to model selection to be estimated as an integral part of the bootstrap procedure, simply by performing model selection at each bootstrap iteration. Buckland *et al.* (1993a) discuss this and other variance and confidence interval estimation methods in more detail.

3.10 Summary: The Conventional LT Problem Solved

The principal problem facing LT estimation based on perpendicular distance after the failure of estimators based on radial distance to provide robust models for general application, was one of model robustness. In other words, without detailed knowlege of the detection process (equivalently of the hazard function), empirical models were required for $f(x)$ which were robust over a wide range of true detection processes.

Each of the two main approaches taken to obtain model robust estimators was at least partially successful. It remained the case, however, that in any particular application a model of the truncated series variety might model the observed data better than one which used a specified functional form for $f(x)$, and *vice-versa*. Further, within each of the two main approaches, one model might be a better model for the data on one occasion but not on another. With the development of a range of more or less suitable models, the problem switched from one primarily of developing robust models for $f(x)$, to one of choosing between the available models.

The truncated series approach already incorporated model selection procedures in the estimation process insofar as model selection criteria were used to decide which terms of the series to include in the model. The unified approach to estimation developed by Buckland (1992a, 1992b) provided the framework and the means by which model selection both within each of the two main approaches, and between the approaches themselves, could be incorporated into the estimation process. This development, together with the development of robust methods of interval estimation, essentially solves the problem of obtaining an adequate empirical model for any LT survey application. What is more, with the use of the bootstrap, estimates of the uncertainty due to model selection (an aspect of inference which is conventionally ignored)

can be incorporated in variance and confidence interval estimation if desired.

Having essentially solved the problem of robust estimation in conventional LT estimation, research effort has moved increasingly to addressing some of the inadequacies of conventional LT theory for many applications, i.e. to developing models and estimators for cases in which one or more of the assumptions of conventional LT theory fail. The literature includes the development of models which incorporate explanatory variables other than x, and the development of estimators of the proportion of animals on the trackline which are detected. I cover these two developments in Chapters 4 and 5, respectively.

Chapter 4

Incorporating Covariates into Conventional LT Models

4.1 Introduction

Although it has long been recognized that variables other than perpendicular distance affect the probability of detecting an animal (see Burnham *et al*, 1980, for example), relatively little attention has been given to developing models which incorporate these additional variables (denoted here by the vector $\boldsymbol{z}$). Instead, attention has been focussed on developing model robust estimators. And while it is true that when $g(0) = 1$, the density of observed perpendicular distances is identical whether or not the detection function depends on explanatory variables other than x (Patil *et al.*, 1993), there are occasions where modelling the detection probability as a function of $\boldsymbol{z}$ is useful even when model robust estimators are available.

One such occasion is when $\boldsymbol{z}$ is of interest in its own right. In this case, there is reason to develop models which incorporate $\boldsymbol{z}$ because they allow inferences to be made about the distribution of $\boldsymbol{z}$. They also allow estimation of the effect of $\boldsymbol{z}$ on the detection function, and this too may be of interest. The distribution, or at least the mean of group/school/pod/cluster size in the population is often of interest both for estimation of the number of individuals in a clustering population, and for the information it gives on the clustering behaviour of the animals concerned. (Note that I use "the population" to refer to the population in the survey region. This includes both sampled and unsampled animals.) It is usually the abundance of animals, rather than of clusters, which is of interest. But when animals cluster (as most do) the physical unit of detection is often the cluster not the animal, because when one animal in the cluster is detected, all other animals in the cluster are likely to be detected. It is by no means always the case that every animal in a cluster is detected with certainty once one animal is detected, but most authors assume that this is the case. For my purposes in this Chapter it is not crucial that it is the case.

When animals cluster, one approach to estimating individual abundance is to estimate the abundance of clusters using conventional LT models which include only x (where in this case x is the centre of a cluster), and to separately estimate mean cluster size in the population without assuming an explicit

parametric form for $g(x, z)$. Burnham *et al.* (1980) were among the first to consider this approach, and Buckland *et al.* (1993a) discuss the various methods which have been developed to date for doing this. The methods all use empirical regression models to model the change in the mean of observed cluster size as a function of x. Under the usually reasonable assumption that true mean cluster size in the population is independent of x, true mean cluster size is estimated as the model prediction of cluster size on the trackline, where $g(0) = 1$ and hence the expected mean of observed cluster sizes is equal to the expected mean of true cluster sizes.

An alternate approach is to develop models for the detection function which incorporate cluster size explicitly. It is this latter approach which is the main subject of this Chapter. In this Chapter, and in the thesis as a whole, I am interested primarily in models for incorporating explanatory variables other than perpendicular distance, rather than in estimating the distribution or mean of one or more of those explanatory variables. In the context of clustered animals, I concentrate on ways of estimating cluster abundance by modelling the detection function as a function of cluster size, rather than on estimating total animal abundance. Cluster size is treated primarily as an explanatory variable, not as a variable of interest in its own right.

4.2 Some New Notation

Before going any further I need to define some notation over and above that used thus far.

- $\pi(\boldsymbol{z})$ is the density of the additional variable(s) $\boldsymbol{z}$ (assumed to be independent of perpendicular distance before detection). $\boldsymbol{z}$ may be a scalar or a vector. If it is a vector, its R elements are denoted $z_1 ... z_R$. The jth observed value of $\boldsymbol{z}$ is written $\boldsymbol{z}_j$, and the rth element of $\boldsymbol{z}_j$ is written z_{rj}.

- $f(x, \boldsymbol{z})$ is the joint density of the **observed** perpendicular distances and additional variables.

- $f(\boldsymbol{z})$ is the marginal density of the **observed** additional explanatory variables, $\boldsymbol{z}$:
  $f(\boldsymbol{z}) = \int_0^W f(x, \boldsymbol{z}) dx$.

- $\omega(\boldsymbol{z})$ is the effective strip width as a function of the additional variable(s):
  $\omega(\boldsymbol{z}) = \int g(x, \boldsymbol{z}) \pi(x) dx$.

- ω is the overall effective strip width:
  $\omega = \int \omega(\boldsymbol{z}) \pi(\boldsymbol{z}) d\boldsymbol{z}$.

- $\boldsymbol{E}$ denotes expectation with respect to the density $\pi(\,)$. Where there is potential for ambiguity, I use subscripts on $\boldsymbol{E}$ to indicate the random variable with respect to which expectation is being taken. For example, if z_1 is the scalar "cluster size", then $\boldsymbol{E}_{z_1}[z_1]$ is the mean cluster size in the population.

$$= \int \frac{1}{\omega(\boldsymbol{z})} \left(\frac{\omega(\boldsymbol{z})}{\omega} \pi(\boldsymbol{z}) \right) d\boldsymbol{z}$$
$$= \frac{1}{\omega} = f(0) \tag{4.4}$$

If the estimator of $f(0|\boldsymbol{z})$ is $\hat{f}(0|\boldsymbol{z})$, then the average of the $\hat{f}(0|\boldsymbol{z}_j)$'s (i.e. of the $\hat{f}(0|\boldsymbol{z})$'s evaluated at the n observed $\boldsymbol{z}_j$'s) is an estimator of $f(0)$:

$$\hat{f}(0) = \frac{1}{n} \sum_{j=1}^{n} \hat{f}(0|\boldsymbol{z}_j) \tag{4.5}$$

So when estimation of $f(0)$ is performed by conditioning on $\boldsymbol{z}$, the estimation consists of the following two parts.

(1) Estimation of $f(0|\boldsymbol{z})$.

(2) Estimation of $f(0)$, given an estimator of $f(0|\boldsymbol{z})$.

4.4 Estimation by Modelling $f(\,)$ directly

The only multivariate estimator in the LT literature which models $f(x, \boldsymbol{z})$ directly is that of Chen (1996). Buckland (1992a) considered a kernel estimator for $f(0)$ for the univariate case and found its performance to be poor in comparison with other available estimators. One unsatisfactory aspect of its performance was the sensitivity of the estimate to the kernel window width used in fitting the model. Chen (1996) extended Buckland's model to include one explanatory variable other than x and estimated kernel window width from the data. In this model, the bivariate density $f(x, \boldsymbol{z})$ is modelled *via* the product of separate kernels for x and $\boldsymbol{z}$. $f(0)$ is estimated by evaluating $\int \hat{f}(0, \boldsymbol{z})\, d\boldsymbol{z}$, where $\hat{f}(0, \boldsymbol{z})$ is the kernel method estimator of $f(x, \boldsymbol{z})$ evaluated at $x = 0$. In a simulation study in which data were generated from the bivariate exponential power series model of Drummer and Mc Donald (1987), Chen (1996) found the estimator to perform as well as the Fourier series model of Quang (1991) in point estimation, and better in interval estimation. (See below for more about Quang's model.)

4.5 Estimation Conditional on the Observed Additional Variables

All estimators in the literature which do not model $f(x, \boldsymbol{z})$ directly use equation 4.5 for step (2). The estimators differ only in the way they estimate $f(0|\boldsymbol{z})$ (although this is not always obvious from the way the estimators were originally formulated). This section therefore deals only with estimation step (1), the ways of estimating $f(0|\boldsymbol{z})$.

Each of two primary types of univariate LT model based on perpendicular distance (namely those based on modelling the detection function and those based on a series expansion of $f(x)$) that appear in the

literature extend the univariate case to the multivariate case by conditioning on $\boldsymbol{z}$. In the multivariate case, it is $g(x, \boldsymbol{z})$ rather than $g(x)$, and $f(x\,|\boldsymbol{z})$ rather than $f(x)$ that are modelled.

4.5.1 Truncated Series Estimation

Quang (1991) developed an estimator for $f(0)$ based on a Fourier series representation of $f(0\,|\boldsymbol{z})$. The estimator is in fact just the Fourier series estimator of Crain *et al.* (1978) applied separately to each observed level/value of the additional variable, in order to estimate $f(0\,|\boldsymbol{z})$ at that level/value. The model is therefore able only to estimate $f(0\,|\boldsymbol{z})$ at the observed levels/values of $\boldsymbol{z}$, but this is sufficient for estimation of $f(0)$ using equation 4.5. Here $f(0\,|\boldsymbol{z}_j)$ is estimated as follows.

$$\hat{f}(0\,|\boldsymbol{z}_j) \;=\; \frac{1}{W}\left[1 \;+\; 2\sum_{m=1}^{M} \alpha_m \cos\left(\frac{\pi m x_j}{W}\right)\right] \tag{4.6}$$

The sum is over the M terms of the series included in the model (see Crain *et al.*, 1978, or Quang, 1991, for more details). Using this estimator together with equation 4.5 yields Quang's (1991) estimator of $f(0)$ (which he calls $\beta(0)$ rather than $f(0)$).

4.5.2 Estimators based on Modelling the Detection Function

Only one form of multivariate model for the detection function has been proposed in the literature. This is a bivariate model which involves the assumption that the additional variable affects only the scale parameter of the detection function. I call this sort of model a "Constant-Shape" ("CS" for short) model.

Particular varieties of CS model were proposed by Ramsay *et al.* (1987) in the context of point transects, and by Drummer and McDonald (1987) and Otto and Pollock (1990) in the context of line transects when $\boldsymbol{z}$ is a scalar, namely cluster size. (The structure of the model is unaltered if $\boldsymbol{z}$ is a vector, so I will continue to work in terms of the multivariate case in which $\boldsymbol{z}$ is a vector rather than a scalar.) CS models are based on parameterizing the scale parameter (only) of the detection function in terms of $\boldsymbol{z}$. (I call the scale parameter σ.) Where the detection function is parameterized in terms of a single scale parameter, this sort of model is the same as the corresponding univariate model, except for the function $\sigma(\boldsymbol{z})$ in place of a constant parameter σ for the univariate model. The effect of $\boldsymbol{z}$ is simply to stretch or shrink the x-axis (and hence to increase or decrease the effective strip width) in proportion to $\sigma(\boldsymbol{z})$. In this sense the additional variables, $\boldsymbol{z}$, do not affect the shape of the detection function.

The parameterizations of the scale parameter of Ramsay *et al.* (1987) and of Drummer and McDonald (1987) as functions of $\boldsymbol{z}$ have the following form. (They were not originally proposed in this form but it is a useful general form in which to put them.)

$$\sigma(\boldsymbol{z}) \;=\; \exp\left\{\beta_0 \;+\; \sum_{r=1}^{R} \beta_r t_r(z_r)\right\} \tag{4.7}$$

where z_r is the rth element of the vector z (the rth explanatory variable) and $t_r(z_r)$ is some transformation of z_r. Drummer and McDonald use $R = 1$ and $t_1(z_1) = \ln(z_1)$, while Ramsay *et al.* use $R = 1$ and $t_1(z_1) = z_1$.

This sort of parameterization provides a flexible means for incorporating additional variables into the detection function (under the assumption of constant shape). Palka (1993), for example, used this form together with the hazard rate functional form of Hayes and Buckland (1983). In this case, $g(x, z)$ can be written in generalized linear model (GLM) terms as a complementary log-log link function:

$$\begin{aligned} g(x,z) &= 1 - exp\left\{-\left(\frac{x}{\sigma(z)}\right)^{-b}\right\} \\ &= 1 - exp\left\{-\left(\frac{x}{exp\left\{\beta_0^* + \sum_{r=1}^{R}\beta_r^* t_r(z_r)\right\}}\right)^{-b}\right\} \\ &= 1 - exp\left\{-exp\{\beta_0 + \sum_{r=1}^{R}\beta_r t_r(z_r) + \beta_x \ln(x)\}\right\} \end{aligned} \tag{4.8}$$

with $\beta_r = -b\beta_r^*$ for $r = 0...R$ and $\beta_x = -b$.

Note that the detection function can't be estimated using GLM methods because only data from observed animals are available (the "successes"). To use GLM methods, one would need to know x and z for the $(N - n)$ undetected animals as well (the "failures"), and if one knew this one would not need to estimate N. Instead, the detection function can be estimated by maximum likelihood from equation 4.3. Note, however, that GLM methods can be used to estimate $g(x, z)$ without knowing N if estimation is conditional on the detections of a second (independent) observer. In this case the second observer generates n_2 (say) "trials" (animals), with observed x and z, and a "success" occurs when observer 1 detects one of these animals. An estimator of this sort is considered in Chapter 7.

4.6 Estimating $Var[\hat{f}(0)]$

The (x_j, z_j)'s $(j = 1 \ldots n)$ are by assumption identically distributed and independent (iid). Therefore, conditional on the estimated parameters of $\hat{f}(0\,|z)$, the $\hat{f}(0\,|z_j)$'s obtained from the methods of Quang (1991) and Drummer and McDonald (1987) are also iid random variables. In the case of Quang's estimator, it would be appropriate to condition on the M terms of the series to be included; in Drummer and McDonald's estimator the conditioning is on the parameters of the detection function. Now $\hat{f}(0)$ is just the average of the $\hat{f}(0\,|z_j)$'s (equation 4.5), so that an unbiased estimator of $Var[\hat{f}(0)]$, conditional on the estimated parameters of $\hat{f}(0\,|z)$, is as follows.

$$\widehat{Var}[\hat{f}(0)] = \frac{1}{n(n-1)}\sum_{j=1}^{n}(\hat{f}(0\,|z_j) - \hat{f}(0))^2 \tag{4.9}$$

It turns out that Chen's estimator of $f(0)$ has the same form as the estimators which condition on the observed z's. The form is as follows.

$$\hat{f}(0) = \frac{1}{n}\sum_{j=1}^{n}\hat{k}(x_j, z_j) \tag{4.10}$$

Here $\hat{k}(x_j, z_j)$ is a function of (x_j, z_j) with parameters estimated from the data (see Chen's equations 3.6, 4.4 and 4.6). Chen (1996) proposed the following equivalent of equation 4.9 for his estimator.

$$\widehat{Var}[\hat{f}(0)] = \frac{1}{n(n-1)}\sum_{j=1}^{n}(\hat{k}(x_j, z_j) - \hat{f}(0))^2 \tag{4.11}$$

In this case the conditioning is on the window widths, which are estimated from the data.

Quang (1991) and Chen (1996) showed that their estimators of $f(0)$ were asymptotically normal. (Chen states, without proof, that the same asymptotic normality holds unconditional on the window widths.) Both authors also proposed bootstrap estimators.

Drummer and McDonald (1987) estimate $Var[\hat{f}(0)]$ conditional on only some of the estimated parameters of the detection function in a similar way to that shown above for the other two estimators. They exploit some of the particular properties of their model to incorporate the variance due to estimation of some of the model parameters in their variance estimator. In particular, they only deal with models which satisfy the following condition.

$$\omega(z) = \omega^* \sigma(z) \tag{4.12}$$

where $\omega^* = \int g^*(x)\,dx$ and $g^*(x)$ is $g(x, z)$ with $\sigma(z) = 1$. In this case equation 4.5 becomes

$$\hat{f}(0) = \frac{1}{n}\frac{1}{\hat{\omega}^*}\sum_{j=1}^{n}\frac{1}{\hat{\sigma}(z_j)} \tag{4.13}$$

Note that any model in which z affects only the scale parameter of $g(x, z)$, and hence of $f(x, z)$, satisfies this condition. This can be seen from the following. Write $f(x, z)$ explicitly in terms of the scale parameter as $f^*(x, \sigma(z)) = f(x, z)$. Then, by definition of a scale parameter,

$$\begin{aligned} f^*(x; \sigma(z)) &= \frac{f^*(x; 1)}{\sigma(z)} \\ \Rightarrow f^*(0; \sigma(z)) = \frac{1}{\omega(z)} &= \frac{f^*(0; 1)}{\sigma(z)} \\ \Rightarrow \omega(z) &= \omega^*\sigma(z) \end{aligned} \tag{4.14}$$

where $\omega^* = f^*(0;1)^{-1}$. Note that ω^* is just the effective strip width of conventional univariate LT models, and it can be estimated using conventional univariate LT methods which ignore the z_j's, as can $Var[\hat{\omega}^*]$.

Armed with the conventional univariate LT estimates $\hat{\omega}^*$ and $Var[\hat{\omega}^*]$, Drummer and McDonald (1987) use the Delta method to obtain the following expression for the approximate variance of $\hat{f}(0)$.

$$Var\left[\hat{f}(0)\right] \approx E[f(0)]^2 \left[\frac{Var[\hat{\omega}^*]}{E[\hat{\omega}^*]^2} + \frac{Var\left[\frac{1}{n}\sum_{j=1}^{n}\hat{\sigma}(z_j)^{-1}\right]}{E\left[\frac{1}{n}\sum_{j=1}^{n}\hat{\sigma}(z_j)^{-1}\right]^2}\right] \tag{4.15}$$

This assumes that the two terms on the RHS of the equation are independent. (Drummer, in his unpublished PhD thesis, showed them to be independent for the estimators he considered.) The expectations are estimated by their observed values. Drummer and McDonald propose the following empirical estimator for $Var[\frac{1}{n}\sum_{j=1}^{n}\hat{\sigma}(z_j)^{-1}]$.

$$Var\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\hat{\sigma}(z_j)}\right] = \sum_{j=1}^{n}\frac{1}{n(n-1)}\left(\frac{1}{\hat{\sigma}(z_j)} - \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\hat{\sigma}(z_k)}\right)^2 \tag{4.16}$$

This estimator is conditional on the estimated parameters of the function $\sigma(z_j)$ (the parameters β_r, $r = 1 \ldots R$), but **not** on the estimates of the parameters of ω^* (i.e. the shape parameters of the conventional univariate LT estimator).

4.7 A Horvitz–Thompson Perspective

It is useful and informative to view the estimators which condition on z for estimation of $f(0)$ as Horvitz–Thompson–like estimators. Horvitz–Thompson estimators are formulated in terms of "inclusion probability" (the probability that a sample unit is included in the sample taken). Conditional on an animal being within the covered region, its inclusion probability in a LT context is the probability that it is observed. Furthermore, under the assumption of independent detections, all animals are included independently of one another. Now the inclusion probability (unconditional on x) for an animal which is within a distance W of the trackline and which has additional variables z associated with it is as follows.

$$p_.(z) = \frac{1}{W}\int_0^W g(x,z)\,dx = \frac{\omega(z)}{W} = \frac{1}{W f(0\,|z)} \tag{4.17}$$

If $f(0\,|z)$ were known a Horvitz–Thompson estimator of N which is not conditional on x would therefore be the following.

$$\hat{N}_H = \sum_{j=1}^{n}\frac{1}{p_.(z_j)} = W\sum_{j=1}^{n} f(0\,|z_j) \tag{4.18}$$

The subscript "H" is for "Horvitz–Thompson". This estimator is, by the properties of Horvitz–Thompson estimators, unbiased and efficient and its variance is as follows.

$$Var[\hat{N}_H] = \sum_{j=1}^{n} \frac{1 - p_.(z_j)}{p_.(z_j)} \tag{4.19}$$

An unbiased estimator of this variance (still assuming $p_.(z)$ to be known) is

$$\begin{aligned} \widehat{Var}[\hat{N}] &= \sum_{j=1}^{n} \frac{1 - p_.(z_j)}{p_.(z_j)^2} \\ &= W^2 \sum_{j=1}^{n} f(0\,|z_j)^2 - \hat{N}_H \end{aligned} \tag{4.20}$$

When $f(0\,|z)$ is estimated, the following is a Horvitz–Thompson–like estimator of N.

$$\hat{N}_H = W \sum_{j=1}^{n} \hat{f}(0\,|z_j) \tag{4.21}$$

The estimator is not neccessarily unbiased, although it will be if the $\hat{f}(0\,|z_j)$ are all unbiased. The corresponding estimator of $\aleph$ is as follows.

$$\hat{\aleph}_H = \frac{A\hat{N}_H}{2LW} = A\frac{1}{2L}\sum_{j=1}^{n} \hat{f}(0\,|z_j) \tag{4.22}$$

This is identical in form to the estimators of Drummer and McDonald (1987) and Quang (1991), which are

$$\hat{\aleph} = A\frac{n}{2L}\hat{f}(0) = A\frac{n}{2L}\left(\frac{1}{n}\sum_{j=1}^{n} \hat{f}(0\,|z_j)\right) \tag{4.23}$$

with the appropriate estimator $\hat{f}(0\,|z)$. These estimators can be seen as Horvitz–Thompson–like estimators similar to those proposed by Huggins (1989) and Alho (1990) in a mark-recapture (MR) context. Then equation 4.20 with $f(0\,|z_j)$ replaced by the appropriate estimator $\hat{f}(0\,|z_j)$ provides an alternate estimator of the variance of the estimators of N of Drummer and McDonald (1987) and Quang (1991) which is conditional on the estimated parameters, $\hat{\beta}$, of $f(0\,|z_j)$.

In the case where $f(0\,|z)$ is estimated by maximum likelihood, recognizing the Horvitz–Thompson–like nature of the estimators has the advantage of yielding an asymptotically unbiased estimator of $Var[\hat{N}_H]$ (and hence $\hat{\aleph}$) and of an asymptotic confidence interval for N (and hence $\aleph$) which is not conditional on

the estimated parameters. This is achieved by using the results of Huggins (1989) but with the likelihood 4.3 rather than Huggins's mark-recapture likelihood. Huggins showed that when the parameters $\boldsymbol{\beta}$ of the detection function $p_.(\boldsymbol{z})$ (which is equal to $(Wf(0|\boldsymbol{z}))^{-1}$ here) are estimated by maximum likelihood, then as long as $p_.(\boldsymbol{z})$ is a sufficiently smooth function of $\boldsymbol{\beta}$, the estimator $\hat{N}_H$ is asymptotically normally distributed with expectation N and a variance which in the cases under consideration in this section is as follows.

$$Var[\hat{N}_H] = \left(\sum_{j=1}^{n} \frac{1-p_.(\boldsymbol{z}_j)}{p_.(\boldsymbol{z}_j)}\right) + D(\boldsymbol{\beta})^T I(\boldsymbol{\beta})^{-1} D(\boldsymbol{\beta}) \tag{4.24}$$

Here $D(\boldsymbol{\beta})$ is the derivative with respect to the parameter vector $\boldsymbol{\beta}$ of $\hat{N}_H$, and $I(\boldsymbol{\beta})$ is the associated information matrix. This variance can be estimated as follows.

$$\widehat{Var}[\hat{N}_H] = \left(W^2 \sum_{j=1}^{n} \hat{f}(0|\boldsymbol{z}_j)^2 - \hat{N}_H\right) + D(\hat{\boldsymbol{\beta}})^T H(\hat{\boldsymbol{\beta}})^{-1} D(\hat{\boldsymbol{\beta}}) \tag{4.25}$$

Here $D(\hat{\boldsymbol{\beta}})$ is $D(\boldsymbol{\beta})$ evaluated at the mle $\hat{\boldsymbol{\beta}}$, and $H(\hat{\boldsymbol{\beta}})$ is the hessian matrix evaluated at $\hat{\boldsymbol{\beta}}$.

4.8 Summary and Discussion

The fact that in contrast to $\pi(x)$, the density $\pi(\boldsymbol{z})$ will generally be unknown presents a difficulty for extending univariate LT estimators to the multivariate case.

Chen's (1996) method avoids the difficulty by modelling $f(x, \boldsymbol{z})$ instead of $g(x, \boldsymbol{z})$. The method appears to be a useful one for estimation of $f(0)$ and abundance when detection of animals on the trackline is certain.

All the estimators for the multivariate case which model the detection function as a function of $\boldsymbol{z}$ avoid this difficultly by using the Horvitz–Thompson–like estimator of equation 4.21 to estimate N, and hence $\hat{N}$. Quang (1991) does this without assuming a functional form for $g(x, \boldsymbol{z})$ and Drummer and McDonald (1987) do it by assuming that $\boldsymbol{z}$ affects only the scale parameter (i.e. by assuming a constant shape for $g(x, \boldsymbol{z})$).

Some empirical support for the constant–shape assumption is contained in the study of Otto and Pollock (1990). They conducted an experiment to investigate whether the constant shape assumption results in adequate models for the detection function. In the experiment, different observers surveyed an area containing a known number of clusters (of differing cluster size) of cans. (In this case, as in the case of all models in the literature, $\boldsymbol{z}$ is a scalar.) Because of the replication, the probability of detection of each cluster size could be estimated independently for each cluster (by the proportion of the observers detecting the cluster). They compared the fit of exponential power series models with both, either, or

neither of the shape parameter and the scale parameter being a function of cluster size and found a model of the sort proposed by Drummer and McDonald (1987) (with the scale parameter only depending cluster size) to fit their data best.

Another interesting aspect of the experimant is that Otto and Pollock (1990) found the observed variation in detection probability to be greater than would apply were detection probability a function of x and cluster size only. This suggests that there were other variables affecting detection probability over and above those included in the model. The result is not really surprising because detection probability is likely to be a function of other variables, but it does indicate that variance estimates based on the assumption of independence between detections will often be negatively biased. In practice, it would be advisable to estimate the variance of the abundance estimator using robust methods which do not rely on strong independence assumptions. The variance of $\hat{f}(0)$ could be based on inter-transect variation when there are replicate transects, for example. (See Buckland *et al.*, 1993a, for a discussion of the available methods for the univariate case. The same methods can be used for the multivariate case, with a multivarite estimator of $f(0)$ in place of the univariate estimator.)

Another reason for not wanting to rely on the analytic estimators of $Var[\hat{f}(0)]$ proposed by Quang (1991) and Drummer and McDonald (1987) is that they are conditional on at least some of the estimated parameters $\hat{\beta}$, and as such will give negatively biased estimates of the sampling variation of $\hat{f}(0)$. Drummer and McDonald (1987) took advantage of the properties of their model to include a component due to estimation of some of the model parameters (those affecting the shape of the detection function) in their variance estimator. Quang (1991) does not include any component of variation due to the estimation of the model parameters in his model.

The Horvitz–Thompson approach and the results of Huggins (1989) provide a means for overcoming the problem of negative bias in variance estimates as a result of conditioning on estimated values of some of the parameters (under the assumption of independent detections). If detection probability can be adequately modelled as a function of a vector $\boldsymbol{z}$ (instead of the scalar z considered in the literature), then the approach provides a means for estimating the (asymptotic) unconditional variance of $\hat{f}(0)$ which incorporates the additional sources of variation (namely explanatory variables other than cluster size), and which is not conditional on the values of estimated parameters. While one could argue that the approach may not be worth pursuing too keenly when $g(0) = 1$ because robust estimators of variance are available, the situation is quite different for the case where $g(0) < 1$ and where the additional explanatory variables may affect $g(0)$. Here, neglecting the additional variables has the very undesirable effect of causing bias in point estimates of $f(0)$ and of abundance. Before considering this situation in detail (Chapter 6), I discuss the estimators of $g(0)$ in the literature (Chapter 5).

Chapter 5

Univariate "$g(0)$" Estimation Methods

5.1 Introduction

The prevalent approach in the literature to abundance estimation when detection on the trackline is not certain is to adjust conventional LT estimates of abundance by an estimated correction factor for the proportion of animals missed on the trackline. This is usually achieved as follows.

(1) Estimate $f(0)$ under the assumption that detection on the trackline is certain.

(2) Estimate the probability of detection on the trackline (call this probability G).

(3) Adjust the $f(0)$ estimate from (1) by dividing it by the estimate from (2) to estimate abundance as follows.

$$\hat{N} = A\frac{n}{2L}\left(\frac{\hat{f}(0)}{\hat{G}}\right) \tag{5.1}$$

(Note that $f(0)/G$ is the inverse of the effective strip width.)

A more integrated approach is to perform steps (1) and (2) simultaneously, i.e. to estimate the inverse of the effective strip width ($\omega^{-1} = f(0)/G$) or the inverse of mean detection probability, unconditional on x ($p^{-1} = W/\omega$), in one step. This latter approach has been used historically far less than the three-step approach, despite the fact that such integrated estimation tends to yield more robust estimates than separate estimation of $f(0)$ and G (Hiby and Hammond, 1989). The reason for the prevalence of the three-step approach is no doubt a result of the fact that conventional LT methods for step (1) are well developed and readily available, and therefore form a convenient starting point for estimation when detection on the trackline is not certain.

In this Chapter I cover step (2) of the three-step approach, i.e. the estimation of G, the detection probability on the trackline. In the following chapter I cover some integrated approaches. Here I also deal only with univariate models (i.e. models in which detection probability depends only on x). Aside

from cue-based LT models, which are not covered in this thesis, there are no multivariate models in the LT literature with uncertain detection on the trackline. (Note that I use the term "multivariate" here to refer to LT models with more than one explanatory variable. In a regression context it is conventionally used to refer to models with more than one response variable.) The mark-recapture models of Huggins (1989) and Alho (1990) are effectively models of this type but they have never been used in a LT context. The application of these methods in analyses of shore-based surveys of gray whales by Buckland *et al.* (1993b) and Laake *et al.* (1994) are closest to a LT application of this method in the literature. However, they are not really LT surveys in that x was not a significant explanatory variable and $\pi(x)$ was not assumed to be constant. I discuss multivariate models with uncertain detection on the trackline in the following Chapter.

There are two approaches in the LT literature to the estimation of G. They are as follows.

(1) Variable effort (VE) methods, which use data from surveys of the same region at two different effort levels to estimate G.

(2) Independent observer (IO) methods, which use data from simultaneous survey by independent observers to estimate G.

VE estimators were developed first and I deal with them first in what follows. Before I do that, however, a short note is provided on the estimation of G in an aerial survey context.

5.1.1 A Short Note on Aerial Survey "Visibility Bias"

The problem of estimating the proportion of animals missed on the trackline has been addressed in an aerial surveys context, where it is called the "visibility bias" problem. Caughley (1974) seems to have been the first to formally identify the problem in an aerial survey context. In many aerial surveys a relatively narrow strip is searched about the trackline and the height of the aircraft is such that the detection functions are effectively flat out to the edge of the strip. In this case aerial transect survey methods differ from strip transect or quadrat surveys only insofar as the (flat) detection function may not be unity. Most of the published methods for estimation of G (or "visibility bias") from aerial surveys are fundamentally different to those developed for LT surveys in that x is not an explanatory variable. In this case it is the height of the (flat) detection function that is estimated. This was at first attempted using Petersen mark-recapture methods (see Pollock and Kendall, 1987) with two independent observers. A removal method for the case in which only one of the observers is independent of the other was developed by Cook and Jacobson (1979). With the removal method one observer (observer 1 say) is aware of what observer 2 detects, but not *vice-versa*. Neither of these methods allowed probability of detection to vary within transects, although it was recognized from an early stage that there were at least two important sources of variation in detection probability within transects. These were cluster size (bigger clusters are more detectable) and animal availability (animals are sometimes hidden from view by vegetation,

water, etc.). Recently Evans and Bonnett (1993) developed a log-linear model which extends the removal method model to allow detection probability to depend on group size and Evans, Bonnett and McDonald (1994) extended the model to cover mark-recapture experiments as well as removal method studies. Both approaches deal with heterogeneity by stratifying with respect to the variables which affect capture probabilities.

Not all aerial surveys assume constant detection probability for all $x < W$. Hiby and Lovell (1996) provide a notable exception. They did not assume certain detection on the trackline, and used LT detection functions to model detection probability as a function of x. They developed a a more sophisticated likelihood-based LT model and estimation method than any currently in use. The model and estimator accommodate both animal movement and uncertain duplicate detection within a likelihood framework. (The final Chapter of this thesis comments further on the model and estimation procedure.)

5.2 A Little More Notation

In a depature from conventional LT notation, I use "p"'s rather than "g"'s for LT detection functions from now on. This is a notation which I find convenient for LT models with uncertain detection on the trackline because it avoids potential ambiguity in what is meant by "$g(\ )$". For example, a form for the detection function used by some authors for the case in which detection on the trackline is not certain, is $G \times g(x)$, with $g(0) = 1$ as in conventional LT models, but with $G < 1$. Using $g(\ )$ for the detection function (with intercept less than 1) is confusing. Using $g(\ )$ with a superscript or a subscript to indicate whether or not $g(0) = 1$, gets messy when other superscripts and/or subscripts have to be added. When there is more than one observer (as there is in the case of IO survey methods) and detection probability depends only on x, the detection function for the ith observer is as follows.

$$p_i(x) \;=\; G_i\, g_i(x) \tag{5.2}$$

Here $g_i(x)$ is the conventional LT detection function for observer i with $g_i(0) = 1$. The function $g_i(x)$ determines the shape of the detection function, while G_i is the probability that observer i detects an animal on the trackline.

Because of the widespread use of "$g(0)$" in the LT literature for probability of detection on the trackline, I occasionally cannot avoid use of $g(0)$ to mean this. Whenever I do this from now on I will use quotation marks: "$g(0)$".

The effective strip width for the ith observer is

$$\omega_i \;=\; G_i \int_0^W g_i(x)\,dx \;=\; G_i\,\omega_i^* \tag{5.3}$$

Here $\omega_i^* = \int_0^W g_i(x)\,dx$ is the conventional LT effective strip width, when detection on the trackline is certain.

Mean detection probability (averaged over x) is denoted by adding a "." subscript after the "i" subscript, as follows.

$$p_{i.} = \int_0^W p_i(x)\,\frac{1}{W}\,dx = \int_0^W G_i\,g_i(x)\,\frac{1}{W}\,dx = \frac{G_i\,\omega_i*}{W} = \frac{\omega_i}{W} \tag{5.4}$$

As with conventional univariate LT methods, the parameters of the detection function can be estimated from grouped or ungrouped data. The conventions used for denoting the number of animals detected in each of these cases is as follows. (The reason for the "$+$" superscript will be made clear in the next Chapter. I use it here for consistency with the notation of the following Chapters.)

In the **binned data** case:

n_{ik}^+ is the number of animals from the kth bin which were detected by observer i ($i = 1, 2$) (some of which may have been detected by the other observer as well),

n_{3k} is the number of animals from the kth bin which were detected by **both observers**, and

$n_{.k} = n_{1k}^+ + n_{2k}^+ - n_{3k}$ is the total number of animals from the kth bin which were detected by observer 1 or 2.

For both **binned and unbinned data**:

$n_{i.}^+$ is the number of animals detected by observer i ($i = 1, 2$) (some of which may have been detected by the other observer as well), and

$n_{..}$ is the number of animals which were detected by observer 1 or 2 (i.e. the total number of animals which were detected).

Whenever there is only one observer, the first subscript (i or ".") is dropped.

5.3 Variable Effort Estimators

The first estimators used for estimating G were developed for variable speed experiments (in which the effective survey effort is varied by varying the speed at which the observer moves through the survey region). The fundamental assumption behind these estimators is that animals on the trackline are missed because there are a limited number of opportunities to detect them before the survey platform passes. Increasing the number of opportunities for detection by increasing the search effort per animal (decreasing speed) increases the probability of detection. In the limit, as speed approaches zero and effort approaches saturation, the probability of detecting stationary animals on the trackline approaches unity.

Estimation is performed on the basis of a parameterization of detection probability as a function of effort. Given a suitable form for the function relating detection probability to effort, one can in principle estimate detection probability at any effort level after making observations at more than one level (provided that detection probability is not exactly proportional to effort over the effort range investigated). Practical considerations generally restrict observations to a couple of points on the effort axis. This makes the detection of model misspecification very difficult.

All variable effort experiments conducted to date have been variable speed experiments based on the use of only two speeds. Butterworth *et al.* (1982) and Cooke (1985) developed estimators of G based on a continuous hazard rate model for the detection process. Zahl (1989a) developed an estimator based on a discrete availability model. While Zahl (1989a) also uses (the inverse of) speed as the measure of effort, he does speculate on the possibility of estimation using other measures of effort.

5.3.1 The Estimators of Butterworth *et al.* and Cooke

Let $p(x|v)$ be the probability of detecting animals at perpendicular distance x when the observer is surveying at a speed v relative to the animals (and all animals are stationary). Butterworth *et al.* (1982) proposed the following general form for the detection function $p(x|v)$ on the basis of a continuous hazard rate model for the detection process.

$$p(x|v) = 1 - e^{-\frac{1}{v}H(x)} \tag{5.5}$$

Here $H(x) = \int_0^\infty h_Y(y|x)dy$, and $h_Y(y|x)$ is the hazard function for an as yet undetected animal at (x, y) when surveying at speed $v = 1$. Writing $H(0)$ as H for brevity, the probability of detecting an animal on the trackline (at $x = 0$) when surveying at the normal survey speed (v_0 say) is as follows.

$$G = p(0|v_0) = 1 - e^{-\frac{1}{v_0}H} \tag{5.6}$$

G is estimated by estimating the parameter H. The key equation relating detection probability at a speed v to that at normal survey speed, v_0, is as follows.

$$p(x|v) = 1 - (1 - p(x|v_0))^{\frac{v_0}{v}} \tag{5.7}$$

Butterworth *et al.* (1982) developed an estimator of G based on the ratio of the expected conventional LT estimated densities at speeds v and v_0. Cooke (1985) developed a similar but more efficient estimator based on the ratio of sighting rates at the two speeds. He also pointed out a contradiction in the way Butterworth *et al.* (1982) applied their estimator using negative exponential detection functions at both

speeds. This arises from the fact that equation 5.5 implies that if $p(x|v)$ has a negative exponential form at one speed, it cannot have the same form at a different speed.

Call the sighting rates at speeds v_0 and v, S_{v_0} and S_v respectively, and let $R_v = S_v/S_{v_0}$. Cooke's (1985) estimator is based on the following equality.

$$\frac{E[S_v]}{E[S_{v_0}]} = \frac{\int p(x|v)\,dx}{\int p(x|v_0)\,dx} \tag{5.8}$$

$$= \frac{\int 1 - (1 - p(x|v_0))^{\frac{v_0}{v}}\,dx}{\int p(x|v_0)\,dx} \tag{5.9}$$

H (and hence G) is estimated by using the ratio of the observed sighting rates, R_v, in place of $E[S_v]/E[S_{v_0}]$.

All applications of these variable speed (VS) estimators have been with $v \approx 0.5v_0$. In his analyses of the IWC VS experiments, Cooke (1985) assumed that $g(x)$ was negative exponential at speed v_0 (i.e. that $p(x|v_0) = Ge^{-\lambda x}$). Letting $W \to \infty$ he derived the following equation and closed form estimator $\hat{G}_c$.

$$G = 2\left(2 - \frac{E[S_v]}{E[S_{v_0}]}\right) \qquad \text{and}$$

$$\hat{G}_c = 2(2 - R_v) \tag{5.10}$$

(The "c" subscript is for "Cooke".) In general, when $v \neq 0.5v_0$, it is not possible to obtain a closed form estimator for G.

Sampling variability in sighting rates can result in a $\hat{G}_c$ outside of the feasible range of (0; 1). Cooke proposed a method which can be used when the length of transect surveyed at each speed is equal, which ensures that $\hat{G}_c$ falls within its feasible range, and Zahl (1989a) proposed an alternate method. Cooke's method is reported in Butterworth (1986).

Variance and Confidence Interval Estimation

Analytic variance estimates for $\hat{G}_c$ obtained under the assumption that sightings occur independently are likely to underestimate variance because of correlation between sightings as a result of variation of variables affecting detectability and/or a non-random distribution of whales. In a comparative study of confidence interval estimation methods for the IWC VS data, Butterworth (1986) concluded that confidence intervals based on the assumption of independence were unreliable and recommended a bootstrap (percentile method) estimator.

If the VS survey is composed of distinct "legs" (pairs of approximately coincident transects, each surveyed at one of the two survey speeds) it may be possible to use an empirical estimator of variance using legs as sampling units. However, if legs with no sightings occur with something more than a negligible

probability, $\hat{G}_c$ may be infinite for some legs and the empirical estimator will be useless. Cooke suggested a solution to this problem for situations in which the length of the two transects which are surveyed at different speeds (and which comprise a leg) are approximately equal. The method is described in Butterworth (1986). The bootstrap method recommendend by Butterworth uses legs as the sampling unit and conditions on the total effort (measured in transect length) covered at each speed.

Cooke (1985) pointed out that the form of $\hat{G}_c$ is such that $Var[\hat{G}_c]$ is substantially larger than $Var[R_v]$, so that it may not be possible to obtain R_v with sufficient precision to get acceptable precision for $\hat{G}_c$. This problem does not necessarily carry over into the estimator of abundance which uses $\hat{G}_c$, because of correlation between $\hat{G}_c$ and the conventional LT abundance estimator; this is one reason to prefer integrated estimation of ω to the three-step approach.

5.3.2 Zahl's Estimator

Zahl (1989a) derived an estimator of $p = \omega/W$, rather than G, for VS experiments which at first sight appears to be based on much less restrictive assumptions about the sighting process than the estimator for G given above. A distinguishing feature of this estimator is that it is does not require "any assumptions about the forms of the detection functions other than the mild assumption of continuity" (Zahl, 1989b, p3). The method effectively estimates the detection function independently in each of a number of perpendicular distance intervals, as shown in Appendix 5.6.1.

Zahl (1989a) derived the key equation relating detection probabilities at different effort levels to one another by discretizing the detection process into units consisting of a single sweep of the observer's eye over the area ahead. He assumed that the observer sweeps the area regularly so that she looks in the direction of any one animal approximately every δt seconds. The availability of the animal for detection is governed by the random variable $\eta(t)$, which is equal to 1 if the animal is available for detection at time t, and is 0 otherwise. It follows that for any particular realization of the availability process before the animal passes abeam, $\boldsymbol{\eta} = \{\eta(t_0), \eta(t_0 + \delta t), \ldots \eta(t_0 + k\delta t)\}$, the probability of failure to detect is

$$\bar{q}(x|\boldsymbol{\eta}) \;=\; exp\left\{\sum_{i=0}^{k} \eta(t_i)\ \ln\{q(x, y(t_i))\}\right\} \tag{5.11}$$

where $t_i = t_0 + i\delta t$, $y(t_i)$ is the animal's position in the along- trackline direction at time t_i, and $q(x, y)$ is the probability that an animal available for detection at $(x; y)$ goes undetected at that point. This is similar to the discrete hazard rate model proposed by Schweder (1977), except that here the animal is only available for detection at discrete, (roughly) equally spaced points $\{(x; y(t_i)); i = 0, \ldots, k\}$. There are no problems so far. However, in subsequently taking the expectation of $\ln\{\bar{q}(x|\boldsymbol{\eta})\}$ with respect to the availability process, Zahl implicitly assumes that the animal's availability at any time is independent of its availability at all other times. This assumption is evident in his equation on page 457, where he has the following.

$$E_\eta\left[\sum_{i=0}^{k}\eta(t_i)\ \ln\{q(x,y(t_i))\}\right] \;=\; c\sum_{i=0}^{k}\ln\{q(x,y(t_i))\} \qquad (5.12)$$

Here $c = E[\eta(t)]$ is the marginal expectation of $\eta(t)$ for any single t. This means that as $\delta t \to 0$, the availability process approaches a Poisson process. This assumption of independence is reflected in the fundamental relationship which he goes on to derive for detection probabilies at different speeds. Except for the fact that this relationship contains an approximate equality rather than an equality, it is identical to equation 5.5 which Butterworth *et al.* (1982) derived by assuming a continuous hazard rate model (which is consistent only with continuous or Poisson availability - see Appendix 5.6.2).

Zahl goes on to treat the approximate equality as an equality, so that his model is indistinguishable from the hazard rate model of Butterworth *et al.* (1982). What of the estimator? With some new notation and rearrangement, his estimator ($\hat{p}_z$; with subscript "z" for "Zahl") is seen to be simply a version of an estimator of p which can be obtained from equation 5.9. The only difference is that $\hat{p}_z$ involves almost no assumptions about the shape of the detection function. In particular, it does not neccessarily satisfy the shape and monotonicity criteria normally applied to LT detection functions. (Details are given in Appendix 5.6.1.)

As is the case with Cooke's estimator, when sample size is small there is a non-negligible probability that the estimate of G may fall outside the feasible range for G because fewer animals may be seen at the lesser speed v than at v_0. Zahl (1989a) deals with this by placing upper and lower bounds on allowable R_v. When the observed R_v is outside of the bounds, the closer bound is used in its place in the estimator. (For robust estimation he recommends the use of a lower bound of 1 and an upper bound of 1.8 when $v = 0.5v_0$.)

Variance and Confidence Interval Estimation

Zahl (1989a) proposed a (percentile method) two-stage parametric bootstrap method for confidence interval estimation for $\hat{p}_z$. In the first stage, the numbers seen by each platform are drawn from binomial distributions, and in the second stage the numbers of sightings within each of the K perpendicular distance intervals are drawn from multinomial distributions, conditional on the numbers obtained in the first stage. This method incorporates a strong independence assumption which may not hold in practice. A more robust procedure would be to use the transect legs as sampling units, possibly conditioning on the total effort (i.e. transect length) at each speed.

5.3.3 A Robust (but biased) Estimator

Polacheck (*pers. commn* to the author, and historically Horwood, *pers. commn* to Butterworth) suggested the following estimator of G, which does not assume any specific form for the dependence of detection probability on effort. (The subscript "HP" for "Horwood/Polacheck".)

$$\hat{G}_{HP} = \frac{\hat{D}(v_0)}{\hat{D}(v)} \tag{5.13}$$

where $\hat{D}(v_0)$ and $\hat{D}(v)$ are conventional LT density estimators (i.e. assuming $G = 1$) obtained from data gathered at speeds v_0 and v respectively (and $v_0 > v$).

The fact that G at speed v may be less than unity will cause positive bias, but the estimator may be useful for placing an estimated upper bound on G.

Variance and Confidence Interval Estimation

The Delta Method can be used to get an estimate of the variance of the estimator, with the variance of the component density estimators obtained by whatever conventional LT methods are appropriate.

5.3.4 The Performance of VS Estimators

Aside from a controlled experiment conducted by Zahl (1989a), the only applications of VS methods are to the 1980/81 and the 1984/85 IWC Antarctic minke whale survey data. Both these surveys used platform speeds such that $v \approx 0.5v_0$.

IWC VS Experiments

The 1980/81 experiment was designed so that the same vessel surveyed both the slow and the fast legs of each transect. The two legs were separated by a perpendicular distance roughly five times the truncation distance W. Two vessels conducted VS experiments in different parts of the ocean. Butterworth (1986) obtained estimates $\hat{G}_c$ from the experiment, after correcting errors in the methods and data used in two earlier analysis (Butterworth *et al.*, 1982, and Cooke, 1985). A negative exponential form was assumed for $p(x|v_0)$, and both vessels were assumed to have the same G. G was estimated to be 1.74 with a 95% confidence interval (CI) (-0.14; 2.18). A confidence region this wide is clearly useless since the feasible range of G is (0; 1). Zahl (1989b) obtained estimates (CI's in brackets) p_z of 0.48 (0.0; 0.96) and 0.81 (-0.28; 1.90) for the two vessels. His estimates may be biased because R_v hit the imposed constraint boundary of 1.8. With *cv*'s close to 100%, the estimates are also too imprecise to be of practical use.

The primary reason for the poor precision of the estimators is thought to be spatial variation in animal density between the two transects, and the next VS experiments were designed to control this problem.

The 1984/85 VS experiments attempted to reduce spatial variation in density by having two vessels survey the same area of ocean at two different speeds, as close in time to one another as was feasible without the one vessel affecting the searching on the other vessel (maximum of 30 minutes separation). Butterworth and McQuaid (1985) obtained a point estimate $\hat{G}_c = -0.34$, with 95% CI (-2.11; 1.06). $\hat{G}_{HP}$ with the hazard–rate form of Hayes and Buckland (1983) is 0.71 for these data, with 95% CI (0.19;

1.23). The confidence intervals are again too wide for the estimator of G to be of practical use - the guilty party on this occasion is variation in density between the time the first and the second vessel surveyed the area. The new design seemed to have substituted temporal variation for spatial variation.

Zahl's Simulation Study

The results of simulation experiments conducted by Zahl (1989a) to investigate the performance of his estimator indicate that the estimator has unacceptably high variance in most situations. When $v_0 = 2v$ and with sample sizes of less than about 150 (which is a respectable sample size for many line transect surveys, and is bigger than that obtained in any of the IWC VS experiments), the cv of the estimator is unacceptably large (between 42% and 328%) except when the true detection function has a very wide shoulder and G is relatively high (0.75); in this one case the cv was 12%. The simulations suggest that the estimator has too high a variance to be of much practical use even under the ideal conditions of a simulation study in which all model assumptions are known to hold, unless the detection function has a very broad shoulder and/or G is at least 0.75 and a sample size in excess of about 200 is obtained. Zahl (1989a) also found that the estimator may be substantially biased when the ratio of expected sighting rates is close to the imposed upper bound of 1.8.

Further Problems with the Method

With the exception of the biased estimator $\hat{G}_{HP}$, all the VS estimators rely on the assumption that an animal's availability for detection at any time is independent of its availability at all previous times. When this is not the case, the estimators may be substantially biased. Cooke (1985) considered the effects on $\hat{G}_c$ of a monotonic increase and of a monotonic decrease in the probability of the animal becoming available for detection, as a function of the time since its last availability, and showed that the former leads to negative bias and the latter to positive bias.

In the case of Antarctic minke whales, animals are not continuously available for detection and their surfacing patterns appear to be more complicated than the types considered by Cooke (1985). Minke whales tend to follow a surfacing pattern which consists of periods of frequent surfacings followed by one or more longer dives (Joyce, 1982, Ward, 1988). The Poisson availability assumption therefore appears to be violated for these animals, but the effect of this violation on the estimators has not been investigated in any detail.

Another cause for serious concern about the utility of VS estimators was raised by Hiby (1986). He showed that random whale movement at a speed of 3 knots (a reasonable guesstimate of average swimming speed of minke whales during the surveys) generated a large bias in sighting rate (as a measure related to density) of an observer travelling at 6 knots (the lower speed, v, for the IWC VS experiments). While this bias could be corrected if the true average speed of animal movement was known, he doubted that this could be estimated reliably.

The unacceptably wide confidence intervals for estimates of G and p alone are sufficient reason to look for other estimation methods. This, together with concerns about the validity of the Poisson availability assumption and the effect of animal movement at low speeds led research effort to be directed increasingly towards IO surveys rather than VS surveys as a way of estimating G and/or p.

5.4 Independent Observer Estimators

The idea behind IO estimation methods is that the proportion of target animals seen by observer 1 at a given perpendicular distance (x), which were also seen by observer 2, can be used to estimate the probability of observer 2 seeing a target animal at that perpendicular distance (or *vice-versa*).

In its simplest form, when the detection probabilities of each observer are constant over the perpendicular distance interval used in estimation, the method is a sort of two-sample mark-recapture method with equal catchability of all animals at each capture occasion, and different capture probability between capture occasions. Sighting by one observer constitutes marking and sighting of a "marked" animal by the other observer constitutes recapture. This corresponds to a version of the mark-recapture model type M_t of Otis *et al.* (1978) in which "time" (subscript t) is indexed by the observer. (LT estimators for G and $Var[G)]$ in this case are given in Buckland, 1987b.) When sighting probabilities vary with perpendicular distance from the trackline (only), the method is a two-sample mark-recapture method with heterogeneous capture probabilities which are determined by the perpendicular distance (x). In this case the model is a version of the "M_{th}" model of Otis *et al.* (1978) in which "time" is indexed by observer and heterogeneity (subscript h) is parameterized in terms of x. This situation (or a more complicated one in which detection probability depends on perpendicular distance **and** other variables) generally applies for line-transect surveys, where detection probability is modelled as a non-increasing function of perpendicular distance from the trackline.

Two varieties of IO models have been developed. The first is designed for situations in which the target animals are continuously available for detection. This type of model is a fairly natural extension of conventional perpendicular distance based LT models insofar as the probability of detection as a function of radial distance need not be considered explicitly in modelling and estimation. The second type of IO model caters specifically for situations in which animals are available to be detected at discrete points in space/time only. These models are extensions of the early conventional LT models based on radial distance, insofar as they generally require detection probability to be modelled as a function of both perpendicular and radial distances. (With discrete availability models one needs to model the probability of detecting an as yet undetected animal, given that it presents itself for detection at a position (r, x).) Aside from a brief mention below, I do not consider discrete availability models in this thesis.

Both sorts of IO model involve two teams of observers surveying the area of interest simultaneously, on platforms which may be separated by some fixed perpendicular distance (Δx - which is small compared to the maximum sighting distance). The observers search in isolation from and independently of one

another. The so-called parallel ship (PS) survey method is a particular form of IO survey in which the observers are separated by a distance $\Delta x > 0$. Aside from this difference PS surveys are the same as other forms of IO survey , and the associated estimators for G are considered together here. For simplicity I deal here only with the case where $\Delta x = 0$. See Buckland (1987b) or Buckland *et al.* (1993a) for details of the case where $\Delta x > 0$.

Three principal methods of estimating G from univariate (x only) IO survey data have been proposed for the continuous availability situation. They are the "Product" method, the "Direct" method and the "Buckland and Turnock" method. Two other estimators have been proposed, which I do not consider in this thesis. Butterworth (1991) proposed $\hat{D}_1/\hat{D}_{all}$ as a positively biased estimator of G, where $\hat{D}_1$ is the conventional LT estimator of density from observer 1 data only and $\hat{D}_{all}$ is the conventional LT estimator of density from all observers. Barlow (1993) proposed a removal method which is based on the assumption that both observers have equal G's. I deal with each of the principal methods below, but first provide a very brief note on discrete availability models.

5.4.1 Discrete Availability Models

Most of the research effort associated with estimating the probability of detection on the trackline has been applied in a marine context, and relates to surveys of cetaceans (whales, dolphins and porpoises) in particular. Cetaceans, and most other marine mammals, are available for detection only at discrete points in time and space, when they blow or surface. In this case some animals are more detectable than others because they become available more frequently or at different times and positions. This is a form of animal heterogeneity which results in correlation between detections by independent observers (see below).

Schweder (1974, 1977) developed LT models for such discrete availability situations, although he did not develop estimators of detection probability or abundance at that time. He did later develop an estimator of abundance and G based on a discrete hazard rate model of the sightings process (Schweder, 1990). The method involves estimation of a hazard function from IO survey in which cues rather than animals are the experimental units. Estimation of the hazard function is in the first instance based on a binary regression in which cues detected by one observer from an animal which has not yet been detected by the other observer are "trials", and the outcome is a "success" if the other observer detects the same cue. Sophisticated estimation methods have been developed which use the positions of first detections as well as these cue-based trials in estimating the hazard function (see Schweder, 1990, Schweder *et al.*, 1991, Schweder and Host, 1992 and Cooke, 1995).

Having estimated the hazard function, abundance and/or G and/or p are estimated by integrating the function over the along-trackline direction using an estimate of the availability process. Schweder (1990) and Schweder and Host (1992) proposed and implemented a computationally intensive simulation procedure to do this. The procedure requires estimates of the availability process. Data for this estimation are not available from LT surveys, and must be collected by other means. In the case of North Atlantic minke

whales, this has been done using radio transmitters attached to individual whales to provide information on their dive patterns (Folklow and Blix, 1993).

I do not consider discrete hazard rate models any further in this thesis.

5.4.2 The "Product" Method Estimator

(I abbreviate "Product method" to "P method".) A special case of this sort of estimator (assuming $g_1(x) = g_2(x)$ and assuming a negative exponential functional form for both $g_1(x)$) and $g_2(x)$ was proposed by Butterworth, Best and Hembree (1984). Kishino, Kasamatsu and Miyashita (1986) extended the theory to allow for $g_1(x) \neq g_2(x)$ while retaining the assumption that $g_i(x)$'s are negative exponential, and Buckland (1987b) generalised the estimator to allow any functional forms to be used for $g_1(x)$ and $g_2(x)$. Butterworth and Borchers (1988) refer to the method as the "PGHR" method when it is used in conjunction with the hazard rate model of Hayes and Buckland (1983) ("PGHR" for "**P**roduct of $\boldsymbol{g}(\,)$'s with **H**azard **R**ate). The estimator can be derived for the general case as follows.

Assuming that animals are uniformly distributed with respect to the trackline, the probability of observer i $(i = 1, 2)$ detecting an animal over a range of perpendicular distances from 0 to W is as follows.

$$\begin{aligned} p_{i.} &= \frac{1}{W} G_i \int_0^W g_i(x)\, dx \\ &= \frac{1}{W} G_i \omega_i^* \end{aligned} \tag{5.14}$$

Assuming that detections of objects by each platform are independent events, the probability that both observers detect the animal is as follows.

$$\begin{aligned} p_{3.} &= \frac{1}{W} G_1 G_2 \int_0^W g_1(x)\, g_2(x)\, dx \\ &= \frac{1}{W} G_1 G_2 \omega_3^* \end{aligned} \tag{5.15}$$

Now (n_3/n_2^+) is a consistent estimator of $(p_{3.}/p_{2.}) = G_1\, \omega_3^*/\omega_2^*$. So the following is a consistent estimator of G_1.

$$\hat{G}_1 = \frac{n_3}{n_2^+} \frac{\hat{\omega_2^*}}{\hat{\omega_3^*}} \tag{5.16}$$

Here $\hat{\omega_2^*}$ and $\hat{\omega_3^*}$ are estimates of ω_2^* and ω_3^* , respectively. The function $g_2(x)$ and ω_2^* can be estimated by applying conventional line transect methods to the n_2^+ perpendicular distances observed by observer 2. The parameters of $g_1(x)$ and ω_1^* can be estimated similarly from observer 1's perpendicular distances. Then ω_3^* can be estimated by evaluating the integral of the product of the estimated detection functions $\hat{g}_1(x)$ and $\hat{g}_2(x)$ (hence the word "Product" in the name of the method).

$$\hat{\omega}_3^* = \int_0^W \hat{g}_1(x)\,\hat{g}_2(x)\,dx \tag{5.17}$$

An alternative procedure is to estimate ω_3^* from the perpendicular distances of duplicate detections (those seen by both platforms) by fitting a detection function ($g_3(x)$, with $g_3(0) = 1$) to these data using conventional LT methods, and then integrate the resulting estimate $\hat{g}_3(x)$ over x. Palka (1993) suggested this method. She calls it the "Direct" method. To avoid confusion with the "Direct" method described below, I refer to it as the "Modified Product" method (MP method). When sample sizes are small it may not be possible to estimate $g_3(x)$ reliably in this way because of the scarcity of duplicate detections. The P method and MP method estimators of G_1 can be written in terms of the probability densities of observed perpendicular distances as follows.

$$\hat{G}_1 = \frac{n_3}{n_2^+}\frac{\hat{f}_3(0)}{\hat{f}_2(0)} \tag{5.18}$$

Here $\hat{f}_2(0)$ is a conventional LT estimator of the intercept of the density of perpendicular distances of observer 2 sightings, and $\hat{f}_3(0)$ is an estimator of the intercept of the density of perpendicular distances of duplicate sightings. The difference between the P method and the MP method estimator is simply a difference in the method of obtaining the estimate $\hat{f}_3(0)$. The MP method uses a conventional LT estimator applied to perpendicular distances of duplicate sightings. The P method estimator uses the inverse of $\int_0^W \hat{g}_1(x)\,\hat{g}_2(x)\,dx$, where $\hat{g}_i(x)$ are conventional LT estimators of $g_i(x)$ $(i = 1, 2)$.

G_2 can be estimated similarly, but for brevity I deal only with the case of G_1. It may or may not be possible to obtain an expression for the variance of $\hat{G}_1$ in closed form, depending on the functional forms chosen for the detection functions. In the case where both $g_1(x)$ and $g_2(x)$ are negative exponential functions with parameters λ_1 and λ_2 respectively, $\hat{G}_1$ and $\hat{G}_2$ can be expressed relatively simply, but obtaining even an approximate expression for variance is messy. In the simplest case, when $\lambda_1 = \lambda_2 = \lambda$, the algebra is more tractable. Butterworth and Hiby (1984) derived estimators of G and $cv[G]$, for this case.

In the more general case, and in particular the case where $g(x)$ has the hazard–rate form of Hayes and Buckland (1983), the variances of $\hat{G}_1$ and $\hat{G}_2$ can't be expressed simply, but can be estimated using a resampling method. Buckland (1987b), in his analysis of the 1984/85 IWC PS experiments, used a bootstrap algorithm, conditioned on $\hat{\omega}_1^*$, $\hat{\omega}_2^*$ and the total number of detections. Butterworth and Borchers (1988) used a jacknife estimator of variance because it is less computationally expensive than bootstrapping, but they note that this estimator may be positively biased if there is substantial variation in the number of sightings per sampling unit (i.e. per transect). In this case the jacknife estimator includes a component of variance due to variable sample size, wheras it may be more appropriate to condition the variance estimate on the achieved sample size as in the bootstrap method of Buckland (1987b).

5.4.3 The "Direct" Method Estimator

(I abbreviate "Direct method" to "D method".) The method was proposed by Butterworth and Borchers (1988). For estimation of G_1, the method conditions on observer 2's detections. Subject to constraints on the form of $g_1(x)$, the proportion of observer 2's detections in each of a number of perpendicular distance intervals which are detected by observer 1, yields an estimate of $p_1(x) = G_1 g_1(x)$ over these intervals.

More formally, let the domain of x $(0 \leq x \leq W)$ be divided into K mutually exclusive intervals, the kth of which is denoted I_k, where $I_k = \{x : c_{k-1} \leq x < c_k\}$, $(k = 1, \ldots, K)$ and the c_k's are the cutpoints between intervals. Then n_{3k} and n_{2k}^+ are the number of duplicate sightings and the number of sightings by observer 2 in the kth interval, respectively. Finally, let $g_3(x) = g_1(x)\, g_2(x)$ and $G_3 = G_1 G_2$. Then the probability that observer 1 sees an animal which has been seen by observer 2 in the kth interval is as follows.

$$p_{1k} = \int_{I_k} G_1\, g_1(x)\, \pi_k\, dx \tag{5.19}$$

Here π_k is the pdf of the perpendicular distances of animals, given that they are in the kth interval. Under the assumption that animals are distributed uniformly with respect to the trackline, π_k is as follows.

$$\pi_k = \frac{1}{(c_k - c_{k-1})} \tag{5.20}$$

Given the numbers detected by observer 2 in each of the K intervals $(n_{2k}^+;\ k = 1, \ldots, K)$, Butterworth and Borchers (1988) and Hiby and Hammond (1989) obtain the Direct estimator of G_1 by maximising the following multinomial likelihood with respect to G_1 and the parameters of $g_1(x)$.

$$L_{Dir} = \prod_{k=1}^{K} \left(\frac{n_{2k}^+!}{n_{3k}!\,(n_{2k}^+ - n_{3k})!} \right) p_{1k}^{n_{3k}}\, (1 - p_{1k})^{(n_{2k}^+ - n_{3k})} \tag{5.21}$$

In fact it is only under special circumstances that this likelihood is the conditional likelihood of n_{3k} $(k = 1, \ldots, K)$ given n_{2k}^+ $(k = 1, \ldots, K)$, although unless intervals are very wide the likelihood would probably be close enough to the correct likelihood in most cases for the difference to be negligible. Consider the conditional likelihood of observer 1 detecting n_{3k} animals in bin k, given that n_{2k}^+ animals were detected by observer 2 in this bin.

$$p_{1k\,|\,\text{seen by 2}} = \frac{\int_{I_k} G_1 g_1(x)\, G_2 g_2(x)\, dx}{\int_{I_k} G_2 g_2(x)\, dx} \tag{5.22}$$

This is equal to p_{1k} only in special circumstances, one of which is when $g_2(x)$ is constant in the kth interval. The correct conditional likelihood for the general case is as follows. (I abbreviate "$p_{1k\,|\,\text{seen by 2}}$" to "$p_{1k\,|\,2}$" and $n_{31} \ldots n_{3K}$ and $n_{21}^+ \ldots n_{2K}^+$ to $\{n_{3k}\}$ and $\{n_{2k}^+\}$, respectively.)

$$L\{p_{1k|2}|\{n_{3k}\},\{n_{2k}^+\}\} = \prod_{k=1}^{K}\left(\frac{n_{2k}^+!}{n_{3k}!\,(n_{2k}^+-n_{3k})!}\right)p_{1k|2}^{n_{3k}}(1-p_{1k|2})^{(n_{2k}^+-n_{3k})} \quad (5.23)$$

Maximising the likelihood L_{Dir} rather than $L\{p_{1k|2}|\{n_{3k}\},\{n_{2k}^+\}\}$ to estimate G_1 will generally result in a positively biased estimate of G_1. Maximising L_{Dir} effectively fits a function to the ratios n_{3k}/n_{2k}^+. Because x varies within each interval, the detection probabilities for the two observers are positively correlated in each interval when both detection functions are monotonically decreasing so that n_{3k}/n_{2k}^+ is a positively biased estimator of the probability of observer 1 detecting an animal in bin k. Except for the effect of possible model misspecification in the chosen form for the detection function, which may constrain the fit, this will result in the fitted detection probability function being positively biased within each interval; in particular this will be the case at the origin so that the G_1 estimate will be positively biased. If the detection functions are relatively flat within the intervals this bias will however be small. When few intervals are used (small K) the detection function will change more over the range of each interval (compared to the case when many intervals are used) and the resulting G_1 estimate may be substantially biased.

A consequence of using the correct likelihood is that $g_1(x)$ and $g_2(x)$ have to be estimated simultaneously (or estimation of G_1 has to be effected conditional on an estimate of $g_2(x)$) since the likelihood is a function of both $g_1(x)$ and $g_2(x)$. Thus maximization of $L\{p_{1k|2}|\{n_{3k}\},\{n_{2k}^+\}\}$ is computationally a more onerous task than maximization of L_{Dir}.

Asymptotic variance estimates for G_1 can in principle be obtained using the empirical information matrix (but only if the correct likelihood, $L\{p_{1k|2}|\{n_{3k}\},\{n_{2k}^+\}\}$, is used); however for the same reasons as given above for the P method estimator, a bootstrap procedure like that of Buckland (1987b) is preferable.

5.4.4 The Method of Buckland and Turnock

(I abbreviate "Buckand and Turnock method" to "BT method".) The method was developed by Buckland and Turnock (1992) in the context of shipboard LT surveys of Dalls porpoise. These are fast moving small cetaceans which are known to react to the presence of the vessel. To apply the method one observer (observer 2 say) is used to confirm the presence of animals which are potentially detectable by observer 1, before they might react to the presence of the vessel. The estimator was developed as one for abundance rather than for G_1, but when animals are stationary it incorporates an estimator of G_1 of the form given below. When animals are not stationary, estimation of G_1 cannot be separated from estimation of a correction factor for animal movement. In this case $\hat{G}_1$ is an estimator of a correction factor for both movement and animals missed on the trackline. (See Buckland and Turnock, 1992, or Buckland *et al.*, 1993a, for a more detailed discussion of this point.) The BT method estimator is as follows.

$$\hat{G}_1 = \frac{n_3}{n_2^+}\frac{\hat{f}_3(0)}{\hat{f}_2(0)} \quad (5.24)$$

Here $\hat{f}_3(0)$ and $\hat{f}_2(0)$ are the estimated densities of observed perpendicular distances **as recorded by observer 2** of duplicate sightings and observer 2 sightings, respectively, evaluated at $x = 0$. The estimated density $\hat{f}_3(x)$ is obtained from the duplicate sightings data by conventional LT methods, where the x's are the perpendicular distance **as recorded by observer 2** of duplicate sightings. The estimated density $\hat{f}_2(x)$ is obtained by conventional LT methods, where x is the perpendicular distance **as recorded by observer 2** of observer 2's sightings.

When animals are stationary, the BT method estimator of G_1 is identical to the MP method estimator (compare equation 5.24 to equation 5.18). The BT method abundance estimator is, however, different from the MP method estimator even when animals are stationary because the BT method estimator of $f_1(x)$ is not a conventional LT estimator, but is instead estimated as follows.

$$\hat{f}_1(0) = \frac{\hat{G}_1}{\int_0^W \hat{p}_1(x)\,dx} \tag{5.25}$$

where

$$\hat{p}_1(x) = \frac{n_3}{n_2^+}\frac{\hat{f}_3(x)}{\hat{f}_2(x)} \tag{5.26}$$

and all x's are as recorded by observer 2. If animals were stationary, this conditioning on the perependicular distances of observer 2 in this way would be unnecessary. When animals move substantially, particularly when their movement is in response to the presence of observer 1, the conditioning is necessary. This is because $\pi(x)$ can be assumed to be uniform before they move (when x is that recorded by observer 2) but not after they may have moved in response to the presence of observer 1 (when x is that recorded by observer 1). Buckland and Turnock (1992) and Buckland *et al.* (1993a) give more details of the assumptions underlying the method and the requirements for its implementation, as well as details of the bootstrap method used to estimate variance and confidence intervals.

5.4.5 The Performance of IO Estimators

Parallel Ship Surveys

The only implementations of PS survey, aside from the experiments conducted for estimation of the hazard function for the discrete availability model of Schweder (1990), are four sets of PS experiments which were conducted on the IWC Antarctic surveys. There are substantial differences in the methods used for these experiments which, makes comparisons of the results from the experiments difficult. The two main differences are in the methods for identifying duplicates (animals seen by both platforms) and the methods used to estimate angles and distances to sightings. Only for the final set of experiments (conducted in the austral summer of 1984/85) was a duplicate identification method developed which was considered adequate for use on subsequent IO surveys. Analysis of the earlier data is further complicated by the

fact that the perpendicular distance distributions of detections from these experiments are spiked at the origin, which makes reliable estimation of G difficult. The problem was rectified on the 1984/85 survey with the introduction of mechanical aids for accurate estimation of angles and distances (angleboards and reticule binoculars).

Butterworth *et al.* (1984b) and Butterworth and McQuaide (1985) estimated G_1 and G_2 from each of the PS experiments, assuming $p_1(x) = p_2(x) = Ge^{-\lambda x}$. Kishino *et al.* (1986) estimated G_1 and G_2 from these data assuming different negative exponentials for each vessel. The estimator of $cv[g(0)]$ proposed by Butterworth and Hiby (1984) was used to estimate cv's. Borchers (1991) contains a summary of the results from each experiment. Because duplicate identification was far from error-free, three sets of duplicate classifications were calculated, namely "Best" (corresponding to what was considered the best estimate of the number of duplicates), "Lower" (corresponding to what was considered to be a reasonable lower bound on the number of duplicates) and "Upper" (corresponding to what was considered to be a reasonable upper bound on the number of duplicates). G's were estimated using each of these classifications. Buckland (1987b) estimated G_1 and G_2 for each of the three 1984/85 PS experiments, using the hazard–rate functional form of Hayes and Buckland (1983) and the "Best" duplicate classification. (The hazard–rate model provides a better fit to the 1984/85 data and yields more feasible estimates of G_i than the negative exponential model.)

While the point estimates of G from the PS experiments are far more reasonable than those from the VS experiments (only one of the 12 G estimates for individual vessels is outside the feasible range), and the estimated cv's are quite acceptable (between 11% and 20%), there remained some serious problems with the method. Primary among these were the following.

Uncertain Duplicate Identification: The proportion of inter–platform pairs of sightings whose duplicate/non-duplicate status was not considered to be certain was high (between 24% and 71%). This uncertainty in duplicate identification might result in biased estimation and would definitely add substantially to the estimated variance of the estimates of G's were it to be included. To get some idea of the magnitude of the component of variance due to uncertain duplicate identification I compared the size of the difference between the "Upper" and "Lower" estimates of G to the length of the estimated 95% confidence interval (calculated under the assumption of normality and ignoring the uncertainty in duplicate identification). On average across all experiments the difference between the "Upper" and "Lower" estimates of G is almost exactly equal to the length of the estimated confidence interval. (It varies from 51% to 180% of the confidence interval length.) One might therefore expect uncertainty in duplicate estimation to contribute something in the region of the reported estimated uncertainty in $\hat{G}$. This uncertainty would substantially inflate the estimated variance of $\hat{G}$ were it to be included in such calculations.

Asymmetries: PS theory assumes that observers behave as if the observers on the other platform were not there and that animals behave as if neither observer was there, at least until they are

detected by both observers or pass abeam. Asymmetries of various sorts in the data indicate failure of these assumptions. Asymmetry was found between the perpendicular distance distributions of animals detected between the vessels' tracklines compared to those outside the tracklines (Butterworth *et al.*, 1982, Butterworth and Rickett, 1986) and in observer effort distributions between and outside the tracklines (Butterworth and Rickett, 1986, Ward *et al.*, 1986). Differences were also found in the mean perpendicular distance of observed animals between the PS experiments and the rest of the survey, raising doubts about the applicability of PS survey estimates of G to the rest of the survey (Butterworth *et al.*, 1982, Butterworth and Rickett, 1986). None of the differences were significant, but the tests are not powerful and the asymmetries remained cause for concern.

Heterogeneity and Possible Model Misspecification: Buckland (1987b) compared the predicted (under the assumption of independence) and observed perpendicular distance distributions of duplicates for the 1984/85 PS set of experiments and found significant differences in two of the three experiments. One possible cause for this discrepency is a lack of independence between sightings by the two platforms as a result of heterogeneous detection probabilities between animals. This and other possible causes are discussed in more detail in the following subsection.

The disappointing results from PS experiments led to them being abandoned in favour of IO experiments in which observers are located on the same platform, but are isolated from one another. This design reduced the size of the duplicate uncertainty problem and effectively solved the problem with asymmetries. (The differences between experiment and other survey could be eliminated by conducting the whole survey as an IO survey, something which is logistically hardly possible for PS surveys.)

IO Surveys

The largest and longest series of IO data for estimation of G comes from the IWC Antarctic surveys. Butterworth and Borchers (1988) estimated G_1 for Antarctic minke whales from the IO surveys conducted on the IWC surveys up to and including the 1986/87 survey using the P and D estimation methods. Borchers (1989) did the same for the 1987/88 survey.

Non-IWC applications of IO methods include the following [the species, location and method involved appear in square brackets]. Barlow (1988) [harbour porpoise, California, Petersen mark-recapture method]; Oien (1990) [minke whales, northeast Atlantic, Petersen method and P method]; Oien (1993) [harbour porpoise, North Sea, P method]; Palka (1992, 1993) [harbour porpoise, Gulf of Maine, P method and MP method]; Buckland and Turnock (1992) [Dalls porpoise, north Pacific, BT method]; Borchers *et al.* (1995) [harbour porpoise, North Sea, BT method].

The IO method estimators from IO surveys with observers on the same survey platform have provided more satisfactory estimates of G than those from VS or PS surveys. (Estimates of G have almost invariably been within the feasible range, and estimated variances have been acceptably small.) From the experience on IWC surveys, placing two observers on the same platform also appears to reduce

the uncertainty associated with duplicate identification substantially. (Butterworth and Borchers, 1988, found that the average percentage of inter-observer pairs which are of uncertain duplicate status on the IWC IO surveys up to 1986/87 had dropped to 17% from the 43% obtained on the PS surveys.) However, there remain some serious problems with the estimators.

One symptom of the problems is the differences obtained between the predicted and observed proportion of duplicate sightings. In order to discuss this problem, I need to define some additional notation. Divide the x-axis into K intervals. Let $d_i(x)$ be a step function with steps at each of the $K-1$ cutpoints between the intervals; $d_i(x)$ equals the number of duplicates in the perpendicular distance interval which contains x, divided by the number of sightings made by the "other" observer (observer $3-i$) in the interval. Note that the D method estimator of G_i is obtained by fitting a smooth function to $d_i(x)$; alternately, $d_i(x)$ is a histogram estimator of the probability that observer i detects an animal at x, given that observer $(3-i)$ detected the animal. Call this conditional probability $p_{i|(3-i)}(x)$. When detections are independent events, as assumed for all the models of this Chapter, $p_i(x) = p_{i|(3-i)}(x)$.

However, it is frequently found that $d_i(x)$ declines slower with x than does $\hat{f}_i(x)$ $(i = 1, 2)$ (see Butterworth and Borchers, 1988, Buckland, 1987b, Oien, 1990 and Palka, 1993, for example). Under the assumption of independence, $f_i(x)$ is simply $d_i(x)$ smoothed and scaled to integrate to unity so that the two functions should decline at the same rate. The observed discrepency has raised doubts about the unbiasedness of the estimator, and has served to focus attention on possible failures of the assumptions of the models. The various hypotheses put forward to explain the discrepancy are discussed below, together with the problem of uncertain duplicate identification.

Unmodelled Heterogeneity: If probability of detection depends on factors other than perpendicular distance, detections by separate platforms are not independent even when observers act independently. Butterworth *at al.* (1982) were probably the first to note the phenomenon in a LT context, although Butterworth *pers. commn* advises that it was discussions with A.R. Hiby which drew the phenomenon to his attention. It is a well known phenomenon in mark-recapture contexts (see Seber, 1982, for example.) One of the ways this lack of indpendence may manifest itself is in the sort of discrepancy between observed and expected distributions of duplicates noted above. (I discuss the phenomenon in more detail in Appendix 6.10.1 of the next Chapter.) Butterworth and Borchers (1988) suggest some possible sources of heterogeneity which might cause the effect observed. Buckland (1992c) and Buckland *et al.* (1993a) discuss sources in more detail, categorizing them into the following four classes.

- *Animal Heterogeneity*: Some animals may be inherently more detectable than others because of their size or behaviour (including surfacing pattern and frequency). There is ample evidence of this sort of heterogeneity in the literature. Examples include Barlow (1988), Schweder, Host and Oien (1990), Polacheck (1991a) and Palka (1993).
- *Environmental Heterogeneity*: Animals passed in periods of favourable environmental condi-

tions will be more detectable than those passed in marginal conditions. In a marine context sea state in particular has often been found to be a critical environmental variable affecting detectability. Examples of environmental heterogeneity in the LT literature include Holt (1987), Holt and Cologne (1987), Barlow (1988), Gunlaugsson (1991a and 1991b) and Palka (1993).

- *Observer Heterogeneity*: Each observer's alertness and efficiency may vary with time, environmental conditions, and platform. (In general I use "observer" to refer to an observation team, which is assumed to be fixed to a single observation platform. In practice, the composition of observation teams varies and individual observers may be swapped between platforms. In the context under discussion here, however, "observer" refers to the actual person making the observations and "platform" refers to the particular vessel/craft and position on the vessel/craft on which the person is located.) Examples appear in Holt and Cologne (1987), Barlow *et al.* (1988), Gunlaugsson and Sigurjonsson (1990) and Polacheck (1992).
- *Platform Heterogeneity*: The relative efficiencies of each platform may vary with environmental conditions and individual observers located on the platform. See Gunlaugsson (1990), Gunlaugsson and Sigurjonsson (1990) and Polacheck (1992) for examples.

Heterogeneity which causes detectabilities of animals to increase and decrease synchronously between platforms (i.e. when an increase in the detectability of an animal from one platform corresponds to an increase in its detectability from the other platform) will induce positive correlation in detection probabilities between platforms and hence positive bias in $\hat{G}_i$. If heterogeneity causes detectabilities of animals to increase and decrease asynchronously between platforms the reverse is true. Synchronous heterogeneity is possibly the more obvious possibility, but asynchronous heterogeneity is possible when, for example, efficient and inefficient observers are swapped between platforms. (Gunlaugsson *pers. commn* was the first to recognize this.) Buckland (1992c) and Buckland *et al.*, (1993a) discuss the phenomenon and give a simple numerical example. Variation in environmental factors like glare, which can temporarily make animals more detectable from one platform than another, is another possible source of negative correlation and hence negative bias in $\hat{G}$.

Unmodelled heterogeneity will in general cause bias in all of the estimators considered in this section. However, in the rather unlikely event that the probability of detection on the trackline is constant over all variables affecting detection probability off the trackline, the MP, BT and D method estimators will remain consistent for G_1 and G_2, providing appropriate functional forms are used for the detection functions. The P method estimator, however, will not be consistent even in this case. (I postpone the demonstration of this to Appendix 6.10.1 of the following Chapter, at which stage I have developed the necessary notation.)

This "constant-G" situation is rather unlikely since unmodelled variables which affect probability of detection off the trackline are likely to affect it on the trackline as well. For example, smaller animals or groups of animals are likely to be less detectable than larger ones both on and off the

trackline; animals are likely to be less detectable in poor sighting conditions than in favourable sighting conditions both on and off the trackline, and so on. When G_1 and G_2 are affected by the unmodelled variables, all the estimators of this section will be biased (see Appendix 6.10.1 of the next Chapter). Note that in this case the D method estimator is estimating the **conditional** probability that observer i detects an animal at 0, given that the other observer has detected it.

Two-Component Detection Functions: Cooke (1987b, 1989) showed that if the probability of detecting an animal consists of two components, one being the probability of the animal being available for detection and the other being the probability of the observer looking in the right direction at the time it presents itself, this can result in discrepencies between the shapes of the observed and predicted distribution of duplicates, even if there is no unmodelled heterogeneity. Cooke (1987b, 1989) presents an analytically tractable example with Poisson availability where observers have identical detection functions. Butterworth and Borchers (1988) discuss the phenomenon in general terms and show that, unlike the P method estimator, the D method estimator remains consistent when this phenomenon is in operation. The MP and BT estimators are not in general consistent for G_i, for the same reasons that the P estimator is not (see Butterworth and Borchers, 1988, for details).

Varying Duplicate Classification Criteria: If a higher percentage of questionable duplicates were called definite duplicates at larger x, the observed distribution of duplicates would appear flatter than that predicted using $f_1(x)$ or $f_2(x)$. This situation might apply if, for example, observers were predisposed to call pairs duplicates unless there was substantial evidence to the contrary. Because there is greater uncertainty in positioning animals at greater distances, observers might tolerate pairs being classed as duplicates which they would have called uncertain were they closer.

This phenomenon will introduce bias in all of the estimators because they are all based on the assumption that duplicates are determined without error.

Non-Uniform Distribution of Animals: If the perpendicular distance distribution of animals in the population (detected and undetected), $\pi(x)$, was peaked at the trackline, $f_i(x)$ would be more peaked than $d_i(x)$. Cooke (1987b, 1989) raises this possibility and notes that this sort of non-uniformity in the distribution of animals could either be real, or an artifact of the way observers define clusters of animals or record data. Real non–uniformity could be a consequence of animals being attracted to the observer, for example, or of clusters fragmenting into more smaller component clusters close to the trackline in reaction to the presence of observers, so generating more clusters at small x. (Remember that when animals cluster, analysis is usually performed treating clusters as "animals" and cluster size as an attribute of the "animal"). Apparent non-uniformity could arise in a number of ways. One possibility is that observers tend to call small clusters in fairly close proximity a single cluster when they are at large perpendicular distance, but would separate them into a number of smaller clusters if they were closer to the trackline. Another possibility is that

observers tend to round small angles to zero, creating an artificial peak in $\pi(x)$ at the trackline.

Whether or not the non-uniformity is real, the D method estimator will provide consistent estimators of G_1 and G_2 because estimation of the detection function for observer i is carried out conditioning on the x's of the other observer's detections. It is not necessary that these x's have any particular distribution. All that is required is that they span the interval $(0; W)$ sufficiently to allow estimation of the detection function. However, the method will not necessarily provide an unbiased or consistent estimator of abundance. For example, if $\pi(x)$ is non-uniform because animals are attracted to the vessel, the sighting rate will be inflated and the abundance estimator will be positively biased.

The BT method is designed to cope with non-uniformity in $\pi(x)$ as a result of responsive movement by the time observer 1 detects animals, and will yield consistent estimates of G_1 and abundance in this case. The BT estimator will, however, not in general be robust to the effects of real or apparent fragmentation of clusters into smaller clusters close to the trackline. The importance of such effects will depend on how clusters detected by observer 2 (before they enter observer 1's detection region) are linked to the smaller clusters after they fragment. The P and MP estimators will in general not remain consistent when $\pi(x)$ is not uniform (for whatever reason).

Random Movement: Kishino (1986) showed that estimates of G can be sensitive to random animal movement. Buckland and Turnock (1992) noted that random animal movement will tend to flatten $d_i(x)$ relative to $f_i(x)$. This is because some of the animals which were detected close to the trackline by one observer will become (or have been) farther from the trackline and so less detectable for at least some of the period during which they were potentially detectable by the other observer. Similarly some of the animals which were detected far from the trackline by one observer will become (or have been) closer to the trackline and so more detectable for at least some of the period during which they were potentially detectable by the other observer. These instances will tend to flatten $d_i(x)$ but will not influence $f_i(x)$. The effect will obviously be larger the greater the speed of animal movement relative to the observer speed. It will also be larger the greater the along-trackline separation between the areas searched by the two observers.

When there is random or responsive movement, the concept of the probability of detecting an animal on the trackline cannot be separated from the effect of movement. Because animals move, the phrase "on the trackline" is not meaningful unless a specific time is attached to it; animals on the trackline one instant may be off it in the next. Estimators of G_i are estimators of the effect of observer i missing some animals which are on the trackline when she is looking on the trackline, **and** of movement between the times the animals are detectable by observers 1 and 2. With random movement this will be lower than the probability of observer i detecting an animal on the trackline at the time she is looking on the trackline (see Buckland and Turnock, 1992).

As a result, all of the estimators yield biased estimators of the probability of observer i detecting an animal on the trackline at the time she is looking on the trackline. The BT estimator is the only

estimator which takes account of this in estimating abundance, and is therefore the only estimator which yields consistent estimates of abundance in the presence of random animal movement.

Uncertain Duplicate Identification Aside from the problems above associated with discrepencies between observed and predicted duplicate proportions, the single biggest difficulty with IO surveys is the duplicate identification itself. IO models assume duplicates are known without error but in practice this is never the case. Uncertainty in duplicate identification may result in biased estimates; but even if it does not, variances and confidence intervals estimated under the assumption that there is no uncertainty associated with duplicate identification will be too small.

By placing both observers on the same survey platform, IO survey proper is an improvement over PS survey because it gives both observers the same perspective on cues and eliminates the error in recording the observers' positions relative to each other. This reduces the uncertainty associated with duplicate identification but does not eliminate it. Uncertainty in point estimates as a result of uncertain duplicate identification remains substantial in comparison with estimated variances conditional on the estimated number of duplicates.

Various methods have been used to identify duplicates. They fall into the following two broad classes. Neither eliminates the uncertainty due to duplicate identification.

(1) *In situ* identification by scientists in the field on the basis of all the information available to them at the time sightings are made. The method is subjective; with the same information, different scientists might reach different decisions on duplicate status. When this method is used the uncertainty in duplicate identification is typically accommodated by classifying potential duplicates on the basis of the scientist's degree of certainty that pairs are duplicates. The classification is then used to give a range of estimates of G spanning the range of classifications. For example, on IWC Antarctic surveys possible duplicate pairs are classified as "Definite" (D), "Possible" (P), "Remotely Possible" (R) or "Non-Duplicate" (N). A range of estimates of G is then presented from the case in which only D's are considered duplicates to that in which all of the D's, P's and R's are considered duplicates (see Butterworth and Borchers, 1988, for example).

(2) *Post hoc* identification on the basis of the times, positions and other attributes of sightings recorded for each platform. This is done either by scientists examining the data after the event (the "Subjective" method) or by using some objective algorithm (the "Objective" method).

Subjective Methods: Duplicate identification is performed making use of data associated with each sighting. The identification would be performed primarily on the basis of recorded species, position and sighting time, but data such as school sizes might also be considered. A classification system like that of *in situ* methods can be used to bound the classification. The technique of scientists performing such classifications "independently" to provide some cross-check has also been applied; examples of this type of classification

are contained in Schweder *et al.*, (1991), Oien (1993) and Palka (1993).

Objective Methods: These methods can be further subdivided into *ad hoc* and probability based methods. The *ad hoc* methods are essentially the same as subjective methods, except that the classification rules are made explicit instead of residing implicitly and invisibly inside the heads of the people making the classifications. An example of this sort of method is contained in Cooke (1995). Probability based methods have been developed by Cooke (1987a) and Hiby and Lovell (1996). The former was shown to perform very poorly on IWC PS and IO datasets (Borchers and Haw, 1989). The latter was developed for aerial surveys and was applied very successfully to aerial surveys of harbour porpoise in the North Sea (Hiby and Lovell, 1996). Their method represents a substantial advance in methods for incorporation uncertain duplicate identification into IO survey models. It is not as yet clear how well it would perform on shipboard or terrestrial datasets.

The form of estimator is tied to the approach for dealing with duplicate uncertainty. For example, neither Cooke's (1987a) nor Hiby and Lovell's (1996) methods involve identifying individual duplicate pairs, but the D method requires such individual pairs to be identified. The development of sound statistical methods for dealing with uncertain duplicate identification is an area which would benefit from more research. I do not, however, deal with this problem any further in this thesis.

5.5 Summary and Discussion

VS estimation methods for G have only ever been applied to IWC surveys of minke whales. Conclusions relating to the success and workability of the estimation methods are therefore based largely on the results of the application of the methods in this context.

VS methods have been shown to perform poorly in practice in cetacean surveys. Reasons for their poor performance include violation of an implicit assumption of the implementation of the methods that there is either no spatial variation or no temporal variation in animal densities, as well as bias caused by random animal movement at low observer speeds. In addition the fundamental equation of the VS models (equation 5.9) may be misspecified because of non–Poisson animal availability.

While IO estimators with PS surveys have performed better than VS estimators, substantial problems in determining duplicates effectively resulted in undesirably wide confidence intervals. There were also indications that PS survey was not representative of normal survey so that estimates obtained from PS survey might not be applicable to normal survey. The use of IO survey proper reduced the size of the duplicate identification problem and largely solved the problem of representativity by making it logistically feasible to survey continuously in IO mode.

The most pressing problems remaining with abundance estimation from IO mode survey are as follows.

(1) **Uncertain Duplicate Identification:** Ultimately, duplicate identification depends on the simi-

larity of the times, positions and associated properties (species, school size, etc.) of the cues detected by the two platforms, and uncertainty in duplicate identification can only be reduced to the extent to which errors in observing and recording times and positions in particular can be reduced. Given that this uncertainty is unlikely to be completely removed by improved survey techniques in the forseeable future, there is a need for models and estimation methods which take account of the uncertainty in a statistically sound manner.

The development of models and estimators which incorporate uncertainty in duplicate identification remains an area of IO LT theory which would benefit from more research, although I do not tackle this particular problem in this thesis.

(2) **Unmodelled Heterogeneity:** This is the main focus of my thesis. In the following Chapters, I develop general models and estimators which incorporate observable heterogeneity. The one sort of heterogeneity which might not easily be incorporated within the framework which I develop is heterogeneity due to discrete availability of animals. If the heterogeneity in availability was due simply to different animals having different mean availability rates, the mean observed rate of availability (the observed number of cues per time unit) contains all the information on the availability of an animal and this statistic can in principle be incorporated as an explanatory variable in the models developed here. However, this statistic might not contain sufficient information on the availability process to incorporate the effect of heterogeneity due to animal availability adequately in the models.

(3) **Animal Movement:** Responsive animal movement is a potentially major source of bias in abundance estimation if it is not taken into account in estimation (see Buckland and Turnock, 1992, for an example). It is often difficult to detect responsive movement (see Leatherwood *et al.*, 1982, Borchers and Haw, 1990, Polacheck, 1990, for example). Random animal movement also leads to bias if the speed of movement is not substantially less than that of the survey platform. Using estimators which are robust to animal movement is therefore critical for unbiased estimation of abundance unless movement at a rate which may affect the estimators is known not to occur.

The problem of uncertain duplicate identification affects all IO estimators. Ignoring it for the moment, how do the P, MP, BT and D method estimators compare in terms of robustness to the difficulties raised in the preceeding section (all of which are failures of model assumptions of one sort or another)?

The P method estimator is in no case less sensitive to failure of a model assumption than the MP method estimator. The MP method estimator of G_1 is identical to the BT method estimator if animals are stationary. But the BT method estimator (used with the BT method survey procedure) outperforms the other estimators in the presence of animal movement as it is the only estimator which is robust to animal movement. (When animals are stationary both the BT and the MP method estimators are consistent estimators of abundance, although the MP method is likely to be the more efficient because it uses the sighting rate information from both observers while the BT method uses the sighting rate information

from observer 1 alone.) The D method estimator of G has the distinction of being the only estimator which remains consistent in the two-component detection function scenario. All these estimators may be biased in the presence of unmodelled heterogeneity.

The nature of IO survey data is such that IO estimators, unlike conventional LT method estimators, cannot in general be robust to unmodelled heterogeneity (see Appendix 6.10.1 of the following Chapter). Unmodelled heterogeneity can in principle be eliminated by modelling the detection functions as functions of the variables causing heterogeneity. In practice this may be difficult, partly because it may not be possible to measure and record all variables affecting detectability, and partly because finite sample sizes and correlation between the various variables affecting detectability will result in variables making a substantial but not statistically significant contribution to heterogeneity in detection probability. Buckland *et al.* (1993a) discuss this problem and suggest some survey procedures which can help to reduce it. Despite these practical difficulties, there is a need to model heterogeneity more adequately, rather than ignore it as is the case with univariate (x only) IO estimators. At worst, adequate modelling of at least some of the effect on the detection functions of variables causing heterogeneity will reduce the size of the resultant bias, at a cost of reduced precision.

What is required is an estimation method which adequately models heterogeneity and combines the "movement robustness" of the BT method with the "two–component detection function robustness" of the D method. I concentrate on the problem of modelling heterogenity in what follows, but in the process of developing estimators for this case, develop an estimator which combines both these aspects of robustness.

5.6 Appendices

5.6.1 A Note on the Similarity of Cooke's and Zahl's Estimators

While Butterworth *et al.* (1982) and Cooke (1985) used variable speed models to obtain estimators of G, the fundamental equations of the models also yield estimators of effective strip width, ω, or mean detection probability at speed v_0, unconditional on x. This mean detection probability, p, is defined as follows.

$$p = \frac{\omega}{W} = \int_0^W p(x \mid v_0) \frac{1}{W} dx \tag{5.27}$$

When $v = 0.5v_o$, an estimator is available in explicit form (see below). Cooke (1985) proposed a version of this estimator based on a negative exponential detection function. For convenience, I call the estimator of p for the more general case $\hat{p}_c$ (subscript "c" for "Cooke") although Cooke did not derive it for the general case. When $v = 0.5v_o$ the similarity between $\hat{p}_c$ and Zahl's (1989) estimator, $\hat{p}_z$ is seen easily. The only difference is in the way $f(x \mid v_0)$ is estimated; Zahl estimates it nonparametrically (without constraining its shape over the range of x), while the $\hat{p}_c$ assumes a smooth functional form for $f(x \mid v_0)$ which constrains its shape (see below).

Obtaining $\hat{p}_c$ from Cooke's equations

Equation 5.9 provides a basis for estimating the effective strip width or mean detection probability at speed v_0 (ω and p, respectively) instead of G. To do this, write $p(x \mid v_0)$ in terms of p and $f(x \mid v_0)$, as follows.

$$f(x \mid v_0) = \frac{p(x \mid v_0)}{\int_0^W p(x \mid v_0) \frac{1}{W} dx} = \frac{p(x \mid v_0)}{\frac{\omega}{W}}$$

$$\Rightarrow p(x|v_0) = \frac{\omega}{W} f(x \mid v_0) = p\, f(x \mid v_0) \tag{5.28}$$

Equation 5.9 can then be written as follows when $v = 0.5v_0$.

$$\frac{E[S_v]}{E[S_{v_0}]} = \frac{\int 1 - (1 - pf(x|v_0))^2 \, dx}{\int pf(x|v_0) \, dx} \tag{5.29}$$

$$= \frac{2\int f(x|v_0) \, dx - p\int f^2(x|v_0) \, dx}{\int f(x|v_0) \, dx} \tag{5.30}$$

$$= 2 - p\int f^2(x|v_0) \, dx \tag{5.31}$$

$$(\text{since } \int f(x|v_0)\,dx = 1.)$$

Now $f(x|v_0)$, the pdf of observed perpendicular distances at speed v_0, can be estimated by conventional LT methods. Conditional on this estimate $\hat{f}(x|v_0)$, and using the observed R_v in place of $E[S_v]/E[S_{v_0}]$, the following is a consistent estimator of p.

$$\hat{p}_c = \frac{\hat{\omega}_c}{W} = \frac{(2 - R_v)}{\int \hat{f}^2(x|v_0)\,dx} \tag{5.32}$$

Cooke (1985) proposed an estimator of this form for the special case in which $p(x|v_0) = Ge^{-\lambda x}$ and $W \to \infty$.

$$\hat{p}_c = \frac{2(2 - R_v)}{\hat{\lambda}} \tag{5.33}$$

(Here $\hat{\lambda}$ is an estimate of λ, which can be obtained using conventional LT methods.)

Zahl's estimator, $\hat{p}_z$

Zahl's estimator of p is obtained after partitioning of the x-axis into K (say) intervals of width W_k $(k = 1, \ldots K)$ up to a truncation distance W $(W = \sum W_k)$. The set of x's in the kth interval is denoted I_k. When $v = 0.5v_0$, the estimator is as follows (equation 4.13, p460 of Zahl, 1989).

$$\hat{p}_z = \frac{(2 - R_v)}{\sum_{k=1}^{K} \frac{\widehat{F_k^2}}{W_k}} \tag{5.34}$$

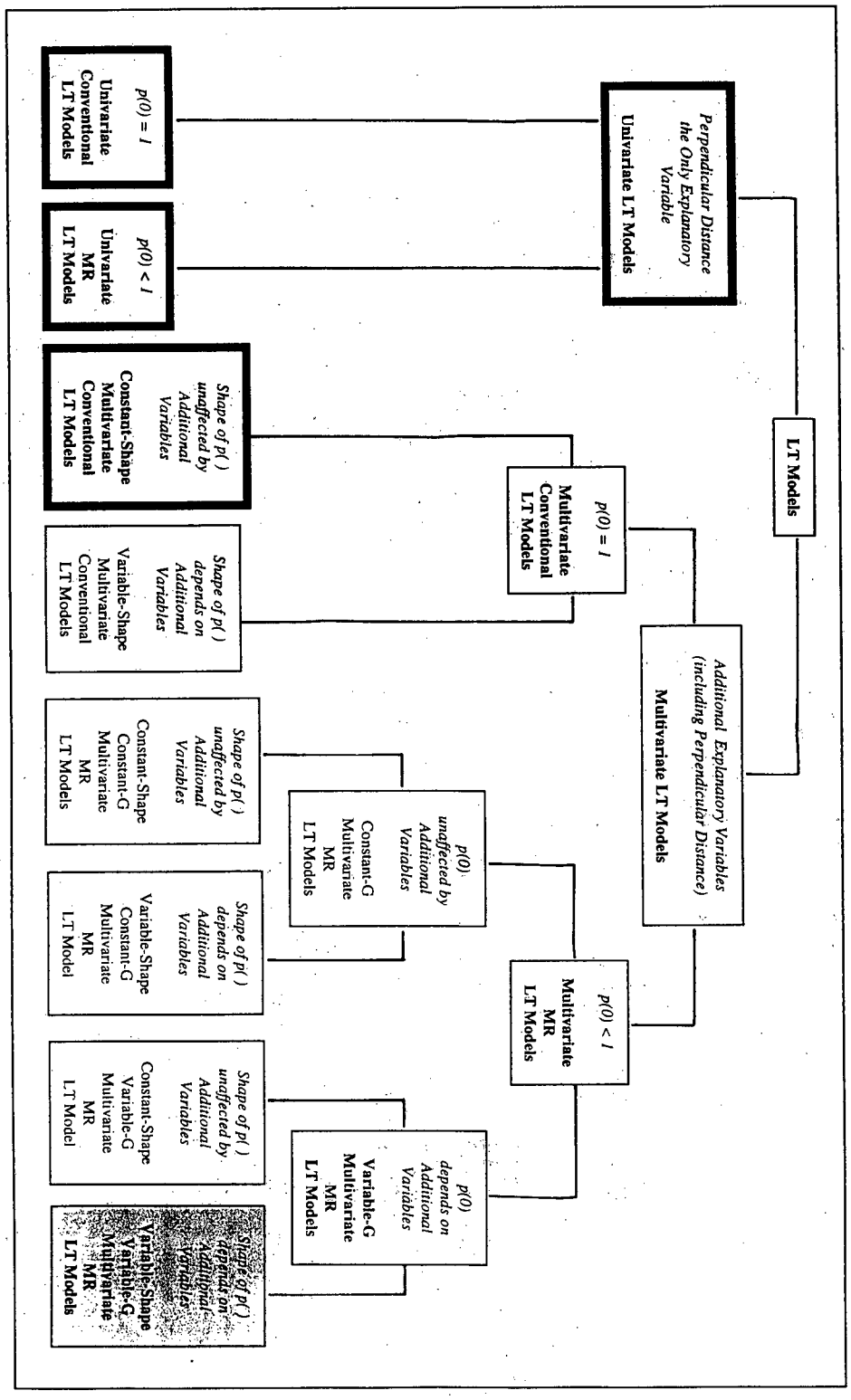
Here $\widehat{F_k^2}$ is an estimator of F_k^2, and $F_k = \int_{I_k} f(v_0|x)\,dx$ is a discrete approximation of equation 5.32 insofar as the term within the sum is a discrete approximation to the integral in equation 5.32. Write F_k as $\bar{f}_k W_k$, where $\bar{f}_k$ is the mean value of $f(x|v_0)$ in the kth interval. Then

$$\lim_{K\to\infty} \sum_{k=1}^{K} \frac{F_k^2}{W_k} = \lim_{K\to\infty} \sum_{k=1}^{K} \bar{f}_k^2 W_k \tag{5.35}$$

$$= \int_0^W f^2(x|v_0)\,dx \tag{5.36}$$

Now, rather than estimating F_k^2 or $f^2(v_0|x)$ by conventional LT methods, Zahl estimates F_k, and hence F_k^2, separately within each interval. (One obvious estimator of F_k is $n_k/\sum n_k$. Zahl, 1989a, proposes an alternate estimator.) This piecewise estimation of the pdf $f(x)$ allows estimation with only weak assumptions about the form of the detection function. One anticipates that the cost of this flexibility in the form of the detection function is increased variance as a result of the fact that unless K is very small, more paramaters need to be estimated; that is a $\hat{p}_z$ will be a less efficient estimator than $\hat{p}_c$.

Figure 6.1: A Classification of Line Transect Models



5.6.2 Continuous *vs* Poisson Availability LT Models

Assume that animals are stationary and that an observer moves at a constant speed v along the trackline (in the y-direction). Define $h_Y(y \mid x)$ and $f_{Y|X}(y|x)$ as in Chapter 2. A discrete availability process is incorporated within this framework as follows. For an animal at perpendicular distance x, assign to the probability density $f_{Y|X}(y|x)$, a component $f_j(x)\delta(y - y_j)$ for an atom $f_j(x)$ at y_j, where $\delta(\,)$ denotes the Dirac delta function. If there is an atom $f_j(x)$ of probability at y_j, $h_Y(y|x)$ contains a component $h_j(x)\delta(y - y_j)$, where $h_j(x) = f_j(x)/F_{Y|X}(y_j|x)$. For a discrete availability process with atoms $f_1(x), f_2(x), \ldots$ at y distances $y_1, y_2, \ldots$ $(y_1 > y_2 > \ldots)$, $h_{Y|X}(y|x) = \sum_j h_j(x)\delta(y - y_j)$. Assume that the availability process (or "cue process") for each animal is Poisson, with rate parameter μ per unit time, and that animals produce cues independently of one another. Consider first an interval of length l from $y = 0$ to $y = l$. The number of times an animal becomes available in this interval is a random variable, called m. The detection function for a single observer can be written as follows.

$$p(x) = 1 - E_m\left[E_{\boldsymbol{y}}\left[\prod_{j=1}^{m}(1 - h_Y(y_j \mid x)) \mid m\right]\right] \tag{5.37}$$

Here

$\boldsymbol{y}$ is a vector of length m containing the along-trackline distances $y_1 \ldots y_m$ at which the animal was available for detection,

$E_m[\ ]$ denotes expectation with respect to m, the number of cues, and

$E_{\boldsymbol{y}}[\ |m]$ denotes expectation with respect to $\boldsymbol{y}$, the positions of the cues, given that there were m cues.

Assume that m is Poisson with rate parameter $\lambda = \mu l/v$ cues per l y-distance units. Then, conditional on m, the points at which the animal presents cues $(y_j;\ j = 1 \ldots m)$ are independently and identically distributed with uniform density $\frac{1}{l}$ (using a result from Parzen, 1962, p140). Thus (using the independence of the y_j's and the fact that each is identically and uniformly distributed),

$$p(x) = 1 - E_m\left[\prod_{j=1}^{m}\left(1 - E_{\boldsymbol{y}}\left[h_Y(y_j \mid x)\right]\right)\right] \tag{5.38}$$

$$= 1 - E_m\left[\left(1 - \frac{1}{l}\int_0^l h_Y(y \mid x)\, dy\right)^m\right] \tag{5.39}$$

$$= 1 - E_m\left[(1 - H(x))^m\right] \tag{5.40}$$

where $H(x) = l^{-1}\int_0^l h_Y(y \mid x)dy$. Now using the fact that m is Poisson with rate parameter λ, $p(x)$ can be expressed as follows.

$$
\begin{aligned}
p(x) &= 1 - \sum_{m=0}^{\infty} \frac{\lambda^m e^{-\lambda}}{m!}(1 - H(x))^m \\
&= 1 - \sum_{m=0}^{\infty} \frac{[\lambda(1 - H(x))]^m e^{-\lambda(1-H(x))}}{m!} \left(\frac{e^{-\lambda}}{e^{-\lambda(1-H(x))}} \right) \\
&= 1 - e^{-\lambda H(x)} \\
&= 1 - exp\left\{ -\frac{\mu}{v} \int_0^l h_Y(y \mid x)\, dy \right\} \qquad (5.41)
\end{aligned}
$$

Let $l \to \infty$, to get the following.

$$
p(x) = 1 - exp\left\{ -\int_0^{\infty} \frac{\mu}{v} h_Y(y \mid x)\, dy \right\} \qquad (5.42)
$$

This is identical to the continuous hazard rate model, but with the detection hazard multiplied by a constant, $\frac{\mu}{v}$, which is the mean number of occasions per unit y-distance at which the animal is available for detection. The detection function for the discrete case with Poisson availability is identical to that from the continuous availability case with the detection hazard scaled by a measure of the mean availability of animals for detection.

Chapter 6

Line Transect as Mark-Recapture: General Models

6.1 Overview

In this Chapter general models are developed for line transect surveys in which two observers search independently for the same target animals. The models are shown to be particular forms of two-sample mark-recapture (MR) models in which detection probabilities are functions of a set of explanatory variables which includes perpendicular distance. Models are developed for both grouped/binned and ungrouped/unbinned data. They represent a marriage of MR and LT models. I call them "Mark-Recapture Line Transect models" (MRLT models). Viewed from one perspective, they are MR models with capture history probabilities parameterized in terms of observable heterogeneity. Viewed from another, they are LT models which incorporate additional explanatory variables to x, and allow for uncertain detection of animals on the trackline

The models extend MR models in the literature which accommodate heterogeneity in individual capture probabilities in the following two main respects.

(1) The MR likelihoods in the literature are conditional likelihoods (conditional on the obseved explanatory variables). The likelihoods developed here incorporate a density for the explanatory variables. This gives the option of estimating abundance either conditional on the obseved explanatory variables or based on the full likelihood.

(2) MR models are available for the unbinned case only. The likelihoods developed here include the full likelihood for binned data, as well as the binned equivalent for the conditional MR likelihood for unbinned data.

Viewed from the LT perspective, the models extend LT models in the literature in two ways.

(1) They include explanatory variables other than perpendicular distance (as well as densities for the explanatory variables).

(2) They include MR components which, with two independent observers, allow detection functions parameterized as functions of x and other explanatory variables to be estimated without having to assume that detection on the trackline is certain.

Existing models for independent observer LT surveys are placed in their appropriate context within the general model. Some current estimators are shown to correspond to maximum likelihood estimators (mle's) for special cases of the general model, and new mle's are derived for other special cases. New abundance estimators for the general model are developed in Chapter 7.

Models are developed for a two-observer/two-capture scenario. They can be extended to accommodate multiple observers/recaptures, but this extension is not discussed here as it has little relevance for line transect surveys as currently defined.

6.2 Introduction

6.2.1 Line Transect as an Experiment

The survey is modelled as an experiment with each animal corresponding to a **trial** in which the possible outcomes or **responses** are {detection by observer 1}, {detection by observer 2}, {detection by both 1 and 2} or {detection by neither observer}. (The last outcome is unobservable.) Associated with each trial is the set of attributes or explanatory variables (one of which would typically be the perpendicular distance of the animal from the trackline).

Conventional LT models with a single observer team are framed in experimental trial-response terms as follows. Each animal is a trial, but in this case there are only two possible outcomes, namely {detection } or {no detection}. (Again the last outcome is not observable.) Perpendicular distance would usually be the only explanatory variable (although some models in the literature have also included cluster size).

Animals are assumed to have different detection probabilities, which depend on the attributes associated with the animal at the time of the survey. In the simplest model, perpendicular distance is the only attribute. In general there is a vector of attributes, which would typically include perpendicular distance. The attributes may be "fixed" to the animals (size, for example) or only temporarily associated with them (perpendicular distance from the trackline; some measure of suitability of environmental conditions for detection at the time the animal is detectable, for example). In the latter instance they are nevertheless assumed to be fixed for the period during which they are available for detection, i.e. for the period during which the animal is considered a "trial".

When animals are not continuously available for detection, one relevant attribute is the sequence of positions at which the animal was available for detection. With detection functions framed in terms of the probability of detection at a perpendicular distance x (as they are here), rather than in terms of a detection hazard (as are the models of Schweder, 1990, for example), incorporating the sequence of positions presents a difficulty. If all, or nearly all, of the information pertaining to detection probability

which is contained in the sequence of positions can be summarized in a scalar variable (mean observed surfacing rate, for example) the sequence of positions can in principle be incorporated in the models through this variable. If this is not the case there are two options. One is to try to destroy the correlation between detections which the variable surfacing patterns induce, by way of survey procedures (in principle separating the areas searched by the two observers in the y-dimension can do this). The other is to use methods designed specifically for discrete availability LT surveys (which were mentioned in Chapter 5 but not covered in any detail in this thesis).

An important difference between conventional LT models (which assume detection on the trackline to be certain and use only one observer) and MRLT models (which do not assume certain detection on the trackline and use two independent observers) is the extent to which explanatory variables which affect detectability can be neglected. This is a major condition in determining the level of complexity needed for the model. When detection of all animals on the trackline is certain, then as long as suitable forms are used for the detection functions, nothing is lost by pooling data over all explanatory variables. This follows from a theorem of Patil *et al.* (1993), which was given in Chapter 3.

Using similar arguments to those of Patil *et al.* (1993), one can show that this is not the case for surveys in which detection on the trackline is not certain. In this case, the pdf of the observed data is not the same when mean detection functions (averaged over $\boldsymbol{z}$) are used in place of detection functions which depend on $\boldsymbol{z}$ - even when the true forms of the mean detection functions are known. (Showing this requires notation which is developed only later in this Chapter, so I have postponed it to Appendix 6.10.1.)

With conventional LT models one can ignore variables other than perpendicular distance even though they may have a large effect on the detection function. In the MRLT scenario unbiased inference requires the detection function to be modelled as a function of all the variables which affect detectability. In general, therefore, if detection on the trackline cannot be assumed to be certain, unbiased inference depends on modelling detection probability as a function of more than one explanatory variable. This adds another dimension (or several dimensions) to the estimation problem.

6.2.2 Types of LT Models

Both the "$g(0)$" problem and methods of incorporating additional explanatory variables in the detection functions have received attention in the LT literature, but usually as separate problems. Work on "$g(0)$" estimation has largely (although not exclusively) neglected modelling explanatory variables other than perpendicular distance. Attempts to incorporate additional explanatory variables have largely been made under the assumption that detection on the trackline is certain. The models of this Chapter simultaneously incorporate additional explanatory variables and allow "$g(0)$" to be a function of those explanatory varaibles, bringing these two areas of research together in a single likelihood framework.

Incorporating explanatory variables other than x when detection on the trackline is not certain requires the formulation of detection probabilities as appropriate functions of these explanatory variables. Devel-

opment of appropriate forms for this case has not received much attention. In fact, the only models for the detection function for this case in the literature have the explanatory variables entering the detection function through the detection function scale parameter only. While this may be an appropriate way for the additional explanatory variables to enter the detection function when detection on the trackline is certain (see the simulation study of Otto and Pollock, 1990), it may not be appropriate when this is not the case. The models developed here allow probability of detection on the trackline to vary as a function of these other explanatory variables. (With suitable parameterization of the detection functions, one can also model this probability as a constant; the general models of this Chapter readily incorporate the specific models of earlier Chapters as special cases.)

The models provide a general likelihood framework within which all current LT models can be placed as special cases. To give an overview of the types of models covered, it is useful to classify LT models according to some criteria. The criteria I have used are as follows.

(1) **Univariate or Multivariate?** A LT model is "univariate" if detection probability depends on x only. If it depends on x and z it is "multivariate". (Note that the use of "multivariate" in this sense is contrary to the conventional use in statistical literature, where it usually implies multiple responses rather than multiple explanatory variables.)

(2) **Conventional or MR?** "MR" (for "mark-recapture") indicates a model in which detection of animals which are on the trackline is not assumed to be certain and two independent observers are used. If a LT model is not an MRLT model, it is a "Conventional" LT model.

(3) **Constant Shape or Variable Shape?** A model has a "constant- shape" if variables other than x affect only the scale parameter of the detection function, if they have any effect at all. Obviously all univariate models are constant-shape models.

(4) **Constant-G** or **Variable-G ?** A model is a "constant-G" model if the probability of detecting animals which are on the trackline is not affected by any explanatory variables. All univariate LT models are by definition constant-G models, as are all Conventional LT models.

This classification of LT models is illustrated in Figure 6.1. The bold boxes correspond to models for which substantial LT theory already exists. Models written in bold type are covered in this thesis (in varying degreees of detail). The shaded box is the most general model, which is the central focus of this thesis.

6.3 The Elements of the Model and the Notation

The notation in this Chapter is necessarily fairly complicated. In this section I describe the notational conventions I use for the data, the detection probabilities, the explanatory variables and their densities. These are the building blocks for the likelihood functions introduced in the following section.

6.3.1 The Data

Say that $n_{..}$ of the N animals in the survey area are detected by at least one observer in the course of the survey. (The rationale for the subscripts is given below.) The observed responses and explanatory variables can be written conveniently as follows.

$$\begin{bmatrix} c_1 \\ \vdots \\ c_{n_{..}} \end{bmatrix} \begin{bmatrix} x_1 & z_{11} & \cdots & z_{1R} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_{..}} & z_{n_{..}1} & \cdots & z_{n_{..}R} \end{bmatrix}$$

Each row corresponds to a trial (i.e. an animal). Here c_j is the observable **response** of interest for the jth detected animal. It is the capture history of the jth detected animal.

c_j = 1 if the jth animal was detected by observer **1 only**.
= 2 if the jth animal was detected by observer **2 only**.
= 3 if the jth animal was detected by **both** observers.

The scalar x_j is the perpendicular distance of the jth detected animal when detected. I develop the models under the assumption that animals do not move, although in many applications there will be some animal movement. It may be possible to accommodate movement using a survey procudure developed by Buckland and Turnock (1992). Aside from a brief discussion of an estimator specifically adapted for this case in section ?? and Appendix 7.10.3 of the following chapter, I do not consider animal movement further in this thesis.

The row vector $\boldsymbol{z}_j = [z_{j1}, \ldots, z_{jR}]$ is a vector of the R attributes associated with the jth animal during the time it is detectable. (The attributes are assumed to stay fixed for the period in which the animal is detectable. They are also assumed to be the same for both observers. Note, however, that although the attributes may be the same for both observers, the way in which different observers' detection functions depend on the attributes may be different.) The vector $\boldsymbol{z}$ might typically be composed of attributes like cluster size, sea state (for marine surveys), survey platform index, stratum index (for a stratified survey), observer index, glare, and so forth.

6.3.2 The Detection Probabilities

The likelihood functions for the models developed here can be expressed either in terms of the LT detection functions or in terms of probabilities of observing particular capture histories. This makes it necessary to define two sorts of probability relating to detection. The first is the LT detection function, and the second is the probability of observing a particular capture history.

Let $p_i(x, \boldsymbol{z})$ be the probability that observer i detects an animal "at" $(x, \boldsymbol{z})$ $(i = 1, 2)$.

The observers are assumed to search independently of each other. Provided that $(x, \boldsymbol{z})$ includes all explanatory variables which affect detectability, detections by each observer are "conditionally independent". This means that they are independent given $(x, \boldsymbol{z})$. In this case, the conditional capture history probabilities, given $(x, \boldsymbol{z})$, are defined as follows.

$P_1(x, \boldsymbol{z})$ is the probability that observer 1 detects an animal "at" $(x, \boldsymbol{z})$, and observer 2 does not detect the animal:

$$P_1(x, \mathbf{z}) = p_1(x, \mathbf{z}) \times (1 - p_2(x, \mathbf{z}))$$

$P_2(x, \boldsymbol{z})$ is the probability that observer 2 detects an animal "at" $(x, \boldsymbol{z})$ and observer 1 does not detect the animal:

$$P_2(x, \mathbf{z}) = (1 - p_1(x, \mathbf{z})) \times p_2(x, \mathbf{z})$$

$P_3(x, \boldsymbol{z})$ is the probability that both observers detect an animal "at" $(x, \boldsymbol{z})$:

$$P_3(x, \mathbf{z}) = p_1(x, \mathbf{z}) \times p_2(x, \mathbf{z}).$$

(Note that the equations for $P_i(x, \boldsymbol{z})$ $(i = 1, 2, 3)$ hold only if detections are independent, given $(x, \boldsymbol{z})$.) A p or P subscripted with "." is used to denote the probability of at least one of the observers observing an animal "at" $(x, \mathbf{z})$ (or equivalently the probability of observing capture history 1 or 2 or 3), i.e.

$P_\cdot(x, \mathbf{z})$ is the probability that at least one of the two observers detects an animal at $(x, \mathbf{z})$:

$$\begin{aligned} P_\cdot(x, \mathbf{z}) &= P_1(x, \mathbf{z}) + P_2(x, \mathbf{z}) + P_3(x, \mathbf{z}) \\ &= p_1(x, \mathbf{z}) + p_2(x, \mathbf{z}) - p_1(x, \mathbf{z}) p_2(x, \mathbf{z}) \\ &= 1 - [1 - p_1(x, \mathbf{z})] \times [1 - p_2(x, \mathbf{z})] \end{aligned}$$

Note that $p_\cdot(\,)$ and $P_\cdot(\,)$ are equivalent and are used interchangeably, depending on the context.

6.3.3 Numbers of Animals Observed

Notation for the numbers of animals observed is similar to that for detection probabilities, with the following two additions.

(i) A "+" superscript is added when I need to indicate that the number includes animals detected by both observers.

(ii) A subscript is added to index individual animals or "bins" (see below). The subscript is different when data are "binned" compared to that used when they are "unbinned". "Binned" data are data for which the explanatory variables are divided into discrete sets spanning the domain of the variables; this is the multi-dimensional equivalent of grouping the data into perpendicular distance intervals in conventional LT models. An additional subscript j is added (for the jth animal) if data are unbinned. An additional subscript k is added (for the kth bin) if data are binned, as described below.

In the **unbinned data** case:

n_{ij}^{+} =1 if the jth animal is detected by observer i ($i = 1, 2$)
=0 otherwise

n_{cj} =1 if the jth animal has capture history c ($c = 1, 2, 3$)
=0 otherwise (in which case it is not detected)

$n_{.j}$ =1 if the jth animal was detected by observer 1 or 2 or both
=0 otherwise.

(Note that $n_{1j}^{+} = n_{1j} + n_{3j}$ and $n_{2j}^{+} = n_{2j} + n_{3j}$.)

In the **binned data** case:

n_{ik}^{+} is the number of animals from the kth bin which were detected by observer i ($i = 1, 2$))
(some of which may have been detected by the other observer as well)

n_{ck} is the number of animals from the kth bin with capture history c ($c = 1, 2, 3$)

$n_{.k}$ is the number of animals from the kth bin which were detected by observer 1 or 2 or both
(i.e. the total number of animals from the kth bin which were detected).

(Note that $n_{1k}^{+} = n_{1k} + n_{3k}$ and $n_{2k}^{+} = n_{2k} + n_{3k}$.)

For both **binned and unbinned data**:

$n_{i.}^{+}$ is the number of animals detected by observer i (some of which
may have been detected by the other observer as well)

$n_{c.}$ is the number of animals with capture history c ($c = 1, 2, 3$)

$n_{..}$ is the number of animals which were detected by observer 1 or 2 or both
(i.e. the total number of animals which were detected).

(Note that $n_{1.}^{+} = n_{1.} + n_{3.}$, $n_{2.}^{+} = n_{2.} + n_{3.}$ and $n_{..}^{+} = n_{1.} + n_{2.} + n_{3.} = n_{1.}^{+} + n_{2.}^{+} - n_{3.}$.)

6.3.4 The Density of the Explanatory Variables

Let the joint density of the explanatory variables x and z be $\pi(x, z)$, the marginals be $\pi(x)$ and $\pi(z)$, and the conditionals be $\pi(x|z)$ and $\pi(z|x)$. Let I_k be the subset of explanatory variable points falling in the kth bin ($k = 1, \ldots K$). Finally let $\int_{I_k} \ldots dx\, dz$ represent integration over the domain of the kth bin, and $\int_{I_k} \ldots dx$ represent integration over the range of perpendicular distances of the kth bin.

6.3.5 Mean Detection Probabilities and Effective Strip Width

Various sorts of mean detection probability arise in the MRLT likelihoods of this section. For example, there are mean detection probabilities given x, and averaged over z, and there are mean detection probabilities given z, and averaged over x. The same symbols "P" and "p" are used to denote mean capture history/detection probabilities as are used to denote these probabilities as functions of all the

explanatory variables. The expectation over some explanatory variables is indicated by making the second subscript "$\cdot$". Mean detection probabilities are a sometimes modelled as a function of one or more explanatory variables. Subscripts are used on P's in the same way as on the n's.

The Mean Detection Probabilities

Mean Capture History Probabilities:

P_{ck} is the mean probability of observing capture history c for an animal in the kth bin:
$P_{ck} = \int_{I_k} P_c(x,z)\,\pi(x,z)\,dx\,dz$

$P_{\cdot k}$ is the mean probability of either observer observing an animal in the kth bin:
$P_{\cdot k} = \int_{I_k} P_{\cdot}(x,z)\,\pi(x,z)\,dx\,dz$

$P_{c\cdot}$ is the mean probability of capture history c, over the whole domain of the explanatory variables:
$P_{c\cdot} = \int\int P_c(x,z)\,\pi(x,z)\,dx\,dz$

$P_{\cdot\cdot}$ is the mean probability of observing animals over the whole domain of the explanatory variables
$P_{\cdot\cdot} = \int\int P_{\cdot}(x,z)\,\pi(x,z)\,dx\,dz.$

Considered as functions of the explanatory variables x or z, the convention used is that expectation has been taken over the explanatory variable(s) which do not appear as arguments of the function.

$P_{c\cdot}(z)$ is the mean probability of capture history c, over the whole domain: of x, **given** z
$P_{c\cdot}(z) = \int P_c(x,z)\,\pi(x|z)\,dx.$

$P_{c\cdot}(x)$ is the mean probability of capture history c, over the whole domain: of z, **given** x
$P_{c\cdot}(z) = \int P_c(x,z)\,\pi(x|z)\,dx.$

$P_{\cdot\cdot}(z)$ is the mean probability of observing an animal over the whole domain of x, **given** z:
$P_{\cdot\cdot}(z) = \int P_{\cdot}(x,z)\,\pi(x|z)\,dx$.

Mean Detection Function Probabilities:

Notation is the same as that for mean capture history probabilities, except that p's are used instead of P's. For example, $p_{1.}$ is the mean probability of observer 1 detecting an animal in the surveyed area, whereas $P_{1.}$ is the mean probability of observing capture history 1, i.e. of observer 1 **only** detecting an animal in the surveyed area.

Note that whenever the first subscript is a dot the "P"s and "p"s are identical.

For example $p_{.k} = P_{.k}$ = the probability of observer 1 or 2 detecting an animal in the kth bin. (Here the expectation is over the domain of the kth bin.)

Effective Strip Width

The symbol ω is used to denote effective strip width (esw). When there are other explanatory variables in addition to perpendicular distance, esw is modelled as a function of the explanatory variables. The notational conventions used for ω are identical to those used for the LT detection functions p. Under the assumption of uniformly iid perpendicular distances of animals from the transect line, there is a simple relationship between ω and p, namely: p (with given subscripts and arguments) is equal to ω (with the same subscripts and arguments) divided by W, where W is the maximum perpendicular distance considered in the model. In other words, if

$$\begin{aligned} \pi(x) &= \frac{1}{w} \quad \text{for } 0 \leq x \leq W \\ &= 0 \quad \text{otherwise} \end{aligned}$$

then, using p_{ik} as an example: $p_{ik} = \omega_{ik}/W$ or $W \times p_{ik} = \omega_{ik}$, and so on.

6.4 The General Likelihoods

The analysis is based on one of two general likelihood functions - one for binned data, the other for unbinned data. The likelihoods are derived under the assumption that all animals are identically and independently distributed with respect to the explanatory variables (x, z). While this independence with respect to x is likely to be violated in most applications (because animals do not generally position themselves completely independently of the positions of other animals), unbiased abundance estimators obtained under the assumption remain unbiased if it is violated. In practice, the variance of the abundance estimates and confidence intervals for abundance can be estimated using methods which do not incorporate this spatial independence assumption, or at least incorporate it to a lesser degree (see Buckland *et al.*, 1993a), so that estimators derived under the independence assumption remain useful.

The likelihoods are formulated with respect to N, the number of animals within W of the trackline at the time of the survey (see Figure 1.1), and not the mean density, D, of animals per unit area in the survey

region. In practice, estimates of N are readily converted into estimates of D by dividing them by the surface area of the covered region ($a = 2LW$ where L is the total trackline length surveyed). Estimates of D are converted to estimates of the population total in the survey region, $\aleph$, by multiplying them by the surface area of the survey region, A.

Conceptually, N is a realization of a counting process with rate parameter D. Assuming an equal coverage probability design and that animals fall in the covered region independently of one another, N is a binomial random variable with parameters $\aleph$ and a/A. In this case the likelihoods for N are easily converted into likelihoods for $\aleph$ by multiplying them by the following binomial likelihood for $\aleph$ given N.

$$L_{\aleph}\{\aleph \mid N\} = \left(\frac{\aleph!}{N!(\aleph - N)!}\right)\left(1 - \frac{a}{A}\right)^{\aleph - N}\left(\frac{a}{A}\right)^{N} \tag{6.1}$$

This approach treats the number of animals in the survey region ($\aleph$) as a constant. An alternative approach is to consider the density D as a constant in the survey area. Under the same assumptions as above the likelihood for D, given N would then be as follows (see also the subsection "A Note on f_1" below).

$$L_D\{N|D\} = \frac{(Da)^N e^{-Da}}{N!} \tag{6.2}$$

Burnham (*pers. commn*) pointed out that equation 6.1 also arises as a Poisson approximation to equation 6.2 if a/A is small and $D = N/A$.

In the absence of a probability model for N, replicate transects in the survey region can be used as sampling units for variance and interval estimation for D or $\aleph$. Buckland *et al.* (1993a) discuss the merits and demerits of modelling the process generating N.

A little more notation: I use a subscripted symbol in braces to denote a list of symbols over the range of the subscript(s). For example $n_{.1}, n_{.2}, \ldots n_{.K}$ is denoted $\{n_{.k}\}$; $n_{11}, \ldots n_{1K}$, $n_{21}, \ldots n_{2K}$, $n_{31}, \ldots n_{3K}$ is denoted $\{n_{ck}\}$. Usually the range of the subscript(s) is apparent from previous appearances of the subscript(s) in the text so that I do not state it explicitly each time I use the subscript(s). I use "L"'s with subscripts for likelihoods and "f"'s for pdf's. In cases where only the pdf is given and the likelihood is referred to, this is the likelihood obtained by treating the pdf as a function of the unknown parameters in question.

6.5 The Likelihood for Binned Data

In this case, detections are grouped or "binned" into K intervals or "bins" by dividing the whole domain of $(x, \mathbf{z})$ into K mutually exclusive "bins" (where the domain of the kth bin is denoted I_k). The appropriate likelihood is the likelihood for N, given that observer 1 only detected n_{1k} animals (i.e n_{1k} with capture

history $c = 1$) from the kth bin, observer 2 only detected n_{2k} animals (i.e n_{2k} with capture history $c = 2$) from this bin, and both observers detected n_{3k} animals (i.e n_{3k} with capture history $c = 3$) from the bin (for $k = 1, \ldots, K$). This likelihood is derived in Appendix 6.10.2). It is as follows (the subscript "B" is for "Binned").

$$L_B\{N \mid \{n_{ck}\}\} = \left(\frac{N!}{\prod_c \prod_k n_{ck}!(N - n_{..})!}\right)(1 - P_{..})^{N-n_{..}} \prod_c \prod_k (P_{ck})^{n_{ck}} \tag{6.3}$$

remembering that:

P_{ck} is the average probability of observing capture history c, over the kth bin ($P_{ck} = \int_{I_k} P_c(x, z)\pi(x, z)\, dx\, dz$).

$P_{..}$ is the combined average detection probability over the domain of (x, z) ($P_{..} = \sum_{k=1}^{K} \sum_{c=1}^{3} P_{ck}$).

While the form of this likelihood corresponds to an almost standard mark recapture likelihood, its parameterization and derivation in terms of observable explanatory variables (x, z) is new.

By partitioning the likelihood appropriately one can easily see the relationship of L_B to more conventional MR likelihoods and LT likelihoods for binned data, as illustrated below.

6.5.1 L_B from a MR Perspective

The MR nature of the model can be seen by partitioning the likelihood as follows:

$$L_B\{N \mid \{n_{ck}\}\} = f_1\{n_{..}|N\} \times f_{B2\&3}\{\{(n_{ck}\}|n_{..}\} \tag{6.4}$$

where:

$$f_1\{n_{..}|N\} = \left(\frac{N!}{n_{..}!(N - n_{..})!}\right) P_{..}^{n_{..}} (1 - P_{..})^{N-n_{..}} \tag{6.5}$$

$$f_{B2\&3}\{\{n_{ck}\}|n_{..}\} = \left(\frac{n_{..}!}{\prod_{c=1}^{3} \prod_{k=1}^{K} n_{ck}!}\right) \prod_{c=1}^{3} \prod_{k=1}^{K} \left(\frac{P_{ck}}{P_{..}}\right)^{n_{ck}} \tag{6.6}$$

(The parameters of the detection functions and of the density $\pi(x, z)$ are not shown explicitly in order to keep the equations simple and readable. They enter the likelihood through the P's.)

f_1 is the probability of observing $n_{..}$ animals, given the number of animals in the covered region (N) and the form and parameters of the detection functions and of the density of the explanatory variables.

$f_{B2\&3}$ is the probability of observing n_{ck} animals from the kth bin with capture history c, given that $n_{..}$ animals were observed in total (and given the detection functions and density of explanatory variables).

Written in this form, it is easy to see the link with MR models: the likelihood L_B is identical in form to the multinomial MR likelihood described in Sandland and Cormack (1984) and Cormack (1989) (and elsewhere). The difference between this and the multinomial MR model lies only in the way the mean capture history probabilities (the P_{ck}'s) are parameterized. In MR models "proper", these probabilities are typically parameterized in terms of biological parameters such as survival rates, trap-response and perhaps a single parameter for "catchability" at each capture occasion. In the MRLT models developed here, they are parameterized in terms of observers' detection functions (which are functions of the observers' efficiency, the attributes associated with the target animals at the time of the survey, and so on) and the density of the explanatory variables in the target population, $\pi(x, z)$. They are specific forms of MR models; ones in which **observable heterogeneity** due to x and z is an absolutely central feature. Heterogeneous detectability due to perpendicular distance from the trackline is a central feature of all LT models, and heterogeneity due to other attributes (z) is usually present in practice although it is often neglected in modelling.

The central role of **observable heterogeneity** in these MRLT models distinguishes them from most MR models in the literature in two ways. The first is that it requires that detectability ("catchability" in MR terms) be modelled as a function of the observed heterogenity between animals (at least as a function of the perpendicular distances of each animal). It is only recently that MR models for doing this have appeared in the literature. Huggins (1989) and Alho (1990) developed estimators for the case where data are unbinned. Pollock *et al.* (1984) developed a model for binned data conditional on N_k, the abundance of animals in the Kth bin. Although his model incorporated other explanatory variables, z, in a (logistic) "detection function", it incorporated them as **factors** (i.e. separate parameters for separate bins). The model of Pollock *et al.* (1984) does not constrain either the detection function or the density of z between bins.

This point relates to the second distinguishing feature of the MRLT models. They focus attention on the role of the distribution of the heterogeneity in the target population, $\pi(x, z)$. Here, as in all LT models, perpendicular distance plays a special role in that it is often reasonable to treat $\pi(x)$ as known. In fact, modelling $\pi(x)$ as a uniform distribution on a perpendicular distance interval $(0, W)$ is so implicit in LT models that it is easy to forget that it is being modelled at all. The opposite is true of MR models "proper". Burnham (1972), Burnham and Overton (1978), Pollock and Otto (1983) and Pollock *et al.* (1984) have all proposed MR models which incorporate a model for the **density of the detection probabilities**, but no attempt has been made in the MR literature to model the **density of explanatory variables** affecting detectability. Is there anything to be gained from doing so? This is one of the questions I try to answer in the following Chapters.

A Note on f_1

The essential equivalence of L_B with the multinomial MR model means that at least some of the model development for the MR situation is available for direct application to the MRLT situation. The particular development I want to mention here is an alternative model for f_1, namely Sandland and Cormack's (1984) Poisson model. Instead of using a binomial with parameters N and $P_{..}$ for f_1, one could take a slightly different perspective and use a Poisson with rate parameter $(P_{..}N)$. Sandland and Cormack (1984) discuss the differences between these two approaches and I'll not repeat that discussion here, except to say that in a LT context the difference is a philosophical one: in the multinomial model the number of animals in the covered region is considered a fixed parameter; in contrast, in the Poisson model the mean density of animals in the covered region is considered a fixed parameter, while the number there at the time of the survey is considered a random variable.

The point to note is that although I use a multinomial f_1 throughout this Chapter, one could equally work with a Poisson f_1.

6.5.2 L_B from the LT Perspective

Now consider L_B as a LT likelihood rather than a MR likelihood. How does it relate to conventional LT models? By partitioning $f_{B2\&3}$ appropriately, the link is easy to see. The partitioning is as follows.

$$f_{B2\&3}\{\{n_{ck}\}|n_{..}\} = f_{B2}\{\{n_{.k}\}|n_{..}\} \times f_{B3}\{\{n_{ck}\}|\{n_{.k}\}\} \tag{6.7}$$

where

$$f_{B2}\{\{n_{.k}\}|n_{..}\} = \left(\frac{n_{..}!}{\prod_{k=1}^{K} n_{.k}!}\right) \prod_{k=1}^{K} \left(\frac{P_{.k}}{P_{..}}\right)^{n_{.k}} \tag{6.8}$$

$$f_{B3}\{\{n_{ck}\}|\{n_{.k}\}\} = \prod_{k=1}^{K} \left(\frac{n_{.k}!}{n_{1k}!n_{2k}!n_{3k}!}\right) \prod_{c=1}^{3} \left(\frac{P_{ck}}{P_{.k}}\right)^{n_{ck}} \tag{6.9}$$

Now note that if x were the only explanatory variable in the model, $f_{B1\&2} = f_1 \times f_{B2}$ would be the multinomial model of Burnham *et al.* (1980) for grouped conventional LT data.

$$\begin{aligned} f_{B1\&2}\{\{n_{.k}|N\}\} &= \left(\frac{N!}{n_{..}!(N-n_{..})!}\right) P_{..}^{n_{..}} (1-P_{..})^{N-n_{..}} \left(\frac{n_{..}!}{\prod_{k=1}^{K} n_{.k}!}\right) \prod_{k=1}^{K} \left(\frac{P_{.k}}{P_{..}}\right)^{n_{.k}} \\ &= \left(\frac{N!}{\prod_{k=1}^{K} n_{.k}!(N-n_{..})!}\right) (1-P_{..})^{N-n_{..}} \prod_{k=1}^{K} (P_{.k})^{n_{.k}} \end{aligned} \tag{6.10}$$

where for conventional LT $P_{.k} = p_{.k} = \int_{I_k} g_{.}(x)dx/W$ with $g_{.}(0) = 1$.

As it is, $f_{B1\&2}$ is an extension of the model of Burnham *et al.* (1980) to include other explanatory variables $\boldsymbol{z}$ (and with them, the density $\pi(\boldsymbol{z}|x)$). The partitioning of L_B in this way shows the MRLT model to be a natural extension of conventional LT models with grouped data - with $f_{B1\&2}$ being the conventional LT component of the model, and f_{B3} containing the MR component which allows estimation of the intercepts of the detection functions (i.e. the probability of detection on the trackline).

6.6 Some Illustrative Examples

In general, some assumption(s) need to be made about $\pi(x, \boldsymbol{z})$ in order to be able to estimate the detection probabilities and N from binned data. Still, given suitable forms for $p_i(x, \boldsymbol{z})$ $(i = 1, 2)$ and $\pi(x, \boldsymbol{z})$, estimation is possible *via* numerical minimization of the negative log-likelihood. Some examples with tractable analytic maximum likelihood solutions are given below for illustration. The examples are all univariate. Maximum likelihood estimation in the multivariate case (when $\boldsymbol{z}$ is included) requires $\pi(x|\boldsymbol{z})$ to be specified (either fully or with unknown parameters which can be estimated) and the detection function form as a function of $\pi(x|\boldsymbol{z})$ to be specified. This goes beyond the intent of this subsection, which is to provide simple illustrative examples.

6.6.1 Conventional Two-Sample Mark-Recapture

Assumptions:

1. No heterogeneity (Detection probability does not depend on x or $\boldsymbol{z}$)

This situation arises in IO LT survey if all animals are equally detectable over all perpendicular distances from the trackline. This lack of dependence on perpendicular distances in the detection function is unlikely, but this sort of model was used frequently in the early days of conventional line transect surveys (the resulting estimates are "Kelker strip estimates" - Kelker, 1945), and it has been used for $f(0)$ estimation from LT surveys as recently as 1988 (Buckland, 1988) - although in this case it was used to place bounds on estimates from a less restrictive model rather than as an estimator in its own right. Leaving aside the likely validity of the model in practice for the moment, it is one which is useful for illustration here. In this case the capture history probabilties $P_c(x, \boldsymbol{z})$ $(c = 1, 2, 3)$ do not depend on $(x, \boldsymbol{z})$, and can be called P_c $(c = 1, 2, 3)$ without loss. Similarly n_{ck} $(c = 1, 2, 3;\ k = 1 \ldots K)$ can be called n_c $(c = 1, 2, 3)$ without loss. The binned likelihood L_B reduces to a multinomial likelihood with a single bin. Dropping the bin subscript, k (since there is only one bin), this likelihood is:

$$L_B = \left(\frac{N}{n_1! n_2! n_3! (N - n_\cdot)!}\right) (1 - P_\cdot)^{N - n_\cdot} \prod_{c=1}^{3} (P_c)^{n_c} \tag{6.11}$$

The parameters of interest are the abundance N, and the detection functions p_1 and p_2 (which, because of the complete lack of heterogeneity, are parameters and not functions). Rewriting the likelihood in

these terms (remembering that $P_{\cdot} = p_{\cdot}$) gives

$$L_B = \left(\frac{N}{n_1!n_2!n_3!(N-n_{\cdot})!}\right)(1-p_{\cdot})^{N-n_{\cdot}} p_1^{n_1^+} p_2^{n_2^+} (1-p_1)^{n_2} (1-p_2)^{n_1} \tag{6.12}$$

(Remember that $n_1^+ = n_1 + n_3$ and $n_2^+ = n_2 + n_3$.)

Maximizing this likelihood with respect to the parameters p_1, p_2, and N yields the Petersen estimator of abundance:

$$\hat{N} = \frac{n_1^+ n_2^+}{n_3} \tag{6.13}$$

and the detection probability estimates

$$\hat{p}_1 = \frac{n_3}{n_2^+} \quad \text{and} \quad \hat{p}_2 = \frac{n_3}{n_1^+} \tag{6.14}$$

6.6.2 Two-Sample Mark-Recapture with Detectability Categories

Assumptions:

1. x is discrete ($x \in \{x_1 \ldots x_K\}$).
2. The (unknown) density of x at x_k, $Pr\{x = x_k\}$, is called π_k.
3. Detection probability does not depend on $\boldsymbol{z}$.
4. The detection probability function for the ith observer in the kth x-category, p_{ik},is unrelated to the detection probability in any other category.

In this scenario, animals belong to one of a finite number (K) of detectability categories, and x takes on values denoted $x_1 \ldots x_K$ in bins $1 \ldots K$, respectively. Animals fall into bin k with probability π_k ($\sum_k \pi_k = 1$). It is difficult to imagine how this scenario might arise in LT surveys, except as an approximation to the true situation when only (say) the centre of the perpendicular distance bin in which animals are detected is recorded and no parametric structure is placed on the detection probabilities as functions of perpendicular distance. It might conceivably arise in an aerial LT survey in which (say) clusters of animals are the unit of detection (N is the number of clusters) and the number of animals in a cluster was the only variable determining detectability. (In this case too one might want to assume some relationship between the π_k, but the aim here is illustration, not realism.) Here x is the number of animals in the cluster. The appropriate likelihood is as follows.

$$L_B = \left(\frac{N!}{\prod_c \prod_k n_{ck}!(N-n_{..})!}\right)(1-\sum_k \sum_c P_{ck}\pi_k)^{N-n_{..}} \prod_c \prod_k (P_{ck}\pi_k)^{n_{ck}} \quad (6.15)$$

(Note that in this context P_{ck} is not really a mean - it is the capture history probability at x_k.) The mle for N is:

$$\hat{N} = \sum_k \hat{N}_k \quad (6.16)$$

where $\hat{N}_k$ is the Petersen mark-recapture estimator for $N_k = N\pi_k$ using data from bin k only. The detection probability mle's are as follows.

$$\hat{p}_{1k} = \frac{n_{3k}}{n_{2k}^+} \quad \text{and} \quad \hat{p}_{2k} = \frac{n_{3k}}{n_{1k}^+} \quad (6.17)$$

(The form of the mle's are obvious when one considers that with the π_k unconstrained (except that $\sum \pi_k = 1$), and the P_{ck}'s similarly unconstrained across categories, one really has K independent single mark- recapture experiments.) This model is similar to that of Pollock *et al.* (1984).

6.6.3 Independent Observer Line Transect with homogeneous targets

This situation arises in IO LT surveys where the probability of detecting animals depends only on their perpendicular distance from the trackline, so that all animals at a given perpendicular distance are equally detectable. This is the scenario assumed for conventional LT models, except that the conventional models also assume that detection of an animal on the trackline is certain. The "Product" and "Direct" estimators of "$g(0)$" of Butterworth and Borchers (1988) were developed for the scenario in which detection on the trackline is less than certain and probability of detection depends only on perpendicular distance.

In this case, x is perpendicular distance from the trackline which for estimation purposes takes on values in the interval $(0; W)$, where W is some maximum perpendicular distance considered. Bins correspond to sub-intervals on $(0; W)$. The detection probability $p_i(x)$ for the ith observer would conventionally be written in the form $G_i g_i(x)$, where G_i is the "$g(0)$" parameter and $g_i(0) = 1$.

I consider two sub-cases of this scenario, all of which involve binned data and detection functions which may vary within each bin.

(a) Both $p_i(x)$ and $\pi(x)$ are unconstrained

In this case no structure is placed on the form of the detection functions across the bins and no assumptions are made about $\pi(x)$. This means that the explanatory variable is effectively k, the bin in which an animal falls, and as a result the likelihood is identical to that given in the preceding section. The mle's

are therefore identical to those given immediately above. For future reference, call the mle of abundance $\hat{N}^{(a)}$ (as this is case (a)), i.e.

$$\hat{N}^{(a)} = \sum_k \hat{N}_k \tag{6.18}$$

where $\hat{N}_k$ is the Petersen estimator for data from the kth bin. The corresponding estimator of the probability of at least one observer detecting an animal in the kth bin is

$$\begin{aligned} \hat{P}_{.k}^{(a)} &= \hat{p}_{1k} + \hat{p}_{2k} - \hat{p}_{1k}\hat{p}_{2k} \\ &= \frac{n_{.k}}{\hat{N}_k} \end{aligned} \tag{6.19}$$

(b) $p_i(x)$ unconstrained, $\pi(x)$ known and uniform

Assumptions:

1. x is continuous ($x \in (0; W)$).
2. $\pi(x)$ is **known** to be $1/W$. The probability of x being in the kth bin (of width W_k) is therefore also known and is $\pi_k = W_k/W$.
3. Detection probability does not depend on $\boldsymbol{z}$.
4. No assumptions are made about the form of the detection functions $p_1(x)$ and $p_2(x)$. The mean probability of observer i detecting an animal in bin k is $p_{ik} = \int_{I_k} p_i(x)/W \, dx = \omega_{ik}/W$. Note that without assumptions about the form of $p_i(x)$, only p_{ik} and not $p_i(x)$ $(i = 1, 2)$ can be estimated.

Although no structure is placed on the form of the detection functions across the bins, $\pi(x)$ is known. The model is typical of LT models in that $\pi(x)$ is treated as a known uniform pdf, but LT models would generally place a specific structure on the detection functions $p_i(x)$, parameterizing them as smooth functions of x. Obtaining mle's in closed form when the detection functions are structured in this way is a non-trivial exercise, even for simple plausible forms for the $p_i(x)$'s. Working with completely unconstrained detection functions is useful for illustration, because closed form mle's are easy to obtain.

The binned likelihood in this case is as follows.

$$L_B = \left(\frac{N!}{\prod_c \prod_k n_{ck}! \, (N - n_{..})!}\right) \left(1 - \sum_k \frac{W_k}{W} P_{.k}\right)^{N - n_{..}} \prod_{k=1}^{K} \left(\frac{W_k}{W}\right)^{n_{.k}} \prod_{c=1}^{3} P_{ck}^{n_{ck}} \tag{6.20}$$

Here $P_{ck} = \int_{I_k} P_c(x) dx$.

The mle of abundance is as follows.

$$\hat{N}^{(b)} = \frac{n_{..}}{\sum_k \{(\pi_k n_{.k})/\hat{N}_k\}} \tag{6.21}$$

where $\hat{N}_k$ is the Petersen estimator for the kth interval (see Appendix 6.10.4 for details). The mle's for $p_1(x)$ and $p_2(x)$ are identical to those of sub-case (a) above. The mle for $P_{..}$ is as follows.

$$\hat{P}_{..}^{(b)} = \sum_{k=1}^{K} \frac{\pi_k n_{.k}}{\hat{N}_k} \tag{6.22}$$

Note that the abundance estimator reduces to the estimator of sub-case (a) above when the mle $\hat{\pi}_k = \hat{N}_k / \sum \hat{N}_k$ replaces π_k. This points to an essential difference between the abundance estimators from (a) and (b), namely that (b) uses information about the distribution of the explanatory variable and (a) does not. One would expect this additional information to translate into improved precision in the case of (b). I have not been able to prove that this is (or is not) the case, but considering one particular extreme scenario lends some support to this conjecture and gives insight into the difference between the two forms of estimator.

The scenario is that in which the parameters of the detection function are known. Then conditional on sample size, $n_{..}$, the estimator $\hat{N}^{(b)}$ has zero variance because it depends on the data only through the parameters of the detection function. The estimator $\hat{N}^{(a)}$, on the other hand, varies from sample to sample even when the parameters of the detection function are known. $\hat{N}^{(a)}$ has a component of variance due to the particular explanatory variables which were sampled, $\hat{N}^{(b)}$ does not.

The reason that $\hat{N}^{(b)}$ does not depend on the sampled x's when the detection functions are known, is that information on the distribution of x (i.e. a known $\pi(x)$) is used to integrate x out of the estimator. $\hat{N}^{(a)}$, on the other hand, does not use information on $\pi(x)$. It effectively uses the sampled x's to estimate $\pi(x)$ (see Chapter 7) and this process gives rise to an additional contribution to the variance. If it is generally true that using information on the distribution of the explanatory variable(s) in this way improves precision of the abundance estimator, this has important implications for estimation from MRLT models. The estimators from MR models available at present do not use information about the density of the explanatory variables; most MR applications do not involve such variables.

There is another point about the comparison between the abundance estimators from (a) and (b) which is worth noting. This is that both estimators can be viewed as Horvitz–Thompson estimators which use estimated "inclusion" probabilities in place of known inclusion probabilities. $\hat{N}^{(b)}$ also has the form of conventional perpendicular distance based LT estimators of abundance, namely $n_{..}/\hat{P}_{..}$, where $\hat{P}_{..}$ is an estimator of the average detection probability over all explanatory variables. (In conventional LT theory $\hat{P}_{..} = (\hat{f}(0)W)^{-1}$.) To see that $\hat{N}^{(a)}$ and $\hat{N}^{(b)}$ are Horvitz–Thompson–like estimators, rewrite them as

follows (remembering that $P_{.k}$ is p the probability of one or other of the observers detecting an animal which is in bin k)

$$\hat{N}^{(a)} = \sum_{k=1}^{K} \hat{N}_k = \sum_{k=1}^{K} \frac{n_{.k}}{\hat{P}_{.k}^{(a)}} = \sum_{k=1}^{K} \sum_{j=1}^{n_{.k}} \frac{1}{\hat{P}_{.k}^{(a)}} \tag{6.23}$$

$$\hat{N}^{(b)} = \sum_{k=1}^{K} \sum_{j=1}^{n_{.k}} \frac{1}{\hat{P}_{..}^{(b)}} \tag{6.24}$$

$\hat{N}^{(a)}$ is not unbiased since all of the $\hat{N}_k$ are biased - and in the same direction. (The $\hat{N}_k$'s are just the Petersen estimators of abundance in the kth bin, and Petersen estimators are known to be biased for finite samples.) This illustrates the fact that although the Horvitz–Thompson estimator is unbiased when the detection probabilities are known (see Thompson, 1992, for example), it is not unbiased in general because of this bias when sample size is small.

6.7 The Unbinned Likelihood

The likelihood for N and the parameters of the detection functions and $\pi(x, z)$, given the $n_{..}$ observed capture histories $\{c_j\}$ together with the associated explanatory variables (x_j, z_j) $(j = 1, \ldots, n_{..})$, when explanatory variables drawn from the density $\pi(x, z)$, can be obtained from L_B by taking the limit as the width of the bins approaches zero. (See Appendix 6.10.3.) The resulting likelihood (with subscript "U" for "Unbinned") is as follows.

$$L_U\{N|\{c_j, x_j, z_j\}\} = \kappa \left(\frac{N!}{n_{..}!(N-n_{..})!}\right) (1 - P_{..})^{(N-n_{..})} \prod_{j=1}^{n_{..}} P_{c_j}(x_j, z_j)\, \pi(x_j, z_j) \tag{6.25}$$

Here κ is a constant of proportionality which does not involve the parameters, and $P_{c_j}(x_j, z_j)$ is the probability of observing capture history c_j, given that the animal has attributes (x_j, z_j). (The term $1/n_{..}!$ could be absorbed into the constant of proportionality. However, for consistency with the notation in the case of binned data it is kept outside the constant.) I again omit the parameters of the detection functions and $\pi(x, z)$ as explicit arguments of the likelihood in an attempt to enhance readability.

It is not as easy to produce illustrative examples which are analytically tractable in the case of unbinned data, but a number of specific cases are discussed in the following Chapter. Both specific methods of abundance estimation for these cases and more general methods for MRLT models are considered there.

As in the case of binned data, one can view the model as an extension of a MR model to incorporate observable heterogeneity, or as an extension of conventional LT models to incorporate additional explanatory variables and uncertain detection on the trackline. Again, partitioning the likelihood L_U appropriately makes the form of the extensions clear:

$$L_U\{N|\{c_j,x_j,z_j\}\} = \kappa f_1\{n_..|N\} \times f_{U2}\{\{x_j,z_j\}|n_..\} \times f_{U3}\{\{c_j\}|\{x_j,z_j\}\} \tag{6.26}$$

where:

$$f_1\{n_..|N\} = \frac{N!}{n_..!\,(N-n_..)!} P_..^{n_..}(1-P_..)^{N-n_..} \tag{6.27}$$

$$f_{U2}\{\{x_j,z_j\}|n_..\} = \prod_{j=1}^{n_..} \frac{P_.(x_j,z_j)\,\pi(x_j,z_j)}{P_..} \tag{6.28}$$

$$f_{U3}\{\{c_j\}|\{x_j,z_j\}\} = \prod_{j=1}^{n_..} \frac{P_{c_j}(x_j,z_j)}{P_.(x_j,z_j)} \tag{6.29}$$

(Note that f_1 is the same for the binned and unbinned cases.)

6.7.1 L_U from a MR Perspective

It is easy to see the similarity between L_U and its equivalent for binned data (the MR formulation of equations 6.5 and 6.6) when the unbinned likelihood is written as $f_1 \times f_{U2\&3}$ (where $f_{U2\&3} = f_{U2} \times f_{U3}$):

$$\begin{aligned} L_U\{N|\{c_j,x_j,z_j\}\} &= \kappa f_1\{n_..|N\} \times f_{U2\&3}\{\{c_j,x_j,z_j\}|n_..\} \\ &= \left(\frac{N!}{n_..!(N-n_..)!}\right) P_..^{n_..}(1-P_..)^{N-n_..} \prod_{j=1}^{n_..} \frac{P_{c_j}(x_j,z_j)\,\pi(x_j,z_j)}{P_..} \end{aligned} \tag{6.30}$$

Other than the fact that one likelihood uses binned data and the other unbinned data, the only notable difference between the appearances of the unbinned pdf $f_{U2\&3}$ and the binned pdf $f_{B2\&3}$ is the explicit presence of the density of the explanatory variables, $\pi(x,z)$, in the unbinned pdf. The difference is, however, only in appearance and not in substance. The density $\pi(x,z)$ is present in the binned pdf (the notation is hiding it inside P_{ck}) so that the difference is really just a consequence of the notation. $f_{U2\&3}$ is the analogue of the binned MR pdf $f_{B2\&3}$.

Huggins/Alho Estimation

Huggins (1989) and Alho (1990) independently proposed the same estimator of $P_.(x,z)$ for MR experiments in which detection probabilities vary between animals as a function of observable explanatory

variables. (Obviously perpendicular distance is not useful in their MR context, but the methods remain the same whatever the particular explanatory variables used.) To do this, they assumed that the capture probability on the ith capture occasion, $p_i(x, \boldsymbol{z})$, has a logistic functional form, and parameterized f_{U3} in terms of the parameters of the i logistic functions. The probability of capturing an animal with associated explanatory variables $(x_j, \boldsymbol{z}_j)$ one or more times was then estimated by treating f_{U3} as a likelihood and maximizing it with respect to these parameters and evaluating $P_.(x_j, \boldsymbol{z}_j)$ using the estimated parameters.

Huggins (1989) and Alho (1990) did not derive or use the rest of L_U (i.e. $f_1 \times f_{U2}$) in estimating N. Instead they proposed a Horvitz–Thompson–like estimator (which I cover in the next Chapter).

6.7.2 L_U from the LT Perspective

The product $(f_1 \times f_{U2})$, considered as a likelihood rather than a pdf, has the same form as conventional LT likelihoods. In conventional LT theory the detection function depends on x alone and $p_.(x) = g(x)$, $g(0) = 1$, $\pi(x) = W^{-1}$ and the jth element of f_{U2} is called $f(x_j)$, so that:

$$\begin{aligned} f(x_j) &= \frac{p_.(x_j)\,\frac{1}{W}}{p_{..}} \\ &= \frac{P_.(x_j)\,\frac{1}{W}}{P_{..}} \end{aligned} \tag{6.31}$$

and

$$f_1 \times f_{U2} = \left(\frac{N!}{n_{..}!(N-n_{..})!}\right) P_{..}^{n_{..}} (1-P_{..})^{N-n_{..}} \prod_{j=1}^{n_{..}} f(x_j) \tag{6.32}$$

This is the pdf derived by Seber (1982). The product $\prod f(x_j)$ is the term which is conventionally maximized for estimation of the detection function. As in the case of the binned data, the MRLT likelihood for unbinned data is an extension of the conventional LT likelihood in the following two main respects.

The first is that it incorporates the additional explanatory variable vector $\boldsymbol{z}$, so that $f(x_j)$ generalises to the following.

$$f(x_j, \boldsymbol{z}_j) = \frac{P_.(x_j, \boldsymbol{z}_j)\,\pi(x_j, \boldsymbol{z}_j)}{P_{..}} \tag{6.33}$$

Including $\boldsymbol{z}$ incorporates $\pi(x, \boldsymbol{z})$ as well - which makes modelling the density of the explanatory variables more of an issue than it is in conventional LT theory, and this has implications for the form of the LT estimator of abundance. By randomising the placement of transects in the survey region in some way, one can reasonably assume $\pi(x)$ to be uniform for any one animal. It is often less reasonable to

assume that the x's from different animals are independent; but even when they are not, this does not introduce bias into the point estimate of abundance in the conventional LT context, although variance estimates obtained under the assumption will be biased. Thus one can control $\pi(x)$ to some extent through survey design. This will generally not be the case for $\pi(\boldsymbol{z})$, even if $\boldsymbol{z}$ can reasonably be assumed to be independent of x. For example, group size would often be a component of $\boldsymbol{z}$, and one would at best be able to postulate a distributional form for $\pi(\boldsymbol{z})$ without being able to reasonably assume any particular values for the parameters of this distribution. To estimate abundance using the LT estimator of the usual form $\hat{N} = n_{..}/(\int\int \hat{p}_{.}(x,\boldsymbol{z})\pi(x)\pi(\boldsymbol{z})dx\,d\boldsymbol{z})$, one needs to estimate the parameters of $\pi(\boldsymbol{z})$ in some way. The Horvitz–Thompson approach described in the following Chapter allows abundance to be estimated without assuming a distributional form or particular parameter values for $\pi(\boldsymbol{z})$. (The approach effectively uses a nonparametric estimator of $\pi(\boldsymbol{z})$ - see next Chapter.)

The second extension to conventional LT theory contained in L_U is the addition of the MR component f_{U3}, which allows estimation of the absolute values of the detection functions without having to know or assume values of the detection functions at any point. This contrasts with conventional LT estimators, which rely on knowing the value of the detection function on the trackline.

6.8 Modelling the Densities of Explanatory Variables

The general MRLT models developed here focus attention on the density $\pi(\,)$ of the explanatory variables, which is central to these models. The way this density is handled conventionally in MR and LT models is quite different. MR models generally avoid it altogether, while LT models usually not only incorporate a model for the density, but assume it to be known. (Conventional LT theory abundance estimators depend on $\pi(x)$ being known.) With MRLT models, estimation of abundance is possible without making any assumptions about $\pi(\,)$. ($\hat{N}^{(a)}$ above does this.) But if estimation is to be effected using the full likelihood, then $\pi(\,)$ needs to be modelled in some way. There is also a potential reduction in variance when information on $\pi(\,)$ is used rather than ignored. When perpendicular distance is the only explanatory variable and it is reasonable to assume uniform (if not independent) distribution of animals with respect to the trackline, using this information about $\pi(\,)$ in point estimation of abundance is straightforward. When there are additional explanatory variables, whose distribution would generally be unknown or only vaguely known, the best way to proceed in estimating abundance is less clear. In the next Chapter, I consider a number of methods of estimation which incorporate assumptions of differing strength about the form of $\pi(\,)$.

Before leaving the subject of the density of the explanatory variables, it is worth noting that this density may in itself be of interest. Using the general likelihoods of this Chapter, it is possible to estimate the distribution. Estimation of $\pi(\,)$ is not possible in conventional LT surveys with only one observer. To convince oneself that this is possible in MRLT surveys, consider equation 6.18 above; here an estimator of the density function in the kth bin, π_k, is simply $\hat{\pi}_k = \hat{N}_k/\hat{N}^{(a)}$.) The actual distribution of explanatory

variables (and/or moments of the distribution and/or moments of functions of the explanatory variables) may be of interest in themselves. Having estimated the density, it is possible to estimate the moments of explanatory variables or moments of functions of the explanatory variables. I discuss this estimation in more detail in the following Chapter. Here, for the moment, are some examples.

(1) **Reactive movement**: In MRLT surveys the distribution of animals with respect to the trackline may be of interest insofar as it indicates attraction or avoidance of the survey platform by animals. (An estimated $\pi(x)$ which was peaked at the origin would suggest attraction; one that had least density close to the origin would suggest avoidance.)

(2) **Shore-based surveys**: In shore based surveys of animals which migrate close to shore, animals would generally not be distributed independently with respect to perpendicular distance from shore. (Animals would presumably have some preferred depth/distance from shore.) The shore–based gray whale migration study analysed by Buckland *et al.* (1993b) is a case in point (although on this particular survey, distance from shore out to some truncation distance W was not a statistically significant explanatory variable).

(3) **Group size**: Say that z_1, the first element of $\boldsymbol{z}$ was group size. Then $\pi(z_1)$, the distribution of group sizes in the population, might be of interest in itself, and the mean group size in the population, $\boldsymbol{E}_{z_1}[z_1]$, would almost certainly be of interest. The density $\pi(z_1)$ can be estimated (call it $\hat{\pi}(z_1)$) and can be used to obtain an estimate of mean group size in the population: $\hat{\boldsymbol{E}}_{z_1}[z_1] = \sum_{z_1} z_1 \hat{\pi}(z_1)$.

(4) **Probability of detection on the trackline**: The functions $p_{1.}(x)$, $p_{2.}(x)$ and $p_{..}(x)$ are often of interest, particularly at $x = 0$, when detection on the trackline is not certain. If $\hat{\pi}(x)$ is an estimator of $\pi(x)$, these can be estimated by $\int \hat{p}_i(x, \boldsymbol{z})\, \hat{\pi}(\boldsymbol{z})\, d\boldsymbol{z}$ $(i = 1, 2)$ and $\int \hat{p}_{.}(x, \boldsymbol{z})\, \hat{\pi}(\boldsymbol{z})\, d\boldsymbol{z}$, respectively. Here x is assumed to be independent of $\boldsymbol{z}$ prior to detection.

6.9 Summary

The general binned and unbinned likelihoods developed in this Chapter provide a framework within which almost all LT likelihoods in the literature can readily be located. Among others, they include the conventional univariate case based on perpendicular distance, the conventional multivariate case based on perpendicular distance, and the univariate case with $G < 1$ based on perpendicular distance. The general likelihoods also provide the framework for generalizing these models. The two types of LT model which are not covered by the general likelihoods are variable effort models and cue-based models (including models based on discrete, non-Poisson animal availability).

The likelihoods generalize current LT models in a number of ways, in particular providing models for the incorporation of explanatory variables other than x while simultaneously accommodating detection probabilities on the trackline which are less than unity. They also highlight the central role of the density of the explanatory variables, $\pi(x, \boldsymbol{z})$, in LT models. Estimation of abundance for the general case is

complicated by the presence of $\pi(\boldsymbol{z})$ (assuming independence of x and $\boldsymbol{z}$) in the likelihoods which, unlike $\pi(x)$, cannot reasonably be assumed known in general. Further, while it is possible to estimate abundance using one component of the general likelihoods without making any assumptions about $\pi(\boldsymbol{z})$, the MR nature of MRLT surveys allows $\pi(\boldsymbol{z})$ to be estimated from the MRLT data. (One consequence of this fact is that with MRLT models, $\pi(x)$ need not be assumed known in order to be able to estimate N.) In the next Chapter I discuss estimation methods for the general MRLT case, as well as some special cases.

6.10 Appendices

6.10.1 The effect of ignoring heterogeneity in MRLT models

Notation is the same as that of the body of this Chapter.

In MRLT surveys, each observed datum consists of a perpendicular distance, x, a vector of "additional" explanatory variables, $\boldsymbol{z}$, and a sighting/capture type, c. In the context of LT and MRLT surveys, $\boldsymbol{z}$ is the source of "heterogeneity" which might be ignored. (Remember that $c = 1$ if the animal is detected by observer 1 only, $c = 2$ if the animal is detected by observer 2 only, and $c = 3$ if the animal is detected by both observers.) The pdf of the observed data is as follows.

$$f(c, x, \boldsymbol{z}) = \frac{P_c(x, \boldsymbol{z})\, \pi(\boldsymbol{z}|x)\, \pi(x)}{P_{..}} \tag{6.34}$$

Rewriting this in terms of the detection functions $p_1(x, \boldsymbol{z})$ and $p_2(x, \boldsymbol{z})$ gives the following.

$$\begin{aligned} f(1, x, \boldsymbol{z}) &= \frac{[p_1(x, \boldsymbol{z}) - p_1(x, \boldsymbol{z})p_2(x, \boldsymbol{z})]\, \pi(\boldsymbol{z}|x)\, \pi(x)}{P_{..}} \\ f(2, x, \boldsymbol{z}) &= \frac{[p_2(x, \boldsymbol{z}) - p_1(x, \boldsymbol{z})p_2(x, \boldsymbol{z})]\, \pi(\boldsymbol{z}|x)\, \pi(x)}{P_{..}} \\ f(3, x, \boldsymbol{z}) &= \frac{[p_1(x, \boldsymbol{z})p_2(x, \boldsymbol{z})]\, \pi(\boldsymbol{z}|x)\, \pi(x)}{P_{..}} \end{aligned} \tag{6.35}$$

The marginal pdf for the observed c and x only is therefore given by the following.

$$\begin{aligned} f(1, x) = \int f(1, x, \boldsymbol{z})\, d\boldsymbol{z} &= \frac{[p_{1.}(x) - \int p_1(x, \boldsymbol{z})p_2(x, \boldsymbol{z})\, \pi(\boldsymbol{z}|x)\, d\boldsymbol{z}]\, \pi(x)}{P_{..}} \\ f(2, x) = \int f(2, x, \boldsymbol{z})\, d\boldsymbol{z} &= \frac{[p_{2.}(x) - \int p_1(x, \boldsymbol{z})p_2(x, \boldsymbol{z})\, \pi(\boldsymbol{z}|x)\, d\boldsymbol{z}]\, \pi(x)}{P_{..}} \\ f(3, x) = \int f(3, x, \boldsymbol{z})\, d\boldsymbol{z} &= \frac{[\int p_1(x, \boldsymbol{z})p_2(x, \boldsymbol{z})\, \pi(\boldsymbol{z}|x)\, d\boldsymbol{z}]\, \pi(x)}{P_{..}} \end{aligned} \tag{6.36}$$

(Remember that $p_{i.}(x)$ is the mean detection funtion for observer i, averaged over $\boldsymbol{z}$, that is $p_{i.}(x) = \int p_i(x, \boldsymbol{z})\, \pi(\boldsymbol{z}|x)\, d\boldsymbol{z}$.) If detection probability did not depend on $\boldsymbol{z}$, the pdf for the observed c and x would be as follows.

$$\begin{aligned} f(1, x) &= \frac{[p_{1.}(x) - p_{1.}(x)p_{2.}(x)]\, \pi(x)}{P_{..}} \\ f(2, x) &= \frac{[p_{2.}(x) - p_{1.}(x)p_{2.}(x)]\, \pi(x)}{P_{..}} \\ f(3, x) &= \frac{[p_{1.}(x)p_{2.}(x)]\, \pi(x)}{P_{..}} \end{aligned} \tag{6.37}$$

In general, however

$$p_{1.}(x)p_{2.}(x) \neq \int p_1(x,z)p_2(x,z)\,\pi(z|x)\,dz \tag{6.38}$$

So, unlike the conventional LT case, the data observed are not the same when the detection functions depend on z as when they are independent of z, and equal to the average value of the detection functions which depend on z. Accordingly, even if the true form of the $p_{i.}(x)$ were known, inference which is unbiased when detection probability does not depend on z will in general be biased if the detection probability actually depends on z and inferences are made on the basis of $p_{i.}(x)$ rather than $p_i(x,z)$.

Equation 6.38 is critical; it says that when detection probability depends on attributes other than x, detections by observers 1 and 2 at any given x are not independent, even though the observers act independently of one another. Variation in z induces covariance between the detection probabilites of the two observers, as is apparent from the following.

$$\begin{aligned} p_{3.}(x) &= \int p_1(x,z)p_2(x,z)\,\pi(z|x)\,dz \\ &= E_z[p_1(x,z)p_2(x,z)] \\ &= Cov_z[p_1(x,z)p_2(x,z)] + E_z[p_1(x,z)]\,E_z[p_2(x,z)] \\ &= \sigma_{12}(x) + p_{1.}(x)p_{2.}(x) \end{aligned} \tag{6.39}$$

where $\sigma_{12}(x) = Cov_z[p_1(x,z)p_2(x,z)]$.

If both $p_1(x,z)$ and $p_2(x,z)$ are either monotonically increasing or monotonically decreasing functions of z then $\sigma_{12}(x) > 0$. In this case animals seen by observer 1 will also tend to be seen by observer 2, because these animals will tend to be those with z's which make them more detectable to both observers. If $p_1(x,z)$ is increasing in z where $p_2(x,z)$ is decreasing in z then $\sigma_{12}(x) < 0$. In this case, animals which observer 1 sees will tend to be missed by observer 2.

If inferences are based on the assumption that detections are independent when they are not, such inferences will in general be biased. So in a MR and a MRLT scenario, z cannot be ignored with impunity. (Note, by the way, that the same applies to x; were one to make inferences from a MRLT survey ignoring x when detection probabilities did in fact depend on x, the inferences would in general be biased.)

The Effect on the Shape of the Duplicate Detection Function

In the previous Chapter I noted that unmodelled heterogeneity could lead to a difference between the shape of the observed perpendicular distance distribution of duplicates and that predicted assuming independent detections. In this section I look very briefly at the diagnostic value of the difference; what does it say about the form of the heterogeneity?

Equation 6.39 can be rewritten as follows.

$$p_{3.}(x) = \delta(x) \times (p_{1.}(x)p_{2.}(x)) \tag{6.40}$$

where

$$\delta(x) = 1 + \frac{\sigma_{12}(x)}{p_{1.}(x)p_{2.}(x)} \tag{6.41}$$

The function $\delta(x)$ is then the proportion by which the unmodelled heterogeneity (z) changes the duplicate detection function from what it would have been at x were there no heterogeneity. The difference in the shape between the observed and predicted perpendicular distance distributions of duplicates (i.e. between $f_{3.}(x) = p_{3.}(x)/\int p_{3.}(x)\,dx$ and $d_{3.}(x) = p_{1.}(x)p_{2.}(x)/\int p_{1.}(x)p_{2.}(x)\,dx$) is contained in $\delta(x)$. If $\delta(x)$ does not vary with x (i.e. if $\sigma_{12}(x)\ \alpha\ p_{1.}(x)p_{2.}(x)$) there will be no difference in the shapes. In principle, there could be strong positive or negative correlation and no difference in the predicted and observed shapes so that a lack of difference does not necessarily imply independence. On the other hand, if there is no correlation there will be no difference in shape, so that a difference in shape does imply a lack of independence (providing none of the other factors mentioned in the previous Chapter are acting to cause the difference).

The Constant-G Case

In the unlikely event that the probability of detection on the trackline, G, is constant over all values of z, the detection function for observer i can be written as $p_i(x,z) = G_i \times g_i(x,z)$, where $g_i(0,z) = 1$. In this case, the RHS of inequality 6.38 can be written as follows.

$$\begin{aligned}\int p_1(x,z)p_2(x,z)\,\pi(z\,|\,x)\,dz &= G_1G_2 \times \int g_1(x,z)g_2(x,z)\,\pi(z\,|\,x)\,dz \\ &= G_1G_2 \times \int g_3(x,z)\,\pi(z\,|\,x)\,dz \\ &= G_1G_2 \times g_{3.}(x)\end{aligned} \tag{6.42}$$

Here $g_3(x,z) = g_1(x,z)g_2(x,z)$. Because $g_{3.}(0) = 1$, the theorem of Patil *et al.* (1993) can be invoked. Thus, providing the correct functional form is used for $g_{3.}(x)$, no bias is introduced by ignoring z and estimating $g_{3.}(x)$ from the observed x's of **duplicate detections** only. In this, case one can therefore estimate the RHS of the inequality 6.38 without any knowlege of, or assumptions about $\pi(z\,|\,x)$. As a result estimation need not be based on the assumption that the inequality 6.38 is an equality; it can instead be based on the equations 6.36. Note that G_1G_2 appears in both the numerator and the denominator of the last of the three equations 6.36, so that it cancels and makes $f(3,x)$ (the pdf of observed **duplicate detection** perpendicular distances) independent of G_1 and G_2.

The MP and BT method estimators estimate $g_{3.}(x)$ directly from the distribution of observed **duplicate detection** perpendicular distances. They do not rely on the assumption that the inequality 6.38 is an equality, and inference is therefore not biased as a result of ignoring z. The P method estimator, on the other hand, does not use the distribution of observed **duplicate detection** perpendicular distances. It is based on the assumption that the inequality 6.38 is an equality, and inferences from the P method will therefore generally be biased when the detection probabilities depend on z.

6.10.2 Derivation of the Likelihood for Binned Observations

This derivation is based on a similar derivation by Zuchini (*pers. commn*), for the case in which there is only one observer. The notation is the same as that in the body of this Chapter.

For brevity I write $(x, \boldsymbol{z})$ as $\boldsymbol{x}$. The full likelihood when observations are binned into K mutually exclusive bins on the basis of the observed covariates $(\boldsymbol{x})$ can be derived as the product of two likelihoods as follows.

Assuming that animals are distributed independently in the plane, the probability of the N animals in the covered region being distributed so that there are N_k $(k = 1, \ldots, K)$ animals in the k'th bin is multinomial, as follows.

$$f\{\{N_k\}|N\} = N! \prod_{k=1}^{K} \left(\frac{\pi_k^{N_k}}{N_k!} \right) \tag{6.43}$$

Here π_k is the probability that the animal falls in the kth bin, and $\{N_k\}$ is the vector of K N_k's. The following notation is used (in addition to the notation of the body of this Chapter) in this derivation.

> $\int_{I_k} d\boldsymbol{x}_j$ is the integral with respect to $\boldsymbol{x}$ (which in general is a vector) over the range of the kth interval. The subscript is for the jth animal. The pdf of $\boldsymbol{x}$ is assumed to be the same for all animals.)

In addition, the following two summations

$$\sum_{\sum n_{cj}=n_{.k}} \quad \text{and} \quad \sum_{\sum N_k=N} \tag{6.44}$$

mean "the sum over the set of all n_{cj}'s in bin k whose sum is equal to $n_{.k}$", and "the sum over the set of all N_k's whose sum is equal to N", respectively.

Given the distribution of the animals in the bins, the probability of observing n_{ck} animals in the kth bin $(k = 1, \ldots, K)$ with capture history c $(c = 1, 2, 3)$ is derived as follows. (Strictly $\boldsymbol{x}_j$ should have another subscript, k, in the derivation below. I have omitted it in an attempt to make the algebra a little less messy, but the reader should keep in mind that $\boldsymbol{x}_j$ in bin k is different from $\boldsymbol{x}_j$ in any other bin.)

$$\begin{aligned} f\{\{n_{ck}\}|\{N_k\}\} &= \prod_{k=1}^{K} f\{\{n_{1k}\}, \{n_{2k}\}, \{n_{3k}\}|\{N_k\}\} \\ &= \prod_{k=1}^{K} \int_{I_k} \cdots \int_{I_k} \left\{ \sum_{\sum n_{cj}=n_{.k}} \prod_{j=1}^{N_k} [1 - P_{\cdot}(\boldsymbol{x}_j)]^{1-n_{\cdot j}} \prod_{c=1}^{3} P_c(\boldsymbol{x}_j)^{n_{cj}} \left(\frac{\pi(\boldsymbol{x}_j)}{\pi_k} \right) \right\} d\boldsymbol{x}_{N_k} \cdots d\boldsymbol{x}_1 \end{aligned}$$

$$
\begin{aligned}
&= \prod_{k=1}^{K} \sum_{\sum n_{cj}=n_{.k}} \prod_{j=1}^{N_k} \left[\int_{I_k} \cdots \int_{I_k} [1-P_{\cdot}(\boldsymbol{x}_j)]^{1-n_{.j}} \prod_{c=1}^{3} P_c(\boldsymbol{x}_j)^{n_{cj}} \left(\frac{\pi(\boldsymbol{x}_j)}{\pi_k} \right) d\boldsymbol{x}_{N_k} \cdots d\boldsymbol{x}_1 \right] \\
&= \prod_{k=1}^{K} \sum_{\sum n_{cj}=n_{.k}} \prod_{j=1}^{N_k} \left(\frac{\pi_k - P_{.k}}{\pi_k} \right)^{N_k - n_{.k}} \prod_{c=1}^{3} \left(\frac{P_{cj}}{\pi_k} \right)^{n_{cj}} \\
&= \prod_{k=1}^{K} \left(\frac{N_k!}{\prod_{c=1}^{3} n_{ck}!(N_k - n_{.k})!} \right) \left(\frac{\pi_k - P_{.k}}{\pi_k} \right)^{N_k - n_{.k}} \prod_{c=1}^{3} \left(\frac{P_{ck}}{\pi_k} \right)^{n_{ck}} \qquad (6.45)
\end{aligned}
$$

Combining this with equation 6.43 we get the following likelihood.

$$
\begin{aligned}
L_B\{N|\{n_{ck}\}\} &= \sum_{\Sigma N_k = N} f\{\{n_{ck}\}|\{N_k\}\} \times f\{\{N_k\}|N\} \\
&= N! \sum_{\Sigma N_k = N} \prod_{k=1}^{K} \left(\frac{N_k!}{\prod_{c=1}^{3} n_{ck}!(N_k - n_{.k})!} \right) \prod_{c=1}^{3} \left(\frac{P_{ck}}{\pi_k} \right)^{n_{ck}} \left(\frac{\pi_k - P_{.k}}{\pi_k} \right)^{N_k - n_{.k}} \left(\frac{\pi_k^{N_k}}{N_k!} \right) \\
&= N! \sum_{\Sigma N_k = N} \prod_{k=1}^{K} \left(\frac{(\pi_k - P_{.k})^{N_k - n_{.k}}}{(N_k - n_{.k})!} \right) \prod_{c=1}^{3} \left(\frac{P_{ck}^{n_{ck}}}{n_{ck}!} \right) \\
&= N! \left[\sum_{\Sigma N_k = N} \prod_{k=1}^{K} \frac{(\pi_k - P_{.k})^{N_k - n_{.k}}}{(N_k - n_{.k})!} \right] \prod_{k=1}^{K} \prod_{c=1}^{3} \left(\frac{P_{ck}^{n_{ck}}}{n_{ck}!} \right) \\
&= N! \left[\frac{\left[\sum_{k=1}^{K} (\pi_k - P_{.k}) \right]^{N_k - n_{.k}}}{(N - n_{\cdot\cdot})!} \right] \prod_{k=1}^{K} \prod_{c=1}^{3} \left(\frac{P_{ck}^{n_{ck}}}{n_{ck}!} \right) \\
&= \left(\frac{N!}{\prod_c \prod_k n_{ck}!(N - n_{\cdot\cdot})!} \right) (1 - P_{\cdot\cdot})^{N - n_{\cdot\cdot}} \prod_{c=1}^{3} \prod_{k=1}^{K} (P_{ck})^{n_{ck}} \qquad (6.46)
\end{aligned}
$$

The multinomial expansion of $[\sum_k (\pi_k - P_{.k})]^{N_k - n_{.k}}$ is used in going from the third-last to the second-last lines, together with the following.

$$
\sum_{\sum N_k = N} = \sum_{\sum (N_k - n_{.k}) = (N - n_{\cdot\cdot})}
$$

6.10.3 Derivation of the Likelihood for Unbinned Observations

Divide the domain of $\boldsymbol{x}$ into K bins of equal size, $\Delta\boldsymbol{x}(K)$. Now let $K \to \infty$ (and $\Delta\boldsymbol{x}(K) \to 0$). The $n_{ck}!$'s in the denominator of equation 6.46 all approach unity and can be absorbed into a constant of proportionality. The limit of the product term can be written as follows.

$$\lim_{\Delta\boldsymbol{x}(K)\to 0} \left\{ \prod_{c=1}^{3} \prod_{k=1}^{K} \Delta\boldsymbol{x}(K) \left(\bar{P}_{ck}\right)^{n_{ck}} \right\} \tag{6.47}$$

where $\bar{P}_{ck}$ is the mean value of $P_c(\boldsymbol{x})\pi(\boldsymbol{x})$ in the kth interval. Absorb $\Delta\boldsymbol{x}(K)$ into the constant of proportionality (call the constant κ) and in the limit as $\Delta\boldsymbol{x}(K) \to 0$, $(\bar{P}_{ck})^{n_{ck}}$ approaches $[P_c(\boldsymbol{x})\pi(\boldsymbol{x})]^1$ at the $n_{..}$ observed $\boldsymbol{x}_j$'s ($j = 1 \cdots n_{..}$) and approaches $[P_c(\boldsymbol{x})\pi(\boldsymbol{x})]^0 = 1$ elsewhere. The likelihood thus becomes

$$L_B\{\{\boldsymbol{x}_j\}|N\} \;=\; \kappa \left(\frac{N!}{(N-n_{..})!} \right) (1 - P_{..})^{N-n_{..}} \prod_{c=1}^{3} \prod_{j=1}^{n_{..}} P_c(\boldsymbol{x}_j)\pi(\boldsymbol{x}_j) \tag{6.48}$$

6.10.4 MLE's for Binned Data with constant detectability in each bin, known π_k's , and no structure on p_{ik}

Note: Following Seber (1982), the following useful approximation for the derivative of $log(N!)$, where N is an integer, is used.

$$\frac{dlog(N!)}{dN} = log(N) \tag{6.49}$$

The likelihood to be maximized (written in terms of the detection probabilities for observers 1 and 2) is as follows.

$$L_B = \left(\frac{N!}{\prod_i \prod_k n_{ik}!(N-n_{..})!}\right)(1-\sum_k p_{.k}\pi_k)^{N-n_{..}} \prod_k \pi_k^{n_{.k}} \prod_{i=1}^{2} p_{ik}^{n_{ik}^+} (1-p_{(3-i)k})^{n_{ik}} \tag{6.50}$$

(Recall that n_{ik} is the number of animals in the kth bin which were detected by observer i **only**.) Differentiating the log of this likelihood with respect to N, p_{1k} and p_{2k}, and setting the resulting equations to zero, gives the following set of equations (the second of which represents $2k$ equations - one for each k, for each of $i = 1, 2$).

$$1 - \frac{n_{..}}{N} = 1 - \sum_{k=1}^{K} p_{.k}\pi_k \tag{6.51}$$

$$\frac{n_{ik}^+}{p_{ik}} - \frac{n_{(3-i)k}^+ - n_{3k}}{(1-p_{ik})} = \frac{(N-n_{..})(1-p_{(3-i)k})}{1-\sum_k p_{.k}\pi_k} \tag{6.52}$$

After some algebraic manipulation of the second of these equations, the following mle for p_{ik} is obtained.

$$\hat{p}_{ik} = \frac{n_{3k}}{n_{(3-i)k}^+} \tag{6.53}$$

So

$$\hat{p}_{.k} = \left(\frac{n_{3k}}{n_{1k}^+ n_{2k}^+}\right) n_{.k} = \frac{n_{.k}}{\hat{N}_k} \tag{6.54}$$

Here $\hat{N}_k$ is the Peterson estimator for the kth interval ($\hat{N}_k = \frac{n_{1k}^+ n_{2k}^+}{n_{3k}}$). From equation 6.51 above it follows then that

$$\hat{N} = n_{..}\left(\sum_k \pi_k \frac{n_{.k}^+}{\hat{N}_k}\right)^{-1} \tag{6.55}$$

Chapter 7

Estimators for MRLT Models

7.1 Overview

This Chapter covers methods of estimating abundance from the general MRLT models developed in the previous Chapter. In the first section, I place some existing estimators in the MRLT model context, and develop generalized forms of some of these. In the second section, I discuss maximum likelihood estimation from the full likelihood and from conditional likelihoods. A central theme of the section is the way in which abundance estimation can be separated conveniently into two components: (a) estimation of detection probabilities, and (b) estimation of abundance conditional on these estimates of detection probability. I consider a number of estimators of detection probabilities, but concentrate on developing estimators for abundance in the surveyed area, conditional on estimates of detection probability. Three sorts of abundance estimators are considered and/or developed in the third section:

1. A Horvitz–Thompson–like estimator, similar to that of Huggins (1989) and Alho (1990). A distinguishing feature of this sort of estimator is that no assumptions are made about the distribution of the explanatory variables, x and z.

2. A Horvitz–Thompson–like estimator which includes the conventional LT assumption of uniform distribution for x with respect to the trackline. (This estimator provides a generalization of the conditional approach of Chapter 4.)

3. An estimator which incorporates an assumption of a parametric form for the distribution of detection probabilities (as opposed to explanatory variables) in the population.

The asymptotic properties of the first estimator were obtained by Huggins (1989) and Alho (1990). The asymptotic properties of the second are obtained easily by the same means. These properties for the last estimator are more difficult to derive, and have not been derived here. Obtaining the properties of all three estimators for finite samples is not analytically tractable. Bootstrap estimators of variance and confidence intervals are proposed.

In the final sections, I consider briefly how the problems of animal movement and animal clustering/grouping relate to each of the estimators, before summarizing the main developments of the Chapter in the conclusion.

7.2 Existing LT Estimators and some Generalizations

7.2.1 Univariate Conventional Models

Model Assumptions:

1. Detection probability is a function of x alone.
2. Detection on the trackline is certain. (For consistency with the notation of the previous Chapter, the detection function for the single observer is written $p_.(x) = g_.(x)$ (with $p_.(0) = g_.(0) = 1$).)
3. Animals are distibuted uniformly with respect to the trackline.

With certain detection on the trackline, the MR components (f_{B3} and f_{U3}) of the general likelihoods (L_B and L_U) are equal to 1 and they do not enter the likelihood. In addition detection probabilities are assumed to depend only on x, so that $\boldsymbol{z}$ does not appear in the likelihood.

Conventional LT estimators of N in the literature have the form $n_../\hat{p}_..$. Abundance estimation is usually performed in two steps. First $p_.(x)$ is estimated, then N is estimated conditional on $\hat{p}_.(x)$. Maximum likelihood estimation of the detection function $p_.(x)$ is effected by maximising f_{U2} or f_{B2} with respect to the paramters of $p_.(x)$. (The density $\pi(x)$ is assumed to be known in order to do this.) Once $p_.(x)$ has been estimated, the number of animals in the area within W of the trackline (N) is estimated as follows (using the notation of the previous Chapter):

$$\begin{aligned} \hat{N}_C &= \frac{n_..}{\frac{1}{W}\int_0^W \hat{p}_.(x)\,dx} \\ &= \frac{n_..}{\hat{p}_..} \end{aligned} \tag{7.1}$$

(The C in the subsctipt is for "Conventional".)

A Horvitz–Thompson–like Estimator

A Horvitz–Thompson–like estimator for the conventional, univariate model is:

$$\hat{N}_H = \sum_{j=1}^{n_..} \frac{1}{\hat{p}_.(x_j)} \tag{7.2}$$

where $\hat{p}_.(x_j)$ is the estimated detection function evaluated at the jth observed perpendicular distance. Were $p_.(x)$ known, the estimator would be a Horvitz–Thompson estimator. $\hat{N}_C$ can be shown to be more efficient than $\hat{N}_H$ when $\pi(x) = 1/W$.

To give a simple example, say for illustration purposes that the detection function was negative exponential:

$$p_.(x) = e^{-\lambda x}$$

Then, letting the cutpoint W approach ∞ and using the mle for λ ($\hat{\lambda} = \overline{x}^{-1}$), the conventional and Horvitz–Thompson–like estimators for N would be as follows.

$$\hat{N}_C = \frac{n_..}{\overline{x}} \qquad \text{and} \qquad \hat{N}_H = \sum_{j=1}^{n_..} \frac{1}{e^{-x_j/\overline{x}}} \tag{7.3}$$

In the conventional univariate LT model case, both types of estimator use knowledge of $\pi(x)$ in estimating N. (Both estimators rely on knowledge of $\pi(x)$ to make estimation of $p_.(x)$ possible; $\hat{N}_C$ also uses $\pi(x)$ in estimating $p_..$) For more general LT models, the Horvitz–Thompson–like estimator is based on fewer assumptions about the density of explanatory variables (see below).

7.2.2 Multivariate Conventional Models

Model Assumptions:

The critical assumptions of the model are as follows.

1. Detection probability is a function of x and $\boldsymbol{z}$.
2. Detection on the trackline is certain ($p_.(0, \boldsymbol{z}) = 1 \; \forall \boldsymbol{z}$).
3. Animals are distibuted uniformly with respect to the trackline.

Drummer and McDonald (1987) and Quang (1991) considered special cases of this sort of model.

As in the case above, with certain detection on the trackline, f_{B3} and f_{U3} are equal to 1 and so effectively disappear from the general likelihoods L_B and L_U. The estimator of the conditional approach of Chapter 4 has the form

$$\begin{aligned} \hat{N} = n_.. W \hat{f}(0) = \; &= n_.. W \left(\frac{1}{n_..} \sum_{j=1}^{n_..} \hat{f}(0|\boldsymbol{z}_j) \right) \\ &= n_.. W \left(\frac{1}{n_..} \sum_{j=1}^{n_..} \frac{1}{\hat{\omega}_..(\boldsymbol{z}_j)} \right) \end{aligned}$$

$$= \sum_{j=1}^{n_{..}} \frac{1}{\hat{p}_{..}(z_j)} \tag{7.4}$$

$(\hat{p}_{..}(\boldsymbol{z}) = \int p_{.}(x,\boldsymbol{z})\frac{1}{W}dx = \hat{\omega}(\boldsymbol{z})/W)$

The estimator is a Horvitz–Thompson–like estimator, but it is similar to conventional LT estimators in the literature in one respect. This is that it is based on the assumption that $\pi(x)$ is known and uniform, and this assumption is used in estimating $\omega(\boldsymbol{z})$ (i.e. in getting $\hat{p}_{..}(\boldsymbol{z})$ from $\hat{p}_{.}(x,\boldsymbol{z})$). The estimator involves no assumptions about $\pi(\boldsymbol{z})$. Without some such assumptions, it is not possible to use an estimator of the form $n_{..} / \int\int \hat{p}_{.}(x,\boldsymbol{z})\pi(x,\boldsymbol{z})\,dx\,d\boldsymbol{z}$, because $\pi(x,\boldsymbol{z})$ is not known.

Note that estimation relies on the assumption that $p_{.}(0,\boldsymbol{z}) = 1$ for estimation of $p_{..}(\boldsymbol{z})$. Without this assumption the perpendicular distance data for any $\boldsymbol{z}$ contain no information on the absolute value of the detection function at any point.

7.2.3 Constant-Shape, Constant-G, Multivariate MRLT Models

Model Assumptions:

Estimators are developed for the case in which the probability of detection on the trackline is less than certain but in which the "additional" explanatory variables, $\boldsymbol{z}$, affect only the scale parameter of the detection functions. In the terminology of the previous Chapter, it is "constant-shape, constant-G, multivariate MRLT Models" that are being considered. The key assumptions are as follows.

1. Detection on the trackline is not certain.
2. The probability of detection on the trackline is constant for each observer. (Call this probability G_i for observer i.)
3. The shape of the detection function does not depend on $\boldsymbol{z}$, i.e., the effect of $\boldsymbol{z}$ on the detection function is only to shrink or stretch the scale of the x-axis.
4. Animals are distibuted uniformly with respect to the trackline.

The model is similar to the conventional multivariate models of Chapter 4, except that detection on the trackline is not certain. Were detection on the trackline certain, the model would be identical to the model of the preceding section. Aside from cue-based models, models in the LT literature which have uncertain detection on the trackline are limited to the case in which detection probability depends only on x (and not on $\boldsymbol{z}$). Two classes of estimator for the probabilities of detection on the trackline exist for this situation. Generalizing the first leads to a generalization of the Petersen estimator of abundance. This is the class I cover in this section. I call them "Generalized Petersen estimators" (GP estimators). The P and MP method estimators fall in this class, as does the BT estimator when animals are stationary.

The D method estimator is the other estimator proposed for the scenario in which detection depends only on x and detection on the trackline is not certain. For univariate models, it provides the equivalent for binned data of the estimator of Huggins/Alho for unbinned data.

A central feature of GP estimators is that because the shape of the detection functions are independent of z, one can **estimate the shapes of the detection functions using conventional LT methods.** Given these estimates, one can obtain mle's of detection probabilities and of N using the general likelihoods of the previous Chapter. (The estimators were not originally derived in this way, but it turns out that they fit neatly into the general likelihood framework developed in the previous Chapter.)

To see this, it is convenient to partition L_U and L_B as follows. (I write the likelihoods as functions of both x and z here for generality, but maintain crucial the assumption that detection on the trackline is not a function of z. In the case where the detection functions depend only on x, the z simply falls away.)

$$L_U = f_1\{n_{..}|N\} \times f_{2P}\{\{n_{c.}\}|n_{..}\} \times f_{U3P}\{\{(x_j, z_j)\}|\{n_{c.}\}\} \tag{7.5}$$

Here f_1 is as in the previous Chapter, and

$$f_{2P}\{\{n_{c.}\}|n_{..}\} = \left(\frac{n_{..}!}{n_{1.}!n_{2.}!n_{3.}!}\right)\prod_{c=1}^{3}\left(\frac{P_{c.}}{P_{..}}\right)^{n_{c.}} \tag{7.6}$$

$$f_{U3P}\{\{(x_j, z_j)\}|\{n_{c.}\}\} = \prod_{c=1}^{3}\prod_{j=1}^{n_{c.}}\frac{P_c(x_j, z_j)\pi(x_j, z_j)}{P_{c.}} \tag{7.7}$$

The likelilhood for binned data (f_B) is identical, except that f_{U3P} is replaced by f_{B3P} where

$$f_{B3P}\{\{n_{cj}\}|\{n_{c.}\}\} = \prod_{c=1}^{3}\left(\frac{n_{c.}!}{n_{c1}\ldots n_{ck}!}\right)\prod_{k=1}^{K}\left(\frac{P_{ck}}{P_{c.}}\right)^{n_{ck}} \tag{7.8}$$

(I use the additional subscript P on L to distinguish the partitioning from that of the previous Chapter, and to associate the pdf with GP estimators.) f_{2P} is the probability of observing $n_{1.}$, $n_{2.}$ and $n_{3.}$ animals with capture histories $c = 1, 2$ and 3, respectively, given that $n_{..}$ animals were observed in all. f_{U3P} is the probability of observing the explanatory variables $\{(x_j, z_j)\}$, given that $n_{1.}$, $n_{2.}$ and $n_{3.}$ animals were observed with capture histories $c = 1, 2$ and 3, respectively. f_{B3P} is the equivalent for binned data.

Without placing some constraints on the relationships between the capture history probabilities (i.e. between the $P_c(x, z)$'s $i = 1, 2, 3$), there is no information in f_{U3P} or f_{B3P} about the intercepts of the $P_c(x, z)$'s. (The pdf's are unchanged by multiplying the $P_c(x, z)$'s by constants.) In this case, f_{U3P} and f_{B3P} contain information about the **shape** of the $P_c(x, z)$'s only. Let $p_i(x, z) = G_i g_i(x, z)$ $(i = 1, 2)$, with

$g_i(0, \boldsymbol{z}) = 1$ $(i = 1, 2)$ - i.e. the probabilies of detection on the trackline are G_1 and G_2 for observers 1 and 2, respectively, irrespective of the values of the other explanatory variables, $\boldsymbol{z}$. A conditional approach to estimation of the G_i's and N would be as follows.

1. Estimate the shape of the $P_c(x, \boldsymbol{z})$'s from f_{U3_P} or f_{B3_P}.
2. Estimate the G_i's (and hence the $P_c(x, \boldsymbol{z})$'s) from f_{2_P}, conditional on the estimated shape parameters
3. Estimate N from f_1, conditional on the $P_c(x, \boldsymbol{z})$'s.

The P and MP method estimators involve an estimation procedure similar to, but not quite the same as this. This estimator was derived in the context of univariate MRLT models (in which x is the only explanatory variable), but it can readily be generalized to include the multivariate case. Here I deal with the multivariate model, of which the univariate model is a special case. It is the shapes of the detection functions $p_i(x, \boldsymbol{z})$ $(i = 1, 2)$, rather than the shapes of $P_c(x, \boldsymbol{z})$ $(c = 1, 2, 3)$, that are estimated in the first step. Estimates of the shapes of the detection functions are obtained by conventional LT methods. This approach can take advantage of the modelling and estimation methods developed for conventional univariate LT theory. In f_{2_P}, information on the shapes of the detection functions is summarized in the effective strip widths $\hat{\omega}_c$ $(c = 1, 2, 3)$. To estimate the G_i's $(i = 1, 2)$, f_{2_P} is maximised with respect to the G_i's, conditional on the estimated effective strip widths.

Write the average effective strip width for observer i (averaged over $\boldsymbol{z}$) as ω_i (where ω_i is defined as $\omega_i = \int\int g_i(x, \boldsymbol{z})\pi(x, \boldsymbol{z})\, dx d\boldsymbol{z}$), and let $\hat{\omega}_i$ be an estimator of ω_i. Note that ω_i involves $\pi(\boldsymbol{z})$ and this will generally not be known. This presents a difficulty, but not an insurmountable one. In Appendix 7.10.2, I develop an estimator of ω_i for the multivariate constant-G case which does not require any knowledge of $\pi(\boldsymbol{z})$. Assume for the moment that estimates of ω_i are available. Maximizing f_{2_P} with respect to G_1 and G_2 conditional on these estimates leads to the following estimators of G_i (see Appendix 7.10.1):

$$\hat{G}_1 = \frac{n_{3.}}{n_{2.}^{+}} \left(\frac{\hat{\omega}_2}{\hat{\omega}_3} \right) \tag{7.9}$$

$$\hat{G}_2 = \frac{n_{3.}}{n_{1.}^{+}} \left(\frac{\hat{\omega}_1}{\hat{\omega}_3} \right) \tag{7.10}$$

The P method estimators of G_i $(i = 1, 2)$ are obtained as a special case when: (1) detection probability depends only on perpendicular distance, x; and (2) ω_3 is estimated as follows.

$$\hat{\omega}_3 = \int_0^W \hat{g}_1(x)\hat{g}_2(x)\, dx \tag{7.11}$$

Here $\hat{g}_i(x)$ is the estimated detection function for the ith observer - estimated by conventional LT methods using the perpendicular distance distribution of the ith observer's detections and the assumption that $g_i(0) = 1$.

The MP and BT methods estimate ω_3 slightly differently, as follows.

$$\hat{\omega}_3 = \int_0^W \hat{g}_3(x)\,dx \tag{7.12}$$

Here $\hat{g}_3(x)$ is the estimated detection function for observed perpendicular distances for the detections by both observers - estimated by conventional LT methods using the perpendicular distance distribution of **duplicate detections** and the assumption that $g_3(0) = 1$. Note that, unlike the P method estimator, this estimator depends on having sufficient duplicate detections for $f_3(x)$ to be estimated reliably. The P method estimator effectively substitutes this requirement with a stronger assumption about the independence of detections. The cost of the stronger assumption is a loss of robustness, although it generally leads to improved precision; MP and BT method estimators remain consistent when detection probability depends on $\boldsymbol{z}$, but the P method estimator does not (see Appendix 6.10.1 of the previous Chapter).

A Generalization of the Petersen Estimator

Maximizing f_1 with respect to N, conditional on the estimates of G_1, G_2, ω_1, ω_2, and ω_3, results in the following estimator of average combined probability of detection by either observer (recall that W is the maximum perpendicular distance used in estimation):

$$\hat{p}_{..} = \frac{n_{..}}{\hat{N}_P}\left(\frac{\hat{\omega}_1\hat{\omega}_2}{W\hat{\omega}_3}\right) \tag{7.13}$$

($\hat{N}_P$ is the Petersen estimator of abundance.) The corresponding mle for N is:

$$\begin{aligned}\hat{N}_{GP} &= \frac{n_{..}}{\hat{p}_{..}} \\ &= \hat{N}_P\left(\frac{W\hat{\omega}_3}{\hat{\omega}_1\hat{\omega}_2}\right)\end{aligned} \tag{7.14}$$

This is a generalization of the Petersen single mark-recapture estimator. The generalization incorporates an adjustment for the lack of independence between detections as a result of the variation in $(x, \boldsymbol{z})$. If detections were unconditionally independent (as opposed to conditionally independent, given $(x, \boldsymbol{z})$, then $p_{3.} = p_{1.} \times p_{2.}$. With constant G_i's, this implies that $\omega_3 = \omega_1\omega_2/W$ and the estimator reduces to the Petersen estimator in this case. With perfect negative correlation, $\omega_3 = 0$ and the abundance estimate is zero, although a better estimator would be the number of animals seen in total in this case. A removal method estimator could also be used. This is a pathological case - in mark-recapture terms it

corresponds to total trap-shyness after initial capture (when an animal is observed by one observer, its probability of being observed by the other observer is zero). With less extreme negative correlation, the Petersen estimator is positively biased while the Generalized Petersen estimator corrects for the effect of correlation. In the case of perfect positive correlation everything seen by observer 1 must be seen by observer 2 as well, so that $G^* = G_1 = G_2$ and $g^*(x,z) = g_1(x,z) = g_2(x,z) = g_3(x,z)$. From this it follows that

$$\begin{aligned} \frac{W\omega_3}{\omega_1\omega_2} &= \frac{W}{\omega^* G^*} \\ &= \frac{1}{p^*_{..}} \end{aligned} \tag{7.15}$$

(where $\omega^* = \int\int g^*(x,z)\,dx\,dz$ and $p^*_{..} = \int\int G^* g^*(x,z)\frac{1}{W}\,dx\,dz$). In this case the Petersen estimator N_P would be equal to $n_{..}$, the number of animals seen, whereas the generalized Petersen estimator corrects for the (estimated) proportion of animals seen by neither observer, as follows.

$$\begin{aligned} \hat{N}_{GP} &= \frac{\hat{N}_P}{\hat{p}^+_{..}} \\ &= \frac{n_{..}}{\hat{p}^+_{..}} \end{aligned} \tag{7.16}$$

If the only explanatory variable is x, ω_3 is estimated by the conventional LT estimator $\hat{f}_3(0)^{-1}$ from the perpendicular distances of animals detected by both observers, and ω_1^{-1} and ω_2^{-1} are estimated by $\hat{f}_1(0)$ and $\hat{f}_2(0)$ respectively, the estimators are as follows.

$$\hat{G}_1 = \frac{n_{3.}}{n^+_{2.}}\left(\frac{\hat{f}_3(0)}{\hat{f}_2(0)}\right) \tag{7.17}$$

$$\hat{G}_2 = \frac{n_{3.}}{n^+_{1.}}\left(\frac{\hat{f}_3(0)}{\hat{f}_1(0)}\right) \tag{7.18}$$

$$\hat{N}_{GP} = \hat{N}_P\left(W\frac{\hat{f}_1(0)\,\hat{f}_2(0)}{\hat{f}_3(0)}\right) \tag{7.19}$$

These estimators for G_1 and N are identical in form to the estimators proposed by Buckland and Turnock (1992) for use in the presence of responsive animal movement. This is immediately apparent in the case of the G_1 estimator, and some re-arranging of their estimator of N (called $\hat{N}_{BT}$ here) makes it apparent for estimation of N:

$$\hat{N}_{BT} = \frac{n^+_{1.} W\,\hat{f}_1(0)}{\hat{G}_1}$$

$$= \frac{n_{1.}^{+} W \hat{f}_1(0)}{\frac{n_{3.}\hat{f}_3(0)}{n_{2.}^{+}\hat{f}_2(0)}}$$

$$= \hat{N}_P \left(W \frac{\hat{f}_1(0)\,\hat{f}_2(0)}{\hat{f}_3(0)} \right) \tag{7.20}$$

The difference between their estimator and the GP estimator, $\hat{N}_{GP}$, lies only in the way $f_1(0)$ is estimated.

The IWC-Type Estimator

Before leaving the subject of Constant-Shape estimators, consider the estimator of N which Butterworth and Borchers (1988), Palka (1993) and others use once G_1 and G_2 have been estimated. The estimator is described in Hiby and Hammond (1989). With this method, N is estimated by separately estimating the shape of **the detection function for all observers combined** ($f_{.}(0)$), using conventional LT methods and assumptions, and then dividing the resultant conventional LT estimate of abundance by the combined estimate of detection probability on the trackline, as follows.

$$\hat{G}_{.} = \hat{G}_1 + \hat{G}_2 - \hat{G}_1\hat{G}_2 \tag{7.21}$$

The abundance estimator is

$$\hat{N}_{IWC} = \frac{n_{..} W \hat{f}_{.}(0)}{\hat{G}_{.}} \tag{7.22}$$

Rewriting this in terms of N_{GP} one gets the following.

$$\hat{N}_{IWC} = \hat{N}_{GP} \left(\frac{n_{..}\hat{f}_{..}(0)}{n_{1.}\hat{f}_1(0) + n_{2.}\hat{f}_2(0) - n_{3.}\hat{f}_3(0)} \right) \tag{7.23}$$

If all of $\hat{f}_{..}(0)$, $\hat{f}_1(0)$, $\hat{f}_2(0)$, and $\hat{f}_3(0)$ are asymptotically unbiased, then asymptotically the term in brackets has expectation $N/(N+N-N) = 1$, and $\hat{N}_{IWC}$ is equivalent to N_{GP}.

7.3 Conditional Likelihood Estimation

Given appropriate models for each component of the likelihood L_U or L_B, the abundance, N, could be estimated by maximizing the likelihood. If the models are correct, the main advantage of this approach is efficiency of estimation and the ability to obtain profile likelihood confidence interval estimates. The utility of such fully parametric estimation depends on having the correct likelihood, and may give misleading results if the assumed likelihood is far from the true underlying likelihood. Buckland *et al.* (1993a) discuss full likelihood estimation in the context of conventional LT models.

As an alternative to estimation based on the full likelihood, estimation of subsets of parameters, conditional on some subset of the parameters of the full likelihood. This approach allows one to take advantage of methods already developed for estimation of some subsets of the parameters. For example, one might want to use robust conventional LT methods for estimating the shape of the detection functions, and then to estimate abundance conditional on these estimates - as in the case of the Generalized Petersen estimator developed above. Alternatively, one might use GLM theory and software to estimate the parameters of the detection functions, and then use the mle of N, conditional on these estimates. Another reason to opt for conditional estimation is that fewer assumptions need to be made about the distribution of statistics of which one has little if any knowledge. Two examples are the distributions of $n_{..}$ and of (x, z). Robust methods which use separate transects in the same survey region as sampling units for estimating the variance of $n_{..}$ and confidence intervals have been developed in the context of conventional LT surveys. Similar methods can be applied in the MRLT context, thereby avoiding the need to model the distribution of $n_{..}$.

With more than one explanatory variable in the detection probability model, another reason to avoid basing inference on the full likelihoods is that the form of the (generally) multivariate density $\pi(x, z)$ must be specified for estimation, but it may be difficult *a priori* to propose a suitable form for this density. With appropriate parititioning of the full likelihood, it is possible to obtain mle's of the detection probabilities with or without making assumptions about $\pi(x, z)$, and then to estimate abundance conditional on the estimated detection probabilities. (This is what the Huggins/Alho estimator does.) Conditional estimation also allows an estimation scheme which falls somewhere between the Huggins/Alho method and a full likelihood approach in this respect, allowing knowledge of the marginal distribution of x to be used while making no assumptions about the joint distribution of x and z other than that x and z are independent. (I develop this sort of estimator below.)

With conditional maximum likelihood estimation, abundance estimation is conveniently separated into two stages.

(1) Estimation of the combined detection probabilities.

(2) Estimation of abundance, conditional on the estimated detection probabilities.

At each stage there is a variety of estimators available, the most appropriate of which will depend on the particular application. Factors affecting the decision on which estimators to use include the form of the data (binned or unbinned), its content (what explanatory variables are available), and what assumptions can reasonably be made about the survey process. In the following sections, I consider only conditional likelihood estimation and I summarize and discuss the options available at stages (1) and (2) above. I concentrate on estimation stage (2) and propose two new estimators for abundance, given estimates of detection probability. I also discuss some properties of the estimators.

7.4 Detection Probability Estimators

The following is a summary of some of the available estimators of detection probabilities. This has been compiled from the estimators available in the literature, and those developed in this and the previous Chapter.

(P1) Estimation of $p_1(x,z)$ and $p_2(x,z)$ by maximum likelihood from f_{U3} (not necessarily using the logistic form for the two detection functions). In this case the estimated combined detection probability at (x,z) is as follows.

$$\hat{p}_{\cdot}(x,z) = \hat{p}_1(x,z) + \hat{p}_2(x,z) - \hat{p}_1(x,z)\hat{p}_2(x,z) \tag{7.24}$$

(Note that $\pi(x,z)$ **need not be estimated or assumed** in order to estimate $p_{\cdot}(x,z)$.)

(P2) Petersen-like estimation in non-overlapping intervals under the assumption that $p_i(x,z)$ is constant within each interval. Denoting the Petersen mark-recapture estimator of abundance in the kth interval by $\hat{N}_k$, the estimator of the combined detection probability at (x,z) is as follows.

$$\hat{p}_{\cdot}(x,z) = \sum_{k=1}^{K} I_k(x,z)\frac{n_{\cdot k}}{\hat{N}_k} \tag{7.25}$$

Here $I_k(x,z)$ is an indicator function which is equal to 1 if (x,z) is in the kth bin, and is zero otherwise.

(P3) Estimation of $p_1(x,z)$, $p_2(x,z)$ and $\pi(x,z)$ (using some appropriate functional form for each) from f_{B3}. Note that if there are explanatory variables other than x, $\pi(x,z)$ must be estimated or assumed in order to estimate of $p_{\cdot}(x,z)$.

(P4) Estimation of $g_1(x,z)$, $g_2(x,z)$, $g_3(x,z)$ using methods which incorporate other covariates, but retain the assumption that all animals on the trackline are detected, followed by estimation of G_1 and G_2 with the Generalized Petersen estimator. In this case, an estimator of combined detection probability at (x,z) is as follows.

$$\hat{p}_{\cdot\cdot}(x,z) = \hat{G}_1\,\hat{g}_1(x,z) + \hat{G}_2\,\hat{g}_2(x,z) - \hat{G}_1\hat{G}_2\,\hat{g}_3(x,z) \tag{7.26}$$

(P5) An IWC-like method, which involves estimation of $f_{\cdot}(x)$ from all observers' data (using methods which assume G to be unaffected by z), together with a combined estimate of $G_{\cdot\cdot}$. The combined estimate is $\hat{G}_{\cdot\cdot} = \hat{G}_1 + \hat{G}_1 - \hat{G}_1\hat{G}_2$, where $\hat{G}_1$ and $\hat{G}_2$ are obtained as in **(P4)**. An estimator of the combined detection probability at (x,z) is as follows.

$$\hat{p}.(x,z) = \hat{G}.\hat{g}.(x,z) \tag{7.27}$$

In general, the assumption that detection probability on the trackline is not affected by z is unlikely to hold. For example, large animals or clusters are likely to be more detectable than small animals or groups both on and off the trackline. Poor sighting conditions are likely to make animals less detectable both on and off the trackline. Observer fatigue is likely to result in fewer detections both on and off the trackline. When the only variable affecting detectability is perpendicular distance from the trackline, the assumption follows; but it is difficult to think of situations in which the constant G assumption is reasonable when probability of detection depends on other variables as well, except when $G = 1$.

P1 through P3 above are not based on the assumption of constant G, whereas P4 and P5 are. This suggests that P1 through P3 are better candidates for a general- purpose estimation method than P4 and P5, although whatever bias the constant G assumption introduces may be offset by reduced variance because fewer parameters need to be estimated. P2 involves a fairly strong assumption about the form of the detection function - namely that it is constant within each bin. In practice this will be an approximation to the real situation, as are all models. Whether it is an adequate approximation will depend on the degree of binning and the shape of the true detection function. Binning may present problems when there are many explanatory variables, because it requires that the joint domain of these variables be partitioned into non-overlapping bins. This introduces an arbitrary element to estimation - where to put the borders between the bins - and with finite sample sizes and many expanatory variables the bins may have to be so wide that the assumption of constant detection probability within each bin becomes untenable. Without extensive binning independent estimation in each bin is likely to be very inefficient. Finally, only P1 and P2 allow estimation of detection probability without having to estimate or assume $\pi(x,z)$.

P1 seems to be the most hopeful candidate for a general-purpose methods for estimation of the detection probability function when there are many explanatory variables affecting detectability. It involves relatively weak assumptions about the form of the detection function, and it does not require estimation of $\pi(x,z)$. The cost of relaxing the questionable assumption of P4 and P5 (of constant detectability on the trackline) is that additional parameters must be estimated (compared to the case where G is constant). These additional parameters come in the form of dependence of G on the explanatory variables. It is possible that P4 or P5 might produce estimators of detection probability with lower mean squared error than P1 when the constant G assumption is violated slightly, since they involve the estimation of fewer parameters.

7.4.1 Finding Suitable Multivariate Detection Function Forms

The published LT literature contains no examples of detection functions which model the dependence of G on explanatory variables. The closest parallels a LT surveys where such dependence has been modelled

are the analyses by Buckland *et al.* (1993b) and Laake *et al.* (1994) of shore-based surveys of migrating California Gray whales. These studies are not typical applications of LT theory because x (for both observed and unobserved animals) is unlikely to be distributed uniformly with respect to the coast and in the case of Buckland *et al.* (1993b), x was not selected as a significant explanatory variable. Like Huggins (1989) and Alho (1990), they both used a logistic form for the detection function. The logistic is not parameterized explicitly in terms of G, but if one of the explanatory variables is x, setting this to zero gives G, which will be a function of the remaining explanatory variables. An advantage of using the logistic form is that it results in a likelihood from the exponential family, so that the full power of generalized linear model theory can be applied in estimation. Its disadvantage when applied to LT surveys is that it may not be flexible enough to model the form of the dependence of the detection function on x adequately.

One aspect of this relative inflexibility is that the logistic is skew- symmetric about $p(x, \boldsymbol{z}) = 0.5$. This is atypical of LT detection functions. It is difficult to say how closely the logistic might be able to model $f(x)$'s from LT surveys in general when it is averaged over other explanatory variables. (Some such shapes are shown in the next Chapter.) However, it seems preferable to start with a model which is close to robust forms for LT detection functions and then generalize it, rather than to start with a detection function which is not close to robust forms for LT detection functions in the perpendicular distance dimension (x). Another aspect of the inflexibility of the logistic functional form becomes apparent when probability of detection on the trackline is substantially less than unity (say 0.5 or less), but the detection function has something of a shoulder. The only situation in which the logistic can produce a shoulder at $x = 0$ is when detection probability is close to unity on the trackline. Borchers *et al.* (1995) found the logistic form to be an inadequate model for harbour porpoise survey data, where average probability of detection on the trackline was estimated to be close to 0.3.

Generalizing conventional LT functional forms to allow dependence of G on the explanatory variables in a sensible way is not trivial without doubling or almost doubling the number of parameters. The generalization can easily be achieved by letting $G(x, \boldsymbol{z})$ be a logistic function (for example) of the explanatory variables other than perpendicular distance, and retaining the model for $g(x, \boldsymbol{z})$ (a hazard rate model with scale parameter a function of the explanatory variables other than perpendicular distance, for example). However, this requires two parameters for each explanatory variable (assuming the same explanatory variables are used in $G(x, \boldsymbol{z})$ and $g(x, \boldsymbol{z})$) included in the model. Alternatively, one could try a model along the lines of $p(x, \boldsymbol{z}) = g(x) \times h(\boldsymbol{z})$, where $g(x)$ is a typical LT detection function model with $g(0) = 1$ and $h(\boldsymbol{z})$ is of logistic form, but this form of dependence of the detection function may also be inadequate. (For example, in the study of Otto and Pollock, 1990, $g(x)$ was found to depend on $\boldsymbol{z}$ *via* its scale parameter.)

Buckland (*pers. commn*) suggests a model for the detection function which incorporates a typical LT $g(x)$ (the hazard rate form of Hayes and Buckland, 1983) within a GLM framework. Borchers *et al.* (1995) used this model to estimate abundance from IO data gathered on the "Small Cetaceans Abundance in

the North Sea" (SCANS) survey of 1994. It is the only model to have been used in a real multivariate MRLT application. Details of the model and estimation procedure are described in Appendix 7.10.3.

7.5 Abundance Estimators for Multivariate MRLT Models

Thus far, only estimators of N have been considered. In any actual survey it is estimators of $\aleph$ that are of interest, so before moving on to the consider general estimators, a few words on the relationship between estimators of N and of $\aleph$ are appropriate.

Here I deal only with the case in which a uniform coverage probability design is used. When coverage probabilities are not uniform over the whole survey region and a Horvitz–Thompson–like estimator of N is used, this is readily converted to a Horvitz–Thompson–like estimator of $\aleph$. This is done by using $(C_j \times \hat{p}_j)$ in place of $\hat{p}_j$ in the estimators, where $\hat{p}_j$ is the estimated detection probability associated with the jth detected animal (given that it is in the covered region) for the Horvitz-Thompson–like estimator of N, and C_j is the (known) coverage probability for the jth detected animal. Analytic estimation of variances and confidence intervals is more complicated. In the case of Horvitz–Thompson–like estimators, point and interval estimation has been considered by Cooke (1984, 1987c) in a LT survey context. The problem has been addressed further in an aerial survey context by Cook and Jacobson (1979), Steinhorst and Samuel (1989), Evans *et al.* (1993), Evans *et al.* (1994), Rivest *et al.* (1995), Hiby and Lovell (1996) and Wong (1996).

With a uniform coverage probability survey design, it is easy to frame LT abundance estimators of $\aleph$ in terms of N. For example, a conventional univariate LT estimator which has the form

$$\hat{\aleph} \;=\; A\hat{D} \;=\; A\left(\frac{n_{..}}{2L}\hat{f}(0)\right) \tag{7.28}$$

can be written as follows.

$$\hat{\aleph} \;=\; \left(\frac{A}{a}\right) \times \hat{N} \tag{7.29}$$

Here $a = 2LW$ and $\hat{N} = n_{..}W\hat{f}(0)$, or $\hat{N} = n_{..}/\hat{p}_{..}$ where $\hat{p}_{..} = [W\hat{f}(0)]^{-1}$ is the estimator of the average detection probability of an animal within W of the trackline. The term A/a is the inverse of the probability of an animal being within W of the trackline (assuming equal coverage probability throughout the survey region). A/a is a known constant from the survey design, and

$$E[\hat{\aleph}] \;=\; E\left[\frac{A}{a}\hat{N}\right] \;=\; \frac{A}{a}E\left[\hat{N}\right] \tag{7.30}$$

If $\hat{N}$ is unbiased for N, then $\hat{\aleph}$ is unbiased for $\aleph$. Also

$$Var[\hat{\aleph}] = Var\left[\frac{A}{a}\hat{N}\right] = \left(\frac{A}{a}\right)^2 Var\left[\hat{N}\right] \tag{7.31}$$

Confidence intervals for N are readily converted to confidence intervals for $\aleph$ by multiplying them by A/a.

The following sections deal with the second step in the two-stage estimation process, namely estimation of abundance given estimates of detection probability functions. An essential difference between the three estimators considered here is the way in which $\pi(x, z)$ is treated. The first estimator, the Horvitz–Thompson–like estimator of Huggins (1989) and Alho (1990), involves no assumptions about $\pi(x, z)$. The second involves the assumption that the marginal density $\pi(x)$ is known but makes no further assumptions about the joint density $\pi(x, z)$. The third is framed in terms of $\pi(p)$, the density of detectabilities in the population, rather than $\pi(x, z)$, the density of the explanatory variables, which in conjunction with the observers' detection functions give rise to $\pi(p)$. This final estimator is based on assuming a functional form for $\pi(p)$.

7.5.1 $\hat{N}_H$: A Horvitz–Thompson–like Estimator

Given estimates of the detection probability function $\hat{p}_.(x, z)$, it is possible to estimate abundance N without making any assumptions about the density of the explanatory variables, $\pi(x, z)$. Huggins (1989) and Alho (1990) do just this, using the first form of N-estimator I consider here, namely the Horvitz–Thompson–like estimator. The distinguishing (and most attractive) feature of this sort of estimator is its almost non-parametric nature - it does not require any assumptions about the form of the density $\pi(x, z)$.

$$\hat{N}_H = \sum_{j=1}^{n_.} \frac{1}{\hat{p}_.(x_j, z_j)} \tag{7.32}$$

If $p_.(x, z)$ were known the estimator would be a Horvitz–Thompson estimator (Horvitz and Thompson, 1952). Huggins,(1989) and Alho (1990) used a finite sampling approach to derive an expression for the asymptotic variance of $\hat{N}_H$ for this case. They did this by treating the explanatory variables $(x_1, z_1) \ldots (x_N, z_N)$ of the N animals in the survey region as fixed. Here I take a philosophically slightly different approach and consider $(x_1, z_1) \ldots (x_N, z_N)$ as N independent samples from the same density $\pi(x, z)$. The difference is of no practical consequence and the results are identical to those of Huggins and Alho. However, treating the explanatory variables as realizations of a stochastic process in this way suggests a generalization of the estimator of N which is a parametric variant of $\hat{N}_H$.

Estimation when the detection functions are known

Like Huggins (1989), I consider first the bias and variance of N_H when the detection probabilities are known and not estimated. Let $p_.(x, z)$ be the function which gives the unique true probability of at least

one of the observers detecting an animal with associated explanatory variables (x, z). Now let (x, z) be a random variable (with density $\pi(x, z)$). Since $p = p.(x, z)$ is a function of (x, z), it too is a random variable - with a density which I call $\pi(p)$, with mean μ and variance σ^2. In observing animals from this density "size-biased" sampling takes place, with the "size bias" being the random variable p itself. The density of observed p's is as follows.

$$f(p) = \frac{p\,\pi(p)}{\mu} \tag{7.33}$$

A word on notation: I use bold type to indicate the moments and expectations with regard to $\pi(p)$, the distribution of p in the underlying population (sampled and unsampled), and normal type to indicate the moments and expectations of the observed distribution of p ($f(p)$). For example, the mean of $\pi(p)$ is $\boldsymbol{E}[p] = \boldsymbol{\mu}$, while the mean of $f(p)$ is $E[p] = \mu$. (Note that $\boldsymbol{\mu} = p_{..} = \int\int p.(x, z)\,\pi(x, z)\,dxdz$.)

Expectations with respect to $f(p)$ are related to the expectations with respect to $\pi(p)$ as follows.

$$E[p^r] = \frac{\mathbf{E}[p^{r+1}]}{\boldsymbol{\mu}} \tag{7.34}$$

In particular:

$$E\left[\frac{1}{p}\right] = \frac{1}{\boldsymbol{\mu}} \tag{7.35}$$

Now $\boldsymbol{\mu}$ is the mean detection probability in the population, so that the expected number of animals detected $E[n_{..}]$ is equal to $N\boldsymbol{\mu}$. Thus $N = E[n_{..}]/\boldsymbol{\mu}$, and

$$\hat{N} = n_{..}\widehat{\left(\frac{1}{\boldsymbol{\mu}}\right)} \tag{7.36}$$

is an unbiased estimator of N when $\widehat{\frac{1}{\boldsymbol{\mu}}}$ is an unbiased estimator of $\frac{1}{\boldsymbol{\mu}}$ (assuming that $\widehat{\frac{1}{\boldsymbol{\mu}}}$ is independent of $n_{..}$). The central problem is the estimation of $\frac{1}{\boldsymbol{\mu}}$. Clearly

$$E\left[\sum_{j=1}^{n_{..}} \frac{1}{p_j}\right] = \frac{n_{..}}{\boldsymbol{\mu}} \tag{7.37}$$

so that $\hat{N}_H$ is unbiased. The model thus suggests estimators of the following form:

$$\widehat{\left(\frac{1}{\boldsymbol{\mu}}\right)} = \hat{E}\left[\frac{1}{p}\right] = \int_{p_0}^{1} \frac{1}{p}\,\hat{f}(p)\,dp \tag{7.38}$$

Here $\hat{f}(p)$ is an estimate of the density of observed p's, and $p_0 > 0$ is a lower limit for possible detection probabilities required for existence of the moments of p. (Alho, 1990, discusses the use of p_0, noting that in most practical applications it is quite reasonable, although it results in negatively biased estimators of abundance if some part of the population is unobservable - he cites an example of this.) $\hat{N}_H$ is an estimator of this form. It can be written as

$$\sum_{j=1}^{n_{..}} \frac{1}{\hat{p}_j} = \int_{p_0}^{1} \frac{1}{p} \left(\frac{1}{n_{..}} \sum_{j=1}^{n_{..}} I_{\hat{p}_j}(p) \right) dp \tag{7.39}$$

where $I_{\hat{p}_j}(p)$ is an indicator function which is equal to 1 when $p = \hat{p}_j$, and zero otherwise. Thus $\hat{N}_H$ uses the following non-parametric estimator of $f(p)$.

$$\hat{f}(p) = \frac{1}{n_{..}} \sum_{j=1}^{n_{..}} I_{\hat{p}_j}(p) \tag{7.40}$$

(One can think of this as an extreme type of kernel estimator which has a window width of zero.) Alternative estimators of abundance could be obtained by using kernel estimators which do more smoothing, or by fitting a smooth parametric density estimator for the interval $(0;1)$ to the observed p's. In fact, any density estimator on the interval $(0;1)$ is a candidate.

The above relates to the situation in which the p_j's are observed. In practice they can't be observed and must estimated (by one of the methods described above, for example). Thus $\hat{f}(p)$ is itself estimated by fitting to the estimated probabilities, $\hat{p}_j = \hat{p}_.(x_j, z_j)$, instead of the unobservable p_j's. In the absence of the requisite analytic results, one could reasonably speculate that an estimator of this type, which uses a smooth form for $\hat{f}(\hat{p})$, might be less variable than $\hat{N}_H$ which uses a totally unsmoothed form - because the variance of the fitted form would be lower. This would be offset to some extent by having to estimate the additional parameters that are associated with $\hat{f}(\hat{p})$, but with reasonable sample sizes and a parsimonious parameterization for $\hat{f}(\hat{p})$ some reduction in variance should be possible. Any reduction in variance may go hand-in-hand with an increase in bias, so that it is not obvious that the RMSE of the smoothed estimator would be lower than that of N_H.

Huggins and Alho treated the set of N true p's in the surveyed region as fixed. Here they are treated instead as a sample of N identically, independently distributed (iid) random variables from the density $\pi(p)$. The jth of these ($j = 1 \ldots N$) is sampled with probability p_j, independently of whichever other animals are sampled. Thus one can think of the sampling process as having two components.

- (1) Sampling the N p's in the sampled region from $\pi(p)$. (Call the set of these N p's "$\{p\}_N$".)
- (2) Sampling the $n_{..}$ observed animals (with probabilities p_j, $j = 1 \ldots n_{..}$) from the N p's in the sampled region.

Huggins and Alho do without step (1). Here I include it in obtaining an expression for $Var[\hat{N}_H]$ for the case in which the p_j's are observed. (The expression turns out to be identical to that of Huggins and Alho.) Let the subscript "(2)" on expectations and variances refer to step (2) above. Expectations and variances with respect to step (1) are shown in bold type, following the convention established above for moments of distributions with respect to the underlying population (observed and unobserved). If the detection probabilities were observable the variance of $\hat{N}_H$ could be obtained as follows.

$$\begin{aligned} Var[\hat{N}_H] &= \boldsymbol{E}\left[Var_{(2)}[\hat{N}_H] \mid \{p\}_N\right] + \boldsymbol{Var}\left[E_{(2)}[\hat{N}_H] \mid \{p\}_N\right] \\ &= \boldsymbol{E}\left[\sum_{j=1}^{n_{..}} \frac{1-p_j}{p_j}\right] + 0 \\ &= \sum_{j=1}^{n_{..}} \boldsymbol{E}\left[\frac{1-p_j}{p_j}\right] = n_{..}\boldsymbol{E}\left[\frac{1-p}{p}\right] \\ &= n_{..}E\left[\frac{1}{p}\left(\frac{1-p}{p}\right)\right] \end{aligned} \tag{7.41}$$

(The second line follows from the first line above by using the properties of Horvitz–Thompson estimators.) It follows that when detection probabilities are observed, the following is an unbiased estimator of this variance.

$$\widehat{Var}[\hat{N}_H] = \sum_{j}^{n_{..}} \left(\frac{1-p_j}{p_j^2}\right) \tag{7.42}$$

However when the detection probabilities are estimated, $\widehat{Var}[\hat{N}_H]$ above will be a negatively biased estimator of the variance of $\hat{N}_H$, because it does not take account of the variability introduced by estimating the detection probabilities.

Estimation with unknown detection probabilities

When detection probabilities are estimated rather than observed, the point estimator $\hat{N}_H$ is not necessarily unbiased and its variance includes a component due to estimation of the detection probabilities. Huggins (1989) and Alho (1990) showed that $\hat{N}_H$ is asymptotically unbiased. Expressions for the bias of the estimator for finite samples are difficult to obtain. I need a little more notation to deal with the variance in this case.

Let $\boldsymbol{c}$ represent the observed capture histories $\{c_j\}$ $(j = 1 \ldots n_{..})$.

Let $\boldsymbol{x}$ represent the observed explanatory variables $\{x_j, z_j\}$ $(j = 1 \ldots n_{..})$.

Make the dependence of $p_{.}(x, z)$ on its parameters, $\boldsymbol{\theta}$, explicit by writing it as $p_{.}(x, z \mid \boldsymbol{\theta})$.

Let $\hat{\boldsymbol{\theta}}$ be the estimate of the parameters $\boldsymbol{\theta}$ of the detection functions. Now $\hat{\boldsymbol{\theta}}$ is a function of both $\boldsymbol{c}$ and $\boldsymbol{x}$, and it is sometimes useful for what follows to make this explicit by writing $\hat{\boldsymbol{\theta}}$ as $\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x})$.

Similarly write $\hat{N}_H$ as $N_H(\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x}), \boldsymbol{x})$. (Given a sample of size $n_{..}$, the capture histories $\boldsymbol{c}$ affect $\hat{N}_H$ only to the extent that they determine the estimated parameters of the detection functions. The explanatory variables, on the other hand, affect $\hat{N}_H$ in determining both the estimated parameters of the detection functions and in determining the values of the detection probability for the jth animal, conditional on the estimated parameters of the detection function, $p_{.}(x_j, z_j \mid \hat{\boldsymbol{\theta}})$.)

With this notation, the variance of the estimator $\hat{N}_H$ can be written as follows.

$$\begin{aligned} Var[\hat{N}_H] &= E_{\boldsymbol{x}}[Var_{\boldsymbol{C}}[N_H(\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x}), \boldsymbol{x}) \mid \boldsymbol{x}]] + \\ & \quad Var_{\boldsymbol{x}}[E_{\boldsymbol{C}}[N_H(\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x}), \boldsymbol{x}) \mid \boldsymbol{x}]] \end{aligned} \tag{7.43}$$

The first term on the RHS is the expected variance in $\hat{N}_H$ due to estimation of the detection functions. (Were the detection functions known, the first term would disappear and the second would reduce to the variance of equation 7.41 above.)

An estimator for $Var[\hat{N}_H]$ can be obtained from this equation as follows. Since $\boldsymbol{\theta}$ is estimated by maximum likelihood (and assuming sufficient regularity for asymptotic mle results to apply) an expression for the asymptotic value of $Var_{\boldsymbol{C}}[N_H(\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x}), \boldsymbol{x}) \mid \boldsymbol{x}]$ is

$$Var_{\boldsymbol{C}}[N_H(\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x}), \boldsymbol{x}) \mid \boldsymbol{x}] = D_H(\boldsymbol{\theta})^T I(\boldsymbol{\theta})^{-1} D_H(\boldsymbol{\theta}) \tag{7.44}$$

where:

$D_H(\boldsymbol{\theta})$	is the (vector) derivative of $\hat{N}_H$ with respect to $\boldsymbol{\theta}$, and
$I(\boldsymbol{\theta})^{-1}$	is the inverse information matrix (and f_{U3}, considered as a function of the parameters, is the relavant likelihood in this case).

To a first order Taylor series approximation, the expectation $E_{\boldsymbol{x}}[N_H(\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x}) \mid \boldsymbol{x}]$ in the second term is $N_H(E_{\boldsymbol{C}}[\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x})])$. If $\hat{\boldsymbol{\theta}}(\boldsymbol{c}, \boldsymbol{x})$ is unbiased for θ, this is equal to $N_H(\boldsymbol{\theta}, \boldsymbol{x})$ and the second term in equation 7.43 above is $Var[N_H(\boldsymbol{\theta}, \boldsymbol{x})]$ - which is the variance of $\hat{N}_H$ when $\boldsymbol{\theta}$ is known (equation 7.41 above). The following is therefore an approximate expression for $Var[\hat{N}_H]$.

$$Var[\hat{N}_H] = D_H(\boldsymbol{\theta})^T I(\boldsymbol{\theta})^{-1} D_H(\boldsymbol{\theta}) + n_{..} E\left[\frac{1 - p_{.}(x_j, z_j \mid \boldsymbol{\theta})}{p_{.}(x_j, z_j \mid \boldsymbol{\theta})^2}\right] \tag{7.45}$$

Huggins (1989) and Alho (1990) showed this to be the asymptotic variance of the estimator. They also showed the estimator to be asymptotically unbiased and normally distributed. An estimator of the variance is obtained by using $\hat{\boldsymbol{\theta}}$ in place of $\boldsymbol{\theta}$:

$$\hat{Var}[\hat{N}_H] = D_H(\hat{\boldsymbol{\theta}})^T I(\hat{\boldsymbol{\theta}})^{-1} D_H(\hat{\boldsymbol{\theta}}) + \sum_{j=1}^{n_.} \left(\frac{1 - p_.(x_j, \boldsymbol{z}_j \mid \hat{\boldsymbol{\theta}})}{p_.(x_j, \boldsymbol{z}_j \mid \hat{\boldsymbol{\theta}})^2} \right) \tag{7.46}$$

In practice sample sizes may be insufficient for the asymptotic results to apply and the variance estimate above may be biased. Exact analytic expressions are not available for either the expectation or the variance of $\hat{N}_H$ when sample size is small. In the absence of such exact analytic expressions, bootstrapping is one way of estimating the variance of $\hat{N}_H$ and confidence intervals for N when sample size is small.

Bootstrap estimates of $Var[\hat{N}_H]$ and of a confidence interval for N can be obtained by resampling transects. This procedure does not rely heavily on the assumption of independent distribution of animals in the survey area because with appropriately large transects most of the spatial correlation between animals' positions is contained within transects, leaving little or no spatial correlation between transects. Buckland *et al.* (1993a) discuss this in the context of conventional univariate LT surveys. The same principles apply to multivariate MRLT surveys.

The bootstrap procedure is simple. Say that the survey contained T transects, of total length L. Bootstrap as follows.

(1) Resample (with replacement) T transects from the original T transects in the survey. One should constrain the total effort (L) in the resampled transects to be approximately equal to that on the original survey if variance is to be estimated conditional on L - which is the usual LT practice. Alternatively, one could condition on sample size by constraining the bootstrapped sample size. In either case, the issue of how to implement the constraint arises because with transects as the sampling unit, both effort and sample size come in transect-sized quanta so that constraining effort or sample size to be exactly equal to that on the original survey is usually not feasible. See Buckland *et al.* (1993a) for a discussion of this small complication.

(2) Estimate $\aleph$ and store the estimate.

(3) If insufficient bootstrap replications have been completed, go to (1); otherwise calculate variance and confidence intervals (CI's) from the stored list of bootstrap point estimates.

7.5.2 $\hat{N}_\omega$: Estimation with $\pi(x) = W^{-1}$

An important way in which MRLT models differ from MR models is that in the former the density of one of the explanatory variables (namely perpendicular distance, x) can often reasonably be assumed to be known and independent of other explanatory variables. The estimator $\hat{N}_H$ ignores whatever information may be available on the density of the expanatory variables. Given that all conventional LT theory is based on the assumption that $\pi(x)$ is known, and that intuitively at least one would expect use of information about $\pi(x)$ to translate into better estimation in some way, it seems worth developing a general estimator for MRLT models which incorporates the assumption that $\pi(x)$ is uniform.

Information on $\pi(x)$ can be incorporated in a Horvitz–Thompson–like estimator simply by using a different estimated probability in the denominator of each term in the estimator. Thus $\hat{p}_.(x_j, z_j)$ is replaced by $\hat{p}_.(z_j)$. I call this estimator $\hat{N}_\omega$:

$$\begin{aligned} \hat{N}_\omega &= \sum_{j=1}^{n_.} \frac{1}{\int_0^W \hat{p}_.(x, z_j)\, \pi(x)\, dx} \\ &= \sum_{j=1}^{n_.} \frac{1}{\hat{p}_.(z_j)} = \sum_{j=1}^{n_.} \frac{W}{\hat{\omega}_.(z_j)} \end{aligned} \tag{7.47}$$

(Recall that $\omega_.(z) = \int p_.(x, z)\, dx$, the "effective strip width" for animals with attributes z, and that $\pi(x) = W^{-1}$.) This estimator has the same form as the estimators of Drummer and McDonald (1987) and Quang (1991). $\hat{N}_\omega$ is a natural generalization of their estimator for the MRLT situation in which detection on the trackline is not certain and detection probability on the trackline is allowed to vary with z. When there is no z, $\hat{N}_\omega$ reduces to $\hat{N}_C$ of the previous Chapter.

Because the estimator is identical in form to that of Huggins (1989), and θ is estimated in the same way, one can follow his derivation of the asymptotic distribution of $\hat{N}_H$ word-for-word, as it were, to obtain the asymptotic distribution of $\hat{N}_\omega$. (The only substantial difference is that whereas Huggins' development requires that the $p_.(x_j, z_j)$'s are sufficiently regular for standard theory for obtaining maximum likelihood estimators to apply, here we require that the $p_.(z_j)$'s are sufficiently regular.) The estimator $\hat{N}_\omega$ is therefore asymptotically unbiased and normally distributed with variance

$$Var[\hat{N}_\omega] = D_\omega(\theta)^T I(\theta)^{-1} D_\omega(\theta) + n_. E\left[\frac{1 - p_.(z_j \mid \theta)}{p_.(z_j \mid \theta)^2}\right] \tag{7.48}$$

where

$D_\omega(\theta)$ is the (vector) derivative of $\hat{N}_\omega$ with respect to θ, and
$I(\theta)^{-1}$ is the inverse information matrix (where f_{U3}, considered as a function of the parameters, is the relevant likelihood in this case).

An estimator of this variance is

$$\hat{Var}[\hat{N}_\omega] = D_\omega(\hat{\theta})^T I(\hat{\theta})^{-1} D_\omega(\hat{\theta}) + \sum_{j=1}^{n_.} \left(\frac{1 - p_.(z_j \mid \hat{\theta})}{p_.(z_j \mid \hat{\theta})^2}\right) \tag{7.49}$$

Bootstrap estimates of variance and confidence intervals for $\aleph$ can be obtained using an algorithm identical to that given above for $\hat{N}_H$.

7.5.3 $\hat{N}_\mu$: Estimation by Modelling π

If $\pi(x, z)$ were known (as it is assumed to be in conventional LT models when x is the only explanatory variable), then this knowledge could be used in estimating N by using a generalization of the conventional LT estimator form $\hat{N} = n_./\hat{p}_.$ (where $\hat{p}_. = \int\int \hat{p}_.(x, z)\pi(x, z)\, dx\, dz$).

One would expect that using knowledge of $\pi(x, \boldsymbol{z})$ in estimation might improve estimation precision. Thus it seems worth considering estimators of this form as an alternative to the Huggins/Alho estimator for the general case in which detection probability depends on x and $\boldsymbol{z}$.

For most applications, it would not be reasonable to assume that $\pi(x, \boldsymbol{z})$ was known, so that at best one might be able to assume a functional form for $\pi(x, \boldsymbol{z})$, with parameters which needed to be estimated. Even if one can reasonably assume that x is independent of $\boldsymbol{z}$, it may be difficult to propose appropriate forms for $\pi(\boldsymbol{z})$. (What, for example, would an appropriate form be for the joint distribution of group size and sea-state on a marine survey?) Furthermore, even if a suitable form for the distribution for $\pi(\boldsymbol{z})$ could be proposed, if $\boldsymbol{z}$ has more than one element (i.e. if there is more than one explanatory variable which needs to be modelled), a substantial number of degrees of freedom may be lost in estimating the parameters of $\pi(\boldsymbol{z})$. As a result, any gain in precision which may accrue from using knowledge of $\pi(x)$ and the form of $\pi(\boldsymbol{z})$ may be lost in estimating the parameters of $\pi(\boldsymbol{z})$.

By estimating the density of detection probabilities, $\pi(p)$, rather than $\pi(x, \boldsymbol{z})$, the number of additional parameters to be estimated over and above those needed for estimation of $\hat{N}_H$ can be kept constant, irrespective of how high the dimension of $\boldsymbol{z}$ might be. Say the dimension of $\boldsymbol{z}$ is $R \geq 1$. The estimation problem is then reduced from that of estimating an R-dimensional density, to that of estimating a 1-dimensional density. (A disadvantage of modelling $\pi(p)$ directly is that one loses whatever information is available on $\pi(x)$. To use that one would need to model $\pi(p|x)$, which I have not attempted here. This is an approach which may be worth pursuing in future.)

Given a suitable functional form for $\pi(p)$ (most obviously a Beta density, but potentially any smooth function on the interval $(0; 1)$ which integrates to 1), its parameters (denoted $\boldsymbol{\phi}$) can be estimated conditional on the estimated parameters of the detection functions ($\hat{\boldsymbol{\theta}}$ say) by maximizing either of the likelihoods below with respect to $\boldsymbol{\phi}$. (One could also perform unconditional estimation using the full likelihood, but I do not consider that here.) The likelihoods are equivalent to the pdf's above with the same subscripts, but are rewritten as functions of $\boldsymbol{\phi}$, conditional on the detection probabilities. To do this I need to define a few quantities, as follows.

$$\hat{p}_j = p_{\cdot}(x_j, \boldsymbol{z}_j \,|\, \hat{\boldsymbol{\theta}}) \tag{7.50}$$

$$p_{\cdot\cdot}(\boldsymbol{\phi}) = \int_0^1 p\,\pi(p \,|\, \boldsymbol{\phi})\,dp \tag{7.51}$$

$$p_{\cdot k}(\boldsymbol{\phi}) = \int_{c_{k-1}}^{c_k} p\,\pi(p \,|\, \boldsymbol{\phi})\,dp \tag{7.52}$$

For grouped estimation of $\pi(p \,|\, \boldsymbol{\phi})$ it is the p-axis that is divided into K intervals, with cutpoints at $c_0, \ldots, c_K$, $c_0 = 0$ and $c_K = 1$. Similarly $n_{\cdot k}$ is redefined to be the number of animals detected with estimated detection probabilities in the kth p-interval, rather than the kth bin in $(x, \boldsymbol{z})$-space. Using

these definitions, the appropriate likelihoods are as follows.

$$L_{U2} = f_{U2}\{p_1 \ldots p_{n_{..}} \mid n_{..}\}$$

$$= \prod_{j=1}^{n_{..}} \frac{p_j \, \pi(p_j \mid \phi)}{p_{..}(\phi)} \tag{7.53}$$

$$L_{B2} = f_{B2}\{n_{.1} \ldots n_{.K} \mid n_{..}\}$$

$$= \left(\frac{n_{..}!}{n_{.1}! \ldots n_{.K}!}\right) \prod_{k=1}^{K} \left(\frac{p_{.k}(\phi)}{p_{..}(\phi)}\right)^{n_{.k}} \tag{7.54}$$

The mean of $\pi(p \mid \hat{\phi})$ (the estimated density of detection probabilities) is the model estimate of $E[p] = p_{..} = \int\int p_{.}(x,z)\pi(x,z)\,dx\,dz$. In practice, the detection probabilities are not observed, so that the estimated detection probabilities are used in their place in estimating $\pi(p \mid \hat{\phi})$. (For example, if $\pi(p \mid \hat{\phi})$ is assumed to be a Beta distribution with parameters α and β, then $\phi = (\alpha, \beta)$ and $\hat{p}_{..} = \hat{\alpha}/(\hat{\alpha} + \hat{\beta})$. Here $\hat{\alpha}$ and $\hat{\beta}$ are the estimators of α and β obtained by maximizing one of the above likelihoods with respect to these parameters, using the estimated detection probabilities as data.) I call this estimator of mean detectability $\hat{\mu}$ and the resulting abundance estimator $\hat{N}_{\mu} = n_{..}/\hat{\mu}$.

Note that one could in principle use $f_{U2\&3}$ or $f_{B2\&3}$ for estimation in a similar way, except that in this case one would have to estimate the bivariate density $\pi(p_1, p_2 \mid \phi_{1,2})$ (where the dimension of $\phi_{1,2}$ is at least that of ϕ). If one used the "Constant-G" assumption, maximization would have to be with respect to G_1, G_2, and $\phi_{1,2}$. There is no obvious gain from this approach - only a loss of precision because more parameters may need to be estimated - so I have not pursued it further.

Obtaining an expression for even the asymptotic variance of this estimator is more difficult than for $\hat{N}_H$. Because of this and the questionable validity of variance and interval estimators based on the strong assumptions of independence when animals are unlikely to be distributed independently of one another, bootstrapping seems a more productive route to follow. The same procedure as that for $\hat{N}_H$ can be used, substituting $\hat{N}_{\mu}$ for $\hat{N}_H$.

7.6 Estimating pdf's and Functions of Explanatory Variables

With both $\hat{N}_H$ and $\hat{N}_{\omega}$, estimation of $\pi(x,z)$ goes hand-in-hand with estimation of abundance. (This is not the case with $\hat{N}_{\mu}$, since $\pi(x,z)$ is not modelled directly in this case.) Consider the simple example in which explanatory variables are discrete and no structure is placed on either the detection probabilities or the density of explanatory variables. In this case, the mle for N is the sum of the Petersen estimates of abundance at each of the explanatory variable values sampled. (A simplified version of this in which x

was the only explanatory variable was considered in the previous Chapter.) The corresponding estimate of density at each of the values of the explanatory variable observed is simply proportional to the Petersen estimate of abundance at that point. Estimated densities at values of the explanatory variable which were not observed in the sample are zero. This sort of density estimator is non-parametric in that no structure is placed on the density being estimated. An alternative would be to assume some smooth form for the density over the range of possible explanatory variables. Then this model could either be incorporated in the detection probability estimation procedure or used to smooth the non-parametric density estimate across all values of the explanatory variable.

When explanatory variables are continuous, one can estimate densities nonparametrically in much the same way as for the discrete case above, using indicator functions to "pick out" only those points that were sampled. Making no assumptions about $\pi(x)$ and $\pi(z)$, they can be estimated as follows.

$$\tilde{\pi}(x) = \frac{\left(\sum_j \frac{I_x(x_j)}{\hat{p}_.(x,z_j)}\right)}{\left(\sum_j \frac{1}{\hat{p}_.(x_j,z_j)}\right)} = \frac{\left(\sum_j \frac{I_x(x_j)}{\hat{p}_.(x,z_j)}\right)}{\hat{N}_H} \tag{7.55}$$

and

$$\tilde{\pi}(z) = \frac{\left(\sum_j \frac{I_z(z_j)}{\hat{p}_.(x_j,z)}\right)}{\left(\sum_j \frac{1}{\hat{p}_.(x_j,z_j)}\right)} = \frac{\left(\sum_j \frac{I_z(z_j)}{\hat{p}_.(x_j,z)}\right)}{\hat{N}_H} \tag{7.56}$$

(Here $I_x(x_j)$ and $I_z(z_j)$ are indicator functions which are equal to 1 when their argument is equal to their subscript, and are zero otherwise.) If the explanatory variables were discrete, the numerator would simply be the estimated abundance at the given observed value of the explanatory variable (using $\hat{N}_H$ evaluated at this point only), while the denominator is the estimate of N (using $\hat{N}_H$ evaluated at all observed points).

If one is prepared to treat $\pi(x)$ as known, an estimator of the density of z is as follows.

$$\hat{\pi}(z) = \frac{\left(\sum_j \frac{I_z(z_j)}{\hat{p}_{..}(z)}\right)}{\hat{N}_\omega} \tag{7.57}$$

The estimator has the same form as $\tilde{\pi}(z)$, except that $\hat{N}_\omega$ is used in place of $\hat{N}_H$ and $\hat{p}_{..}(z)$ is used in place of $\hat{p}_.(x_j, z_j)$.

To generalize, partition the explanatory variables (x, z) into two exclusive sets u and v, where it is u-space in which we are interested. (For example, if we were interested in the joint density of sea-state (z_1, say) and group size (z_2, say), we would have $u = (z_1, z_2)$, with v containing x and any other explanatory variables in z.) Estimators of the density of u when no assumptions are made about $\pi(x)$, and when $\pi(x)$ is assumed to be known are, in that order

$$\hat{\pi}(\boldsymbol{u}) = \frac{\left(\sum_j \frac{I_{\boldsymbol{u}}(\boldsymbol{u}_j)}{\hat{p}_.(\boldsymbol{u},\boldsymbol{v}_j)}\right)}{\hat{N}_H} \tag{7.58}$$

and

$$\tilde{\pi}(\boldsymbol{u}) = \frac{\left(\sum_j \frac{I_{\boldsymbol{u}}(\boldsymbol{u}_j)}{\int_0^W \hat{p}_.(\boldsymbol{u},\boldsymbol{v}_j)\,\pi(x)\,dx}\right)}{\hat{N}_\omega} \tag{7.59}$$

The density of $\boldsymbol{v}$ can be estimated similarly.

Using these estimators, it is possible to estimate both moments of the explanatory variables and any function of the explanatory variables which may be of interest. An example of the former is mean group size estimation, and of the latter the expected value of the detection function in the $\boldsymbol{u}$-dimension.

Mean Group Size

Say that explanatory variable z_1 was group size (and for simplicity assume that group size is observed without error). Then, treating $\pi(x)$ as known, $\boldsymbol{u} = z_1$ and mean group size is estimated as follows.

$$\hat{E}[z_1] = \sum_{j=1}^{n_.} z_{1j}\hat{\pi}(z_{1j}) \tag{7.60}$$

Here z_{1j} is the size of the jth group observed. Using equation 7.57, this reduces to the following.

$$\hat{E}[z_1] = \frac{\hat{N}_{\omega_{indiv}}}{\hat{N}_\omega} \tag{7.61}$$

where $\hat{N}_{\omega_{indiv}}$ is the Horvitz–Thompson–like estimator of animal abundance under the assumption that groups are distributed uniformly over perpendicular distances 0 to W (see section 7.8 below).

Marginals of the Detection Functions

The detection function is an obvious example of a function of the explanatory variables which is of interest. Treating $\pi(x)$ as known, the expected value of the detection function $p_.(\,)$ at $\boldsymbol{u}$ can be estimated by

$$\hat{p}_.(\boldsymbol{u}) = \sum_{j=1}^{n_.} \hat{p}_.(\boldsymbol{u},\boldsymbol{v}_j)\,\hat{\pi}(\boldsymbol{v}_j) \tag{7.62}$$

Plots of the marginals of the detection function in one dimension ($\boldsymbol{u} = x$, for example) show how detection probability varies with respect to the variable (x in this example) on average. Plots when $\boldsymbol{u}$ has more than one dimension show how detection probability varies jointly with the components of $\boldsymbol{u}$. Confidence intervals for functions of the explanatory variables estimated in this way can be obtained from the bootstrap procedure described above, as can estimates of variances.

7.7 An Estimator for Dealing with Animal Movement

Both random and responsive movement will in general cause estimators of abundance from MRLT surveys to be biased. The one estimator for which this is not the case is the BT method estimator for the univariate situation, when it is used in conjunction with the survey procedures of Buckland and Turnock (1992). Central to the method is the estimation of the probability of observer 1 (who searches close to her survey platform) detecting animals, given their **positions at the time they were available for detection by observer 2.** (Recall that searches ahead of observer 1 at a distance at which animals will not yet have reacted to the observers' presence; animals are assumed not to have reacted by the time they become detectable to observer 2). It turns out that $\hat{N}_\omega$ can readily be adapted to provide an estimator of abundance which deals with reactive or random animal movement. This is done as follows.

Condition on observer 2's detections and explanatory variables $(x_{2j}, \boldsymbol{z}_{2j})$ $(j = 1 \ldots n_{2\cdot}^+)$. (An additional subscript, $_2$, has been added to indicate detection by observer 2. The detection function of observer 1 can be estimated by binary regression with an appropriate functional form (see Appendix 7.10.3) using observer 2's detections as the trials. (Detection by observer 1 is a "success", non-detection a "failure".) Call the resulting estimated detection probability function $\hat{p}_1(x, \boldsymbol{z})$, as usual. An estimator of abundance is then readily obtained by adapting $\hat{N}_\omega$ to use only observer 1's detections, as follows.

$$\hat{N}_{\omega_1} = \sum_{j=1}^{n_{1\cdot}^+} \frac{W}{\hat{\omega}_1(\boldsymbol{z}_{1j})} \tag{7.63}$$

Here $\hat{\omega}(\boldsymbol{z}_{1j}) = \int_0^W \hat{p}_1(x, \boldsymbol{z}_{1j})\, dx$, and the additional subscript $_1$ on the $\boldsymbol{z}$'s indicates $\boldsymbol{z}$'s of observer 1's detections. (Note that it is assumed that the $\boldsymbol{z}$'s remain constant over the period during which they are detectable by either observer 1 or observer 2. In practice, this may not always be the case.) This is the estimator that was used in Borchers *et al.* (1995) in conjunction with a detection probability estimation method similar to that of Appendix 7.10.3 to estimate small cetacean abundance in the North Sea.

Note that $\hat{N}_H$ could not be adapted in this way because it involves the x's associated with observer 1's detections in the denominators of the terms in the sum. The problem is that these x's are the x's of observer 1's detections **before the animals (possibly) reacted to the survey vessel, i.e. at the time they were available for detection by observer 2**, and these are known only for duplicate detections (in which case the x is that recorded by observer 2). In principle, the estimator $\hat{N}_\mu$ could be adapted in a similar way to the adaptation of $\hat{N}_\omega$ above (equation 7.63).

7.8 Estimation When Animals Cluster

The models and estimators developed in this thesis thus far have been framed in terms of units/target objects called "animals". When these target objects (animals/plants/etc.) cluster into identifiable schools/pods/groups/etc. (collectively referred to as either "clusters" or "groups" in this thesis), the unit of detection is often the group rather than the individual animals within the group. This situation can be handled by defining the group as the "animal", and then estimating mean group size by some procedure, as discussed in Chapter 4. Interval estimation of the number of individuals is then conventionally based on the assumption that mean group size in the population (sampled and unsampled) is independent of group density. The estimators $\hat{N}_H$ and $\hat{N}_\omega$ provide a method for simultaneous estimation of mean group size and group density which does not require the assumption of independence between the two. This is achieved simply by replacing the 1 in the numerator of $\hat{N}_H$ and $\hat{N}_\omega$ by z_{1j}, where z_{1j} is the group size of the jth detected group, to give the following estimators of *individual* abundance in the covered region.

$$\hat{N}_{H_{indiv}} = \sum_{j=1}^{n_{..}} \frac{z_{1j}}{\hat{p}(x_j, z_j)} \qquad \text{and} \tag{7.64}$$

$$\hat{N}_{\omega_{indiv}} = \sum_{j=1}^{n_{..}} \frac{z_{1j}}{\hat{\omega}_{.}(z_j)} \tag{7.65}$$

Analytic estimation of the variance of $\hat{N}_{H_{indiv}}$ under the assumption of independence of group detections is readily obtained by replacing $\hat{p}_{.}(x_j, z_j)^{-1}$ by $z_{1j} \times \hat{p}_{.}(x_j, z_j)^{-1}$ in the second term of equation 7.46 and redefining $\boldsymbol{D}_H(\hat{\boldsymbol{\theta}})$ to be the vector derivative of $\hat{N}_{H_{indiv}}$ with respect to $\hat{\boldsymbol{\theta}}$. Similarly, an estimator of the variance of $\hat{N}_{\omega_{indiv}}$ under the assumption of independence of group detections is readily obtained by replacing $\hat{\omega}_{.}(z_j)^{-1}$ by $z_{1j} \times \hat{\omega}_{.}(z_j)^{-1}$ in the second term of equation 7.49 and redefining $\boldsymbol{D}_\omega(\hat{\boldsymbol{\theta}})$ to be the vector derivative of $\hat{N}_{\omega_{indiv}}$ with respect to $\hat{\boldsymbol{\theta}}$.

As in the case of group abundance estimation, it is usually unlikely that groups are located independently in the survey region, so that transect-based bootstrap estimators of variance and CI's may be more appropriate than the analytic estimators. This is particularly the case if sample size (number of groups detected, not number of transects covered) is small, since the analytic estimators are based on the asymptotic properties of the estimators. (When the number of transects covered is small, the bootstrap estimates may themselves be unreliable as transects are the resampling units for the bootstrap.)

Estimates of mean group size are readily obtained as follows.

$$\hat{E}[z_1]_H = \frac{\hat{N}_{H_{indiv}}}{\hat{N}_H} \qquad \text{and} \tag{7.66}$$

$$\hat{E}[z_1]_\omega = \frac{\hat{N}_{\omega_{indiv}}}{\hat{N}_\omega} \tag{7.67}$$

The variance of these estimators and CI's for mean group size can be estimated using a transect-based bootstrap procedure.

Borchers *et al.* (1995) estimated group and individual abundance as well as mean group size using $\hat{N}_{\omega}$, $\hat{N}_{\omega_{indiv}}$ and $\hat{E}[z_1]_{\omega}$, respectively, for three species of small cetacean from North Sea survey data. For these data, they found that mean group size was negatively correlated with group density, a result which raises doubts about the validity of the conventional practice of estimating the variance of individual abundance estimates under the assumption that mean group size is independent of group density.

A final question relating to mean group size estimation, which I must mention but do not address in any detail, is that frequently group size cannot be observed without error, so that group sizes have to be estimated. In addition to containing uncertainty, group size estimates may be biased. If the true group sizes of a representative or random sample of estimated group sizes can be obtained (representative or random with respect to the factors which affect errors and bias in group size estimation), this can be used to estimate a factor to correct bias in estimated group sizes. Furthermore, when variance estimation is by way of the bootstrap method, the variance due to this correction factor can be incorporated into estimates of mean group size and individual abundance by recomputing the correction factor at every bootstrap iteration. This is the procedure adopted by Borchers *et al.* (1995).

7.9 Conclusion

The likelihood equations of the MRLT models provide a consolidated framework for existing conventional LT estimators, existing "$g(0)$" estimators, and generalizations which include explanatory variables other than perpendicular distance. This framework is one which makes apparent relationships between various LT estimators in the literature (and some MR estimators too), and which facilitates the development of maximum likelihood and other estimators of abundance for a very wide range of LT scenarios.

With suitable assumptions it is possible to obtain mle's for N using the full likelihood. Alternatively, one can obtain estimators by using mle's based on conditional likelihoods for one subset of the model parameters, in conjunction with mle or non-mle estimators of another subset of the model parameters. The broad approach taken here has been to (1) estimate detection probabilities using maximum likelihood methods and then to (2) estimate abundance conditional on the estimates obtained from (1). A variety of estimators for both steps were considered and/or developed in this Chapter. One advantage of the conditional approach is that estimation of the detection probabilities can potentially be performed within a generalized linear model framework, using the theory and software available for GLMs. The usual GLM link functions may not, however, provide appropriate forms for LT detection functions. One method of adapting a GLM link function for the purposes of LT surveys, while still allowing GLM software to be used for estimation, is described in Appendix 7.10.3. The development of suitable LT link functions for GLMs is an area which would benefit from more research.

A distinguishing feature of the estimators developed for the general MRLT situation is the degree to

If the statistical performance of all three estimators of group abundance were identical, $\hat{N}_\omega$ would be first choice as a robust general-purpose MRLT abundance estimator. Obviously this judgement will be affected by whatever statistical properties the three estimators turn out to possess. In the following Chapters, I concentrate on Q3, addressing the question of the statistical properties of $\hat{N}_H$, $\hat{N}_\omega$ and $\hat{N}_\mu$. I do this by way of a simulation study. Chapter 8 describes the details of the study, while Chapter 9 reports the results.

7.10 Appendices

7.10.1 Derivation of the "Constant G" estimators from f_{2_P}

For brevity, I denote the explanatory variables $\boldsymbol{x} = (x, \boldsymbol{z})$. I write the detection function $p_i(x, \boldsymbol{z})$ as $G_i\, g_i(x, \boldsymbol{z})$ (with $g_i(0, \boldsymbol{z}) = 1$ for all $\boldsymbol{z}$). For notational ease, I dispense with the "k" subscript, so that the total number of animals seen by observer i only is denoted n_i, for example (rather than $n_{i.}$).

Assume that estimates $\hat{\omega}_i$ of

$$\omega_i = \int\int g_i(x, \boldsymbol{z})\, \pi(\boldsymbol{z})\, dx\, d\boldsymbol{z} \tag{7.68}$$

(for $i = 1, 2$) and an estimate $\hat{\omega}_3$ of

$$\omega_3 = \int\int g_1(x, \boldsymbol{z})\, g_2(x, \boldsymbol{z})\, \pi(\boldsymbol{z})\, dx\, d\boldsymbol{z} \tag{7.69}$$

are available. When x is the only explanatory variable, these estimates can be obtained from the application of conventional line transect methods using the perpendicular distance distribution data of each of the observation platforms separately for ω_1 and ω_2, and of the duplicate detections for ω_3. When there are additional explanatory variables $\boldsymbol{z}$, extensions of conventional LT methods which retain the $g(0, \boldsymbol{z}) = 1$ assumption (by modelling the scale parameter of $g(0, \boldsymbol{z})$ as a function of $\boldsymbol{z}$, for example) must be used.

Now f_{2_P} can be rewritten in these terms:

$$f_{2_P} = \left(\frac{n_.!}{n_1! n_2! n_3!}\right) \frac{\{G_1(\hat{\omega}_1 - \hat{\omega}_3 G_2)\}^{n_1} \{G_2(\hat{\omega}_2 - \hat{\omega}_3 G_1)\}^{n_2} \{G_1 G_2 \hat{\omega}_3\}^{n_3}}{\{G_1\hat{\omega}_1 + G_2\hat{\omega}_2 - G_1 G_2 \hat{\omega}_3\}^{n_.}} \tag{7.70}$$

Differentiating $\log(f_{2_P})$ with respect to G_1 and G_2 and setting the results equal to zero yields the following.

$$\frac{n_1^+}{G_1} - \frac{\hat{\omega}_3(n_2^+ - n_3)}{\hat{\omega}_2 - \hat{\omega}_3 G_1} = \frac{n_.(\hat{\omega}_1 - \hat{\omega}_3 G_2)}{G_1\hat{\omega}_1 + G_2\hat{\omega}_2 - G_1 G_2 \hat{\omega}_3} \tag{7.71}$$

$$\frac{n_2^+}{G_2} - \frac{\hat{\omega}_3(n_1^+ - n_3)}{\hat{\omega}_1 - \hat{\omega}_3 G_2} = \frac{n_.(\hat{\omega}_2 - \hat{\omega}_3 G_1)}{G_1\hat{\omega}_1 + G_2\hat{\omega}_2 - G_1 G_2 \hat{\omega}_3} \tag{7.72}$$

Multiply the first equation by $(\hat{\omega}_2 - \hat{\omega}_3 G_1)$ and the second by $(\hat{\omega}_1 - \hat{\omega}_3 G_2)$, simplify (recognising that $n_. = n_1^+ + n_2^+ - n_3$) and reorganize to get:

$$\frac{n_1^+\hat{\omega}_2 - n_.\hat{\omega}_3 G_1}{G_1} = \frac{n_.(\hat{\omega}_1 - \hat{\omega}_3 G_2)(\hat{\omega}_2 - \hat{\omega}_3 G_1)}{G_1\hat{\omega}_1 + G_2\hat{\omega}_2 - G_1 G_2\hat{\omega}_3} \tag{7.73}$$

$$\frac{n_2^+\hat{\omega}_1 - n_.\hat{\omega}_3 G_2}{G_2} = \frac{n_.(\hat{\omega}_1 - \hat{\omega}_3 G_2)(\hat{\omega}_2 - \hat{\omega}_3 G_1)}{G_1\hat{\omega}_1 + G_2\hat{\omega}_2 - G_1 G_2\hat{\omega}_3} \tag{7.74}$$

Equating the left hand sides of the equations and simplifying gives the relationship between G_1 and G_2:

$$G_2 = \left(\frac{n_2\hat{\omega}_1}{n_1\hat{\omega}_2}\right) G_1 \tag{7.75}$$

Substituting this into equation (7.72) and simplifying again eventually gives the results required.

$$\hat{G}_1 = \frac{n_3\hat{\omega}_2}{n_2^+\hat{\omega}_3} \tag{7.76}$$

$$\hat{G}_2 = \frac{n_3\hat{\omega}_1}{n_1^+\hat{\omega}_3} \tag{7.77}$$

The corresponding estimated probability of detection by either observer is as follows.

$$\begin{aligned} \hat{p}_. &= \hat{G}_1\frac{\hat{\omega}_1}{W} + \hat{G}_2\frac{\hat{\omega}_2}{W} - \hat{G}_1\hat{G}_2\frac{\hat{\omega}_3}{W} \\ &= n_.\left(\frac{n_3}{n_1^+ n_2^+}\right)\left(\frac{\hat{\omega}_1\hat{\omega}_2}{W\hat{\omega}_3}\right) \end{aligned} \tag{7.78}$$

Maximizing f_1 conditional on this estimate yields the mle of abundance:

$$\hat{N} = \frac{n_.}{\hat{p}_.} = \left(\frac{n_1^+ n_2^+}{n_3}\right)\left(\frac{W\hat{\omega}_3}{\hat{\omega}_1\hat{\omega}_2}\right) \tag{7.79}$$

The first term in brackets is the Petersen mark-recapture estimator.

7.10.2 An Estimator of Effective Strip Width when G is Constant

Ignoring the subscript indexing observer for notational convenience, write the detection function as $G\,g(x,z)$, where $g(0,z) = 1$. Assume that x and z are independent, and that $\pi(x) = W^{-1}$. It is $\omega = \int\int g(x,z)\,\pi(z)\,dx\,dz$ that we want to estimate. Having estimated $g(x,z)$, $\omega(z)$ can be estimated by integrating over x as follows.

$$\omega(z) = \int_0^W \hat{g}(x,z)\,dx \tag{7.80}$$

Generally $\pi(z)$ would not be known, so that $\omega = \int \omega(z)\,\pi(z)\,dz$ can't be estimated by taking the expectation of $\hat{\omega}(z)$ with respect to $\pi(z)$. Nevertheless, the results of section 7.5.1 can be used to estimate ω as follows.

Let $\pi(p)$ be the density of detection probabilities ($p = Gg(x,z)$) in the population (sampled and unsampled). Let $\boldsymbol{E}[p]$ be the expectation of p in the population of animals from which a sample is taken, and $E[p]$ be the expectation of the "size-biased" sample of observed p's. Then from section 7.5.1, $E[p^{-1}] = 1/\boldsymbol{E}[p]$. Hence

$$E[p^{-1}] = E\left[\frac{W}{G\,\omega(z)}\right] = \frac{1}{\boldsymbol{E}[p]} = \frac{W}{G\,\omega} \tag{7.81}$$

Because G and W are constants, it follows that

$$E\left[\frac{1}{\omega(z)}\right] = \frac{1}{\omega} \tag{7.82}$$

If $\omega(z)$ were observed rather than estimated, then given a sample size of n, the following is an unbiased estimator of ω^{-1}.

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\omega(z)} \tag{7.83}$$

Hence, if $\hat{\omega}(z)$ is a consistent estimator of $\omega(z)$ then

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\hat{\omega}(z_j)} \tag{7.84}$$

is a consistent estimator of ω^{-1} and

$$\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\hat{\omega}(z)}\right]^{-1} \tag{7.85}$$

is a consistent estimator of ω.

7.10.3 The "Modified Logistic Regression" Method

This Appendix contains a description of the so-called "Modified Logistic Regression" (MLR) method of Borchers *et al* (1995) for estimating the detection function. The method was suggested by Buckland (*pers. commn*) and developed in Borchers *et al.* (1995) for abundance estimation from data from the survey method of Buckland and Turnock (1992). The distinguishing feature of the application of Borchers *et al.* (1995) from a theoretical LT perspective is that the detection function is estimated for only one of the observers, conditional on the detections by the other observer. With some small changes to the method, it is extended here to the case where the detection functions for both observers are estimated. The original method is described in Borchers *et al.* (1995).

The problem of model-misspecification in the x-dimension of detection functions with logistic functional form is overcome by using a typical conventional LT detection functional form to transform x (to say $t(x)$, or t for brevity). The logistic regression method of Buckland *et al.* (1993b) is then used with (t, z) instead of (x, z) as explanatory variables. Because the data are used to estimate the parameters of $t(x)$, the transformation step and regression step are performed iteratively, until the estimated parameters from both steps converge. This procedure is described in more detail below. First the details of the transformation $t(x)$ are given.

The Transfromation, $t(x)$

In this subsection, I dispense with the subscripts which distinguish between the two observers. The transformation is framed in "generic" terms, and the contents of this subsection apply to both observers.

If x was the only explanatory variable, we would expect the hazard rate model of Hayes and Buckland (1983) together with an intercept parameter (called G here) to fit the observed data adequately in the x-dimension. In this case, the detection function, $p(x)$ (or p for brevity) would be

$$p = p(x) = G\left[1 - \exp\left\{-\left(\frac{x}{\sigma}\right)^{-b}\right\}\right] \tag{7.86}$$

where σ and b are the parameters of the detection function. If p is now transformed to t using the logit transfomation, i.e.

$$t = \ln\left\{\frac{p}{1-p}\right\} \tag{7.87}$$

then a detection function with logistic functional form, as a function of t, is exactly equal to the hazard rate detection function, as a function of x, with intercept parameter G, i.e.

$$p(t) = \frac{\exp\{t\}}{1+\exp\{t\}} = G\left[1 - \exp\left\{-\left(\frac{x}{\sigma}\right)^{-b}\right\}\right] \tag{7.88}$$

This is the rationale for using the transformation t. Obviously if x were the only explanatory variable, the transformation would be unnecessary. It is only useful in the presence of other explanatory variables, $\boldsymbol{z}$. In that case, it allows the detection function to be formulated in logistic regression terms while retaining something close to the hazard rate form in the x-dimension. In particular, it provides a form of logistic regression which accommodates a point of inflection at any detection probability value, where the point can lie anywhere between $x = 0$ and $x = W$.

The Estimation and Variable Selection Procedure

With R explanatory variables $\boldsymbol{z} = (z_1 \dots z_R)^T$ included in addition to x, the detection function for the ith observer ($i = 1, 2$) is as follows.

$$p_i(x, \boldsymbol{z}) = \frac{\exp\left\{\beta_0 + \beta_1 t(x) + \sum_{r=2}^{R} \beta_r z_{r-1}\right\}}{1 + \exp\left\{\beta_0 + \beta_1 t(x) + \sum_{r=2}^{R} \beta_r z_{r-1}\right\}} \tag{7.89}$$

The model has $2(R + 5)$ parameters. There are $(R + 2)$ β's for each observer, as well as the three transformation parameters: G, b and σ.

Parameter estimation is performed iteratively as follows.

(1) For each observer, estimate the two parameters, b and σ, which determine the hazard rate functional form. (Only the x's of each observer's detections are used to estimate these parameters.)

(2) For each observer, choose an initial estimate of G.

(3) Using the logistic regression procedure of Buckland *et al.* (1993b), estimate the $(R+2)$ β's for each observer, conditional on the estimates of G, b and σ.

(4) Select variables to be included in the model by backward stepwise selection (or some other method), conditional on G, b and σ.

(5) Minimize the scaled deviance of the logistic regression (of (3) above, including only those variables selected in step (4) above) with respect to the G's and the β's.

(6) If the two G's estimated in (5) above are different from that in (3) above, go to (3) and repeat using the new G's, else exit.

Chapter 8

The Simulation Experiment Design

8.1 Introduction

The purpose of the simulation experiment is to compare the statistical properties, and the small–sample properties in particular, of the three general MRLT estimators of Chapter 7 ($\hat{N}_H$, $\hat{N}_\omega$ and $\hat{N}_\mu$) under a specific set of assumptions about the detection process. The assumptions are assumptions (1) to (6) of conventional LT models given in Chapter 1, but with assumption (1) relaxed to allow the probability of detection to depend on the observable variable $\boldsymbol{z}$ in addition to x, and assumption (2) removed to allow probability of detection of animals on the trackline to be less than unity.

Although these assumptions are much less restrictive than those of conventional LT models, they remain very much an idealization of real survey environments. (I have discussed the effects of failure of some of the remaining assumptions elsewhere in this thesis, and I mention them again in the concluding Chapter.) Although the environment is idealized, examining the properties of estimators in this environment is a useful starting point for understanding the behaviour and properties of the three estimators in more realistic contexts.

All analyses are conducted using some truncation at perpendicular distance (i.e. discarding data beyond some maximum perpendicular distance W). Buckland *et al.* (1993a) recommend that univariate LT analyses (in which perpendicular distance is the only explanatory variable) be carried out using truncation at perpendicular distance. The results of the following Chapter indicate strongly that the same is true for MRLT analyses. Given a truncation distance W, estimated animal density in the survey region (with surface area A) is $\hat{N}/(2wL)$ (where $\hat{N}$ is one of the three estimators under consideration).

Given that an equal coverage probability survey design is to be used, and that animals are distributed independently in the plane, the survey design is determined by L and W. Once L and W are fixed in the design, the process governing the distribution of animals in space contributes a component of variance to $\hat{\mathrm{N}}$ as a result of variance in N. I am less interested in the properties of the estimators as a function of the stochastic process determining the distribution of animals in the surveyed region, than as a function of the stochastic nature of the detection process, and therefore choose to condition the simulations on N. This conforms to the way in which the likelihood functions developed in Chapter 6 are formulated;

N is a fixed parameter rather than a random variable. With N constant, variation in the estimates of N comes from two sources, namely variation in (x, z) (governed by $\pi(x, z)$) and sampling variation given the N realized (x, z)'s.

A difficulty with simulation studies is that they can't provide results which are as general as those from analytic evaluations. To conduct them one has to choose specific simulation models with specific values of the model parameters, and inference beyond the combinations of models and values chosen is to a large extent a matter of faith. Choice of a "representative" set of models and parameters for simulation is important in order to be able to make reasonable inferences about a more general set of models and parameters.

At the broadest level, the general set includes detection models in which the following apply.

- Detection on the trackline is less than certain.
- Detection probability depends on prependicular distance from the trackline (x) and other variables (z), both on and off the trackline.

More specifically, the set is determined by the following.

- The detection function form (logistic/hazard rate/etc.).
- The values of the parameters (β) of the detection functions.
- The joint distribution of x and z ($\pi(x; z)$, which is assumed to be equal to $\pi(x)\pi z$) since x and z can be made independent by design).
- The abundance in the surveyed region (N).

8.2 The Explanatory Variables

8.2.1 A Note on $(x; z)$ and Truncation at Perpendicular Distance

The MRLT scenario of interest is one in which there are multiple sources of heterogeneity (i.e. one in which z is a vector). However, the higher the dimension of z, the more assumptions need to go into the form and parameter values of $\pi(z)$. Furthermore, the higher the dimension of z, the more variable selection becomes an issue in estimation. Like others (Alho, 1990, in particular), I do not address the issues of model selection or variable selection in this simulation study. With these considerations in mind I have chosen to model heterogeneity (other than that attributable to perpendicular distance) through a scalar z rather than a vector z. Conceptually the scalar z is a composite variable which combines all sources of heterogeneity in detection probability other than that due to perpendicular distance, and it is implicitly assumed that no sources of heterogeneity are omitted in estimation. With this interpretation of z, variable selection becomes less of an issue.

The scale of x and z is arbitrary, so for simplicity I use $x \in (0;1)$. The upper limit of x used in simulations is the truncation point for perpendicular distance; only data with perpendicular distances less than this value are used in estimation. The choice of the truncation point has implications for robust estimation of the detection function. Guidelines for truncation of perpendicular distance are available for the conventional univariate case (in which detection probability depends only on perpendicular distance and detection on the trackline is certain); truncation criteria for the more general MRLT case are less well developed. For robust estimation when fitting conventional LT detection functions, Buckland *et al.* (1993a) recommend that either perpendicular distance data be truncated about at the point that $p(x) = 0.15$ or the greatest 5% of perpendicular distances be discarded. Detection functions used in simulation were therefore designed so that after taking expectation over z, the detection functions were in the region of 0.15 at $x = 1$ (as far as this was feasible given the other constraints on the detection functions).

8.2.2 The Joint Density of x and z

Perpendicular distances from the trackline are assumed to be independent uniform random variables over the range of perpendicular distances used in estimation.

The variable representing the heterogeneity, z is assumed to be a normal random variable centered on the origin, and independent of x. I do not claim any realism in modelling heterogeneity in this way. The assumption of normality is made in the absence of any compelling reason to assume any other specific functional form.

Three levels of heterogeneity were chosen, corresponding to three values for the standard deviation of z (σ). The values chosen for σ are 1 (level a: "high heterogeneity"), 0.5 (level b: "moderate heterogeneity") and 0.25 (level c: "low heterogeneity"). The choice of levels is unavoidably somewhat arbitrary, given that there are no data available in the literature on typical levels of heterogeneity. The values were chosen to span a wide range of feasible scenarios. Some idea of the way in which σ translates into variation in the detectability of animals can be obtained from Figure 8.1. The Figure shows detection functions of the central 95% of animals in each case. With $\sigma = 1$ this range includes animals which are close to undetectable (in the foreground of the Figure) and animals which are almost certain to be detected (in the background). (The range obviously also varies according to the detection function models 1, 2 or 3.) With $\sigma = 0.25$, the low heterogeneity case, the range of detection functions does not vary substantially about that for $\sigma = 0$.

To summarize, the (independent) distributions of the explanatory variables x and z are as follows.

$$x \sim \text{Uniform}(0;1)$$

$$z \sim \text{Normal}(0;\sigma^2) \qquad \sigma \in \{1, 0.5, 0.25\}$$

The two-dimensional densities are shown on the same scale as the detection functions, in Figure 8.1.

8.3 The Detection Functions

8.3.1 The Form of the Detection Functions

In order that simulation results are applicable to a general MRLT scenario in which the probability of detection depends on z both on the trackline and away from the trackline, simulations to compare the performance of the different abundance estimators were conducted using a detection function in which the probability of detection depends on z both on the trackline and away from the trackline. At the time the simulations were conducted the MLR method detection function (Appendix 7.10.3) had not yet been developed and there were no detection functions with this property developed specifically for LT survey data in the literature. The simulations were therefore conducted using the logistic functional form for the detection functions. While this form appears to be insufficiently flexible to be a useful general-purpose LT model, it does yield plausible shapes in the x- dimension and it also has the property that detection on the trackline is a function of z.

The detection function for the ith observer ($p_i(x;z)$, $i = 1,2$) therefore has the following form.

$$p_i(x;z) = \frac{\exp\{\beta_0 + \beta_1 x + \beta_2 z\}}{1 + \exp\{\beta_0 + \beta_1 x + \beta_2 z\}} \tag{8.1}$$

8.3.2 The Values of the Detection Function Parameters

The detection function parameters were chosen to give detection functions which

- are monotonically decreasing in x,
- are monotonically increasing in z,
- have expectation on the trackline ($E_z[p(0;z)]$) approximately equal to a pre-specified value (see below for details), and
- have expectation at the maximum x, (i.e. $E_z[p(1;z)]$) in the region of 0.15.

While most MRLT surveys conducted to date have involved two platforms with different detection functions (because of different heights and/or methods of searching and/or numbers of observers on each platform), the detection functions for the two platforms are taken to be identical in these simulations. This is for reasons of practicality; simulating combinations of a range of different detection functions for each platform increases the dimension of the simulation experiment to an impractical level.

Three sets of detection functions were simulated, with mean detection probabilities on the trackline, G_i ($i = 1,2$), approximately equal to 0.5 (Model 1), 0.7 (Model 2) and 0.85 (Model 3) respectively. In

Table 8.1: The Simulation Model Parameters

	Simulation Model								
	1			2			3		
Parameter	a	b	c	a	b	c	a	b	c
β_0	0	0	0	1	1	1	2	2	2
β_1	-3	-3	-3	-3	-3	-3	-3	-3	-3
β_2	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
σ	1.00	0.50	0.25	1.00	0.50	0.25	1.00	0.50	0.25

retrospect it may have been better to extend the range of simulation models to include models with lower probabilities of detection on the trackline. Unfortunately this is not a trivial task and I have not, therefore, pursued it here. In any case, the results of the current simulation study do give some indication of the properties of the estimators for situations in which animals are less detectable than the animals simulated here.

The parameters of the nine simulation models are listed in Table 8.1.

The shapes of the bivariate detection functions are shown in Figure 8.1. The detection functions get higher (i.e. animals are more detectable) as one moves from model 1 through to model 3. Viewed as functions of perpendicular distance (x, the axis running left to right), the model with least detectable animals (model 1) has animals ranging from moderately detectable ($0.9 \geq p_i(x,z) \geq 0.5$ at $z = 2$) to virtually undetectable ($0.05 \geq p_i(x,z) \geq 0.0$ at $z = -2$). The model with most detectable animals (model 3) has animals ranging from extremely detectable ($1.0 \geq p_i(x,z) \geq 0.9$ at $z = 2$) to moderately undetectable ($0.3 \geq p_i(x,z) \geq 0.03$ at $z = -2$). As one moves from the a-models (with high heterogeneity) to the c-models (with low heterogeneity), the range of the detection functions within 2σ of the mean $z = 0$ contracts about $z = 0$ and the more extreme detection functions (corresponding to animals with extreme z values) become increasingly rare.

The detection functions averaged over z are shown in Figure 8.2. The corresponding values of the detection functions at $x = 0$ at $x = 1$, and averaged over both x and z are shown in Table 8.2. The proportion of animals in the surveyed region which are detected by a single observer and by both observers combined are roughly 25% and 40%, respectively, for model 1; 40% and 60%, respectively, for model 2; and 60% and 80%, respectively, for model 3.

Note that at both x-values shown in Table 8.2 ($x = 0$ and $x = 1$), $p_{3.}(x)$ ($= E_z[p_3(x,z)]$) is greater than $p_{1.}(x)p_{2.}(x)$ ($= \{E_z[p_1(x,z)]\}^2$ because the detection functions for the two observers are identical), indicating positive correlation between $p_{1.}(x)$ and $p_{2.}(x)$ as a result of variation in z. Positive correlation (rather than negative correlation) is inevitable when the two detection functions used for simulation are identical. (Note also that the $E_z[p(1,z)]$ values for Model 1 are less than the target value of 0.15. Making

Figure 8.1: $\pi(x,z)$ and the Bivariate Detection Functions. The scales for x and z are the same in all the plots. The range of z is $-2 \leq z \leq 2$ for plots of $\pi(x,z)$, while it is $-2\sigma \leq z \leq 2\sigma$ for plots of $p(x,z)$. In both cases the range of x is $0 \leq x \leq 1$.

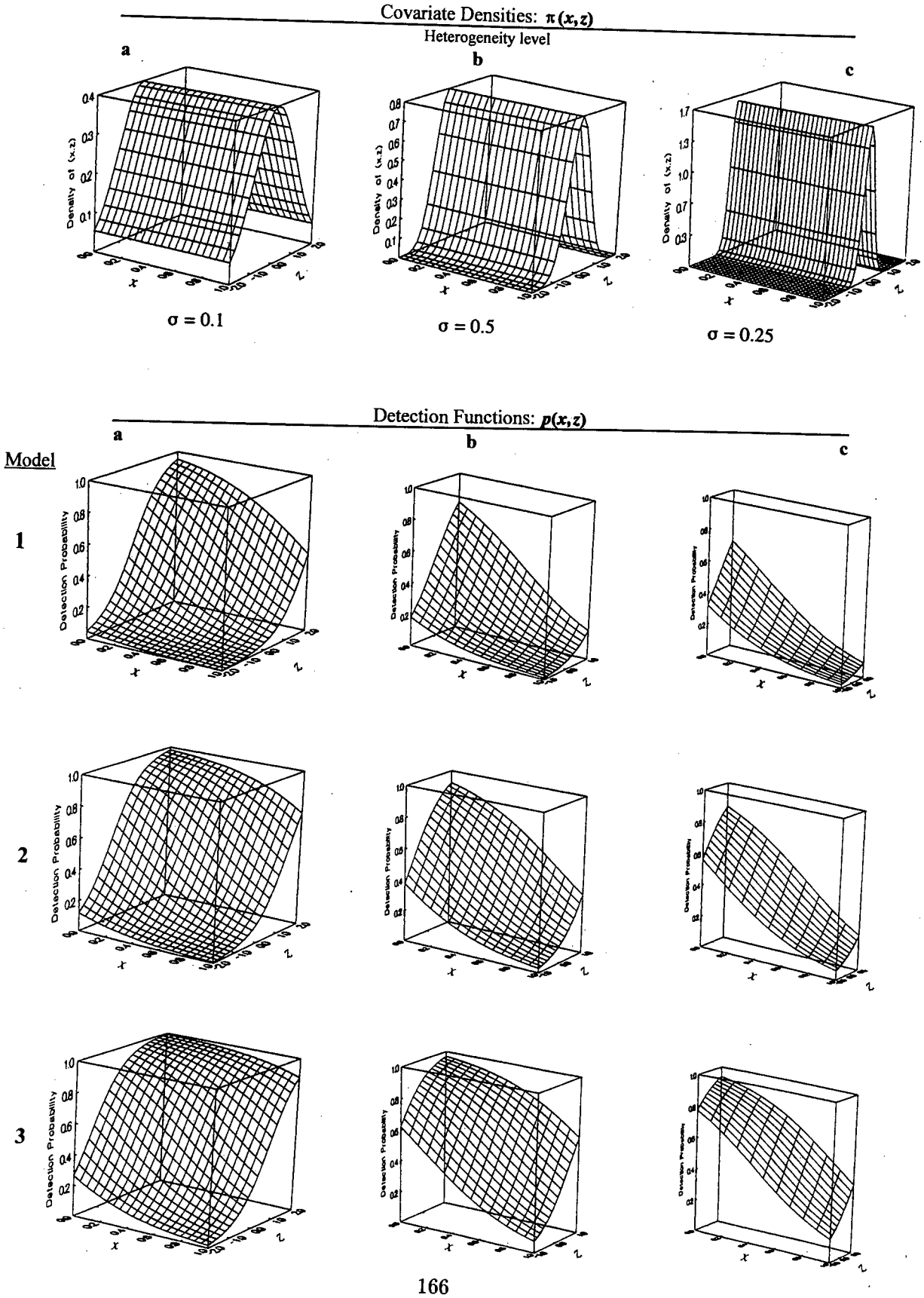


Figure 8.2: The Mean Detection Functions in the x-Dimension. The top curve is for both platforms, the middle is for a single observer, the bottom is for duplicates.

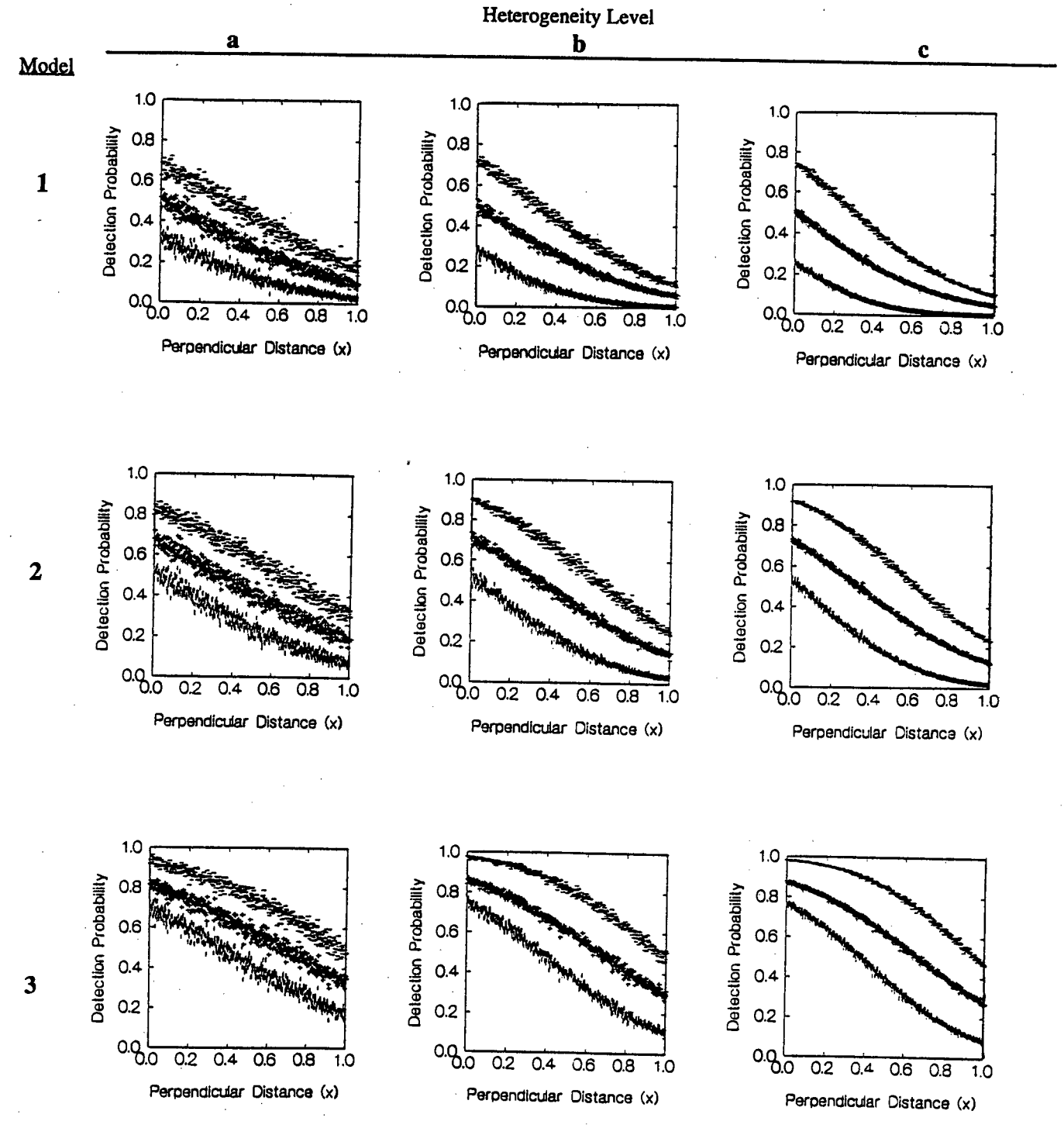


Figure 8.3: The Density of $p_i(x, z)$ for each Model.

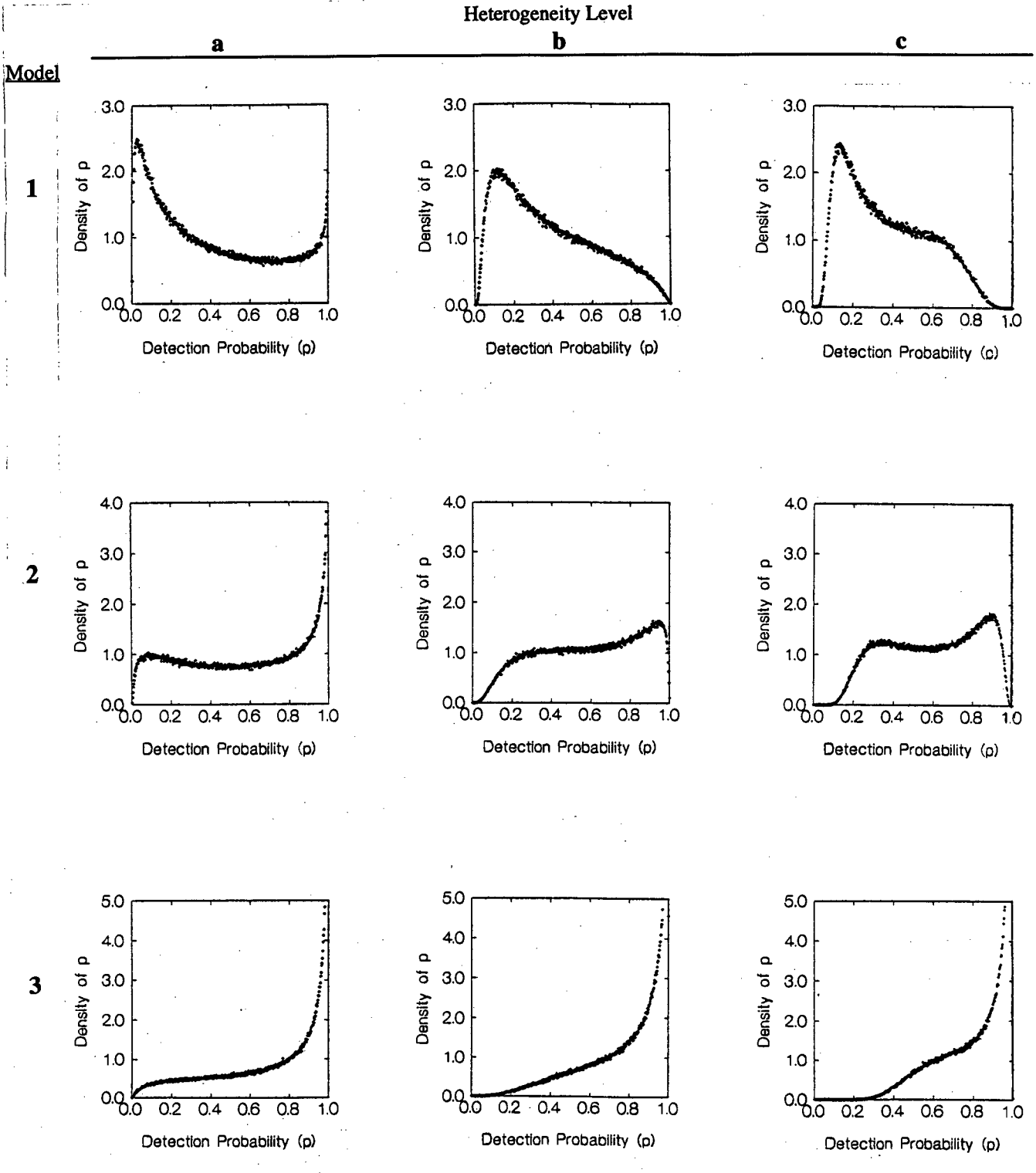


Figure 8.4: Some Shapes of The Beta Density Function

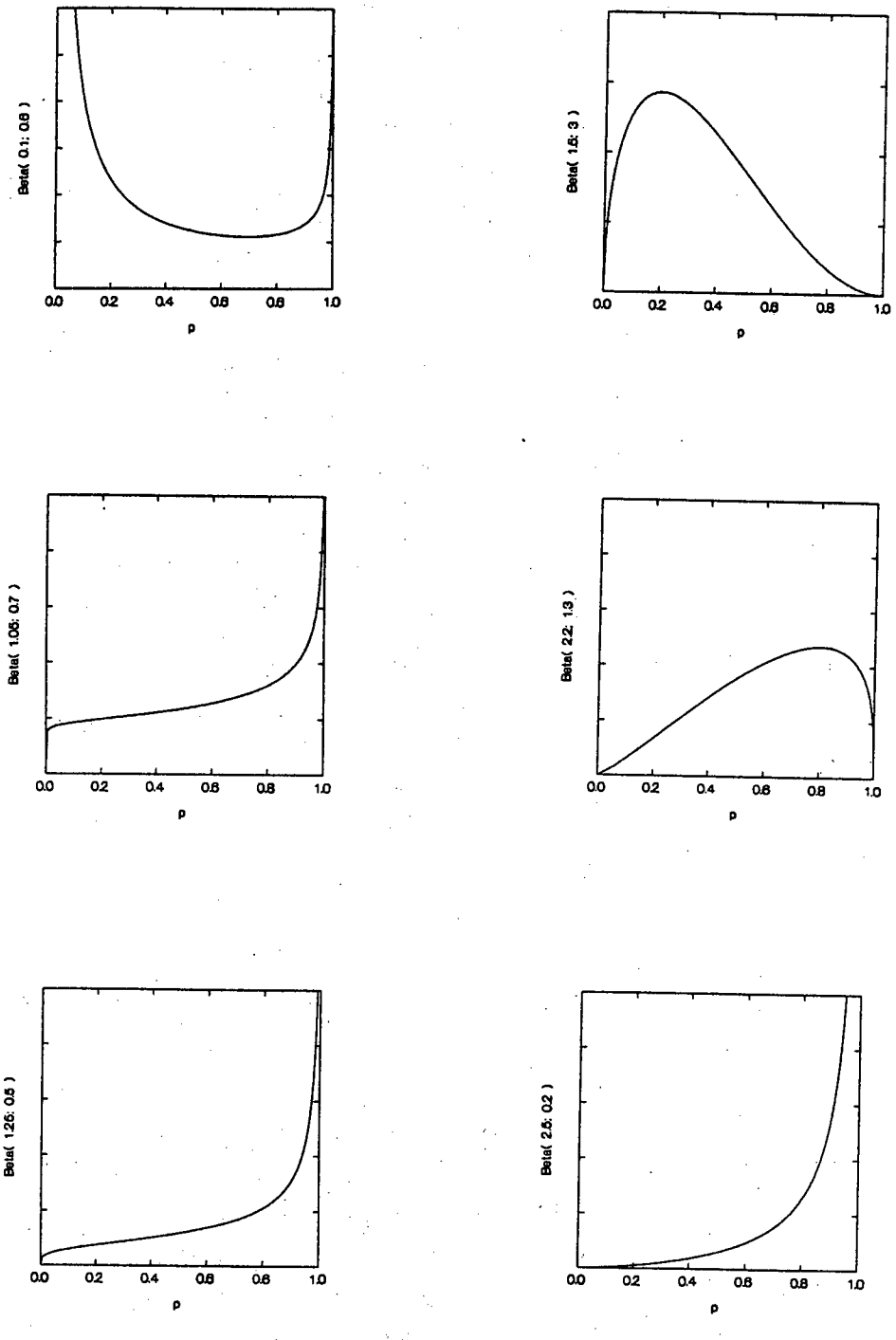


Table 8.2: Some Properties of the Simulated Detection Functions

Description	Quantity	Simulation Model								
		1			2			3		
		a	b	c	a	b	c	a	b	c
p on trackline ("$g(0)$")	$E_z[p_\cdot(0,z)]$	.67	.70	.73	.83	.90	.91	.92	.97	.98
	$E_z[p_1(0,z)]$	.51	.50	.49	.69	.70	.72	.88	.86	.87
	$E_z[p_3(0,z)]$	.33	.28	.25	.53	.53	.52	.70	.74	.76
minimum p	$E_z[p_\cdot(1,z)]$	.19	.11	.10	.30	.25	.23	.50	.48	.47
	$E_z[p_1(1,z)]$	.10	.06	.05	.19	.15	.13	.27	.29	.28
	$E_z[p_3(1,z)]$	.025	.005	.003	.075	.030	.018	.18	.11	.08
average p (μ)	$E_{(x,z)}[p_\cdot(x,z)]$	.40	.38	.37	.58	.59	.60	.74	.79	.80
	$E_{(x,z)}[p_1(x,z)]$	.27	.23	.22	.42	.40	.40	.58	.60	.60
	$E_{(x,z)}[p_3(x,z)]$	.14	.09	.07	.26	.21	.20	.42	.41	.40

them higher would have entailed what was felt to be too little change in the detection function over the range of x.)

The density of $p_\cdot(x,z)$ for each of the models is shown in Figure 8.3. The functions in both Figure 8.2 and Figure 8.3 have been evaluated by simulation, without smoothing the resulting points, which is why the curves are "blurred". With more simulation and/or smoothing the "blur" could be reduced or removed, but since the curves are used only to illustrate the shapes of the functions this was thought unnecessary. Incidentally, the degree of "blurring" gives an indication of the degree of heterogeneity.

8.4 The Simulated Abundance, N

Three separate levels of abundance of animals in the interval $x \in (0;1)$ were simulated, namely $N = 100$, $N = 250$ and $N = 500$. These values were chosen to span total and duplicate sample sizes ranging from a minimum below which estimation would frequently be difficult or impossible (because the probablity of obtaining zero duplicates is substantial), to sample sizes which by the standards of MRLT experiments in the literature would be considered large or very large (i.e. up to hundreds of duplicate detections).

The total and duplicate sample sizes associated with each of the levels of abundance, for each of the detection functions and each of the levels of heterogeneity, are shown in Table 8.3

8.5 The Responses

The responses of primary interest in the simulation experiment are listed below. Here T is the true value (as simulated) of a parameter of interest, $\hat{T}$ is the point estimator of T, and $\widehat{CI}[\hat{T}]$ is the estimated 95%

Table 8.3: Expected Sample Sizes (to nearest integer) from the Simulation Models

Seen by	N	Simulation Model 1 a	1 b	1 c	2 a	2 b	2 c	3 a	3 b	3 c
1 and/or 2	100	40	38	37	58	59	60	74	79	80
	250	101	95	93	145	148	150	185	197	201
	500	202	190	185	290	297	299	371	393	402
1	100	27	23	22	42	40	40	58	60	60
	250	68	58	55	105	101	100	145	149	151
	500	135	116	110	210	202	199	291	299	301
1 and 2 (duplicates)	100	14	9	7	26	21	20	42	41	40
	250	34	21	18	65	54	50	106	102	101
	500	68	43	35	130	107	99	211	204	201

confidence interval for T.

% Bias$[\hat{T}\ ] = 100 \times |E[\hat{T}] - T|/T$.

%CV$[\hat{T}\ ] = 100 \times \sqrt{Var[\hat{T}]}/E[\hat{T}]$. (The percentage coefficient of variation of the estimator $\hat{T}$.)

%RMSE$[\hat{T}\ ] = \left(\sqrt{Var[\hat{T}] + \text{Bias}[\hat{T}]^2}\right)/T$. (The root mean square error of the estimator $\hat{T}$, as a percentage of T.)

%Cover$[\hat{T}\ ] = 100 \times Pr\{$T is contained in its estimated confidence interval $\widehat{CI}[\hat{T}]\}$. This is the percentage coverage probability of the interval estimator for T.

Given estimates $\hat{T}_s$ $(s = 1 \ldots S)$ obtained from S simulations, $E[\hat{T}]$ and $Var[\hat{T}]$ are estimated as follows:

$$\widehat{E}[\hat{T}] = \frac{1}{S}\sum_{s=1}^{S}\hat{T}$$

$$\widehat{Var}[\hat{T}] = \frac{1}{S}\sum_{s=1}^{S}(\hat{T} - E[\hat{T}])^2$$

%Cover$[\hat{T}]$ is estimated by the percentage of the S interval estimates $\widehat{CI}_s[\hat{T}]$ $(s = 1 \ldots S)$ which contain T.

The number of simulations, S, was set to 500. While this may not be sufficient to ignore Monte carlo variation in the quantities of interest, the simulations are extremely time-consuming and conducting more simulations was not feasible. As aresult, the estimates of $E[\hat{T}]$, $Var[\hat{T}]$ and Cover$[T]$ are treated as if

they were the true values in this analysis. While this is not ideal, it is nonetheless informative. Further, given this relatively high value for S, the standard $S/(S-1)$ multipier to correct for bias in $Var[\hat{T}]$ has been ignored.

8.6 The Estimators

8.6.1 Point Estimators of N

The three sorts of abundance estimator tested are $\hat{N}_H$, $\hat{N}_\omega$ and $\hat{N}_\mu$. The detection probabilities are in all cases estimated by logistic regression (using the ungrouped data). The estimator $\hat{N}_\mu$ uses a Beta probability density function to model the density of $p_.(x,z)$. The Beta density for p is defined as follows.

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \qquad 0 \le p \le 1 \tag{8.2}$$

This provides a flexible two-parameter model for random variables on $(0,1)$, which can take on a wide variety of shapes. The actual densities generated by the simulation model are shown in Figure 8.3 and some appropriate shapes which the Beta density can take are shown in Figure 8.4. While the actual p-densities are not as smooth as the Beta densities, they follow the broad range of shapes which the Beta density function can take. Further, remembering that it is the estimated **mean** of the Beta density that is used in $\hat{N}_\mu$, the Beta density seems *a priori* a reasonable model.

$\hat{N}_\mu$, as implemented here, incorporates a constraint on the first parameter of the Beta density, and uses the binned likelihood to estimate the Beta density parameters. (Details, and the reasons for doing this are given in Chapter 9.)

$\hat{N}_H$ and $\hat{N}_\omega$ were implemented with three different forms of data truncation. The reasons for this are discussed in Chapter 9. For the moment note that the truncations are as follows.

(1) **"Untruncated"**: Truncation only at $x = 1$, the maximum simulated perpendicular distance.

(2) **"t-truncation"**: Truncation at the maximum duplicate x, i.e. the maximum x of animals detected by both observers. Call this value x_{3max}. The abundance estimate obtained from this truncation is divided by x_{3max} to give an abundance estimate for the strip of width 1 (comparable to the abundance estimate with truncation at $x = 1$). (The abundance estimators when "t-truncation" is used are denoted $\hat{N}_{Ht}$ and $\hat{N}_{\omega t}$ respectively.)

(3) **"T-truncation"**: Truncation at x_{3max} **and** at the minimum duplicate z (call it z_{3min}). The abundance estimate obtained from this truncation is divided by x_{3max} to correct for the x- truncation, but no correction is made for the z-truncation because the density of z is unknown to the estimator. (The abundance estimators when "T- truncation" is used are denoted $\hat{N}_{HT}$ and $\hat{N}_{\omega T}$ respectively.)

(The rationale for the notation is as follows. Subscripts "T" or "t" are for "truncation" and lowercase "t" indicates less severe truncation than uppercase "T".)

8.6.2 Interval Estimators

Only in the cases of $\hat{N}_H$ and $\hat{N}_\omega$ were the asymptotic variance and the corresponding confidence intervals for N estimated using the information-matrix based approach given in Chapter 7. As has been noted several times, this method of estimation relies on strong independence assumptions which are unlikely to be met in practice, and are therefore of limited practical value in real LT surveys. However, asymptotic interval estimators are sometimes used with relatively small sample sizes and it would be useful to have some idea of how quickly these estimators converge to their asymptotic values as sample sizes increase.

Analytic estimators of variance and confidence intervals for N are not available for $\hat{N}_\mu$ or for any of the truncated versions of the other two estimators. (No analytic estimators which incorporate the variance due to choosing the truncation point are available.) In addition, variance and interval estimators which are not based on the assumption of independent detections are more appropriate for real LT surveys in which there is usually some spatial correlation in detections. The bootstrap estimation methods of Chapter 7 will be more applicable to real LT surveys in which the assumption of independent distribution of animals in the surveyed area may not hold. However, calculation of coverage probabilities with bootstrap interval estimation is extremely computationally expensive. This method has therefore been used to estimate variance and obtain confidence interval estimates here for N using $\hat{N}_\mu$ and $\hat{N}_{\omega t}$ only. (The reason for choosing the latter is that it was judged to perform best among the Horvitz-Thompson–like estimators in many situations; see Chapter 9.)

In a real LT survey (as opposed to simulated one), it is advisable to use transects as bootstrap resampling units for variance and confidence interval estimation, because this involves weaker independence assumptions than does resampling individual animals or groups. In this simulation study, however, only the case in which animals are distributed independently of one another in space (and with respect to z) is considered. The bootstrap was therefore implemented using individual animals as sampling units.

Two sorts of interval estimator are considered, as follows. ("$\hat{N}$" refers to whichever of the estimators is under consideration at the time.)

- **Normal 95% Confidence Interval**: $(\hat{N}-c;\ \hat{N}+c)$, where $c = 1.96\sqrt{\widehat{Var}[\hat{N}]}$ and $\widehat{Var}[\hat{N}]$ is the estimated variance of $\hat{N}$.
- **Percentile Method 95% Confidence Interval**: $(\hat{N}_{0.025}\,;\ \hat{N}_{0.975})$ where $\hat{N}_{0.025}$ and $\hat{N}_{0.975}$ are the 2.5% and 97.5% bootstrap percentiles, respectively. Three hundred bootstrap samples were generated to obtain these estimates. Generating more bootstrap samples was infeasible in the time available.

Chapter 9

Simulation Results

9.1 Overview

In this Chap ter I present and compare the results of the simulation experiment with regard to point and interval estimation of N. I deal with point estimation in the first section and interval estimation in the second.

9.1.1 Point Estimation of N

Twenty seven different surveys were simulated (Models 1a through 3c, each at three different N's) and seven estimators were applied to each simulation ($\hat{N}_H$ and $\hat{N}_\omega$ - each with no truncation, t–truncation and T–truncation, as well as $\hat{N}_\mu$). For each of these 189 {simulation; estimator} sets, three different statistics are considered (%bias, %cv and %RMSE). It is difficult to digest this many results at once, so I present them in the following more–or–less heirarchical fashion.

I consider the untruncated estimator $\hat{N}_H$ first. The results for this estimator indicate a serious problem when sample size is small and mean detection probability is low. Investigation of the causes of the problem (which provides the motivation for "t-truncation" and "T-truncation") is followed by a comparison of the properties of the "H-estimators" ($\hat{N}_H$, $\hat{N}_{Ht}$ and $\hat{N}_{HT}$). Then I consider a comparison of the properties of the "ω-estimators" ($\hat{N}_\omega$, $\hat{N}_{\omega t}$ and $\hat{N}_{\omega T}$). Truncated forms of the estimators are found to improve performance in most cases where the %bias or %cv of the untruncated estimator is large. The final set of comparisons is between $\hat{N}_{\omega T}$, $\hat{N}_{\omega t}$ (which were judged to perform best among the Horvitz-Thompson–like estimators) and $\hat{N}_\mu$.

In an attempt to make the Tables of results easier to digest I have highlighted some of the entries. When the results for a single estimator only are tabulated, entries which have "high" bias or %cv are shown in bold text. For this purpose "high" bias is defined as more than 20% (positive or negative bias), and "high" %cv as more than 25%. These definitions are fairly arbitrary, but they do help structure the Tables and the discussion. %RMSE is used only to compare different estimators, so that Tables of %RMSE's are not presented when one estimator only is under consideration. When estimators are being

compared to one another in a single Table, the lowest %bias, lowest %cv and lowest %RMSE for each simulation are shown in bold text. In all Tables, numbers are rounded to the nearest integer. Numbers are displayed in bold on the basis of their value before rounding.

9.1.2 Interval Estimation of N

Analytic Interval Estimation

The first set of simulation experiments to investigate interval estimation for N is conducted for the estimators $\hat{N}_H$ and $\hat{N}_\omega$ using the analytic estimators of the asymptotic variance of equations 7.46 and 7.49. There are two reasons for performing this set of simulation experiments despite the fact that truncated forms of the H–estimator and the ω–estimator prove to be better point estimators than the untruncated forms, and despite the fact that the independence assumption implicit in the analytic estimators of variance is unlikely to be met in most real LT surveys. Both reasons relate to work in the published literature. The first is to extend the simulation results of Alho (1990) to a LT situation. Alho (1990) considered only $\hat{N}_H$. He found the that estimated 95% confidence intervals for N based on the analytic variance estimator (equation 7.46) had a true coverage of between 93% and 97%. However Alho's (1990) simulations were conducted in the context of MR experiments, and he used univariate detection functions with a normally distributed explanatory variable. The mean detection function for both observers ($p_{..}$) is 0.92 in the case of his first model, and 0.55 in the case of his second. These mean detection probabilities are not very different from those for some of the models simulated in this thesis, but the models themselves are different. How does the coverage of the confidence interval in a LT scenario compare to that evaluated by Alho (1990)? Since it can be done relatively easily, it seems worth investigating here how the analytic interval estimators perform in a LT context when x is not the only explanatory variable.

The second reason for performing this first set of simulations is to investigate the effect of neglecting the component of variance due to estimation of the detection probabilites, on interval estimators of N. This is of interest is that $\hat{N}_H$ has been applied in shore–based surveys (Buckland *et al.*, 1993b; Laake *et al.*, 1995) using interval estimators which neglect this component of variance (i.e. using the second term only on the RHS of equation 7.46 for variance estimation). Do the resulting nominal 95% confidence intervals have true coverage substantially less than 95%?

Although only $\hat{N}_H$ appears in the literature, it seems worth comparing the performance of the analytic interval estimators for $\hat{N}_\omega$ to those of $\hat{N}_H$, given that $\hat{N}_\omega$ is found to perform better than $\hat{N}_H$ in the simulations conducted to investigate point estimation. Simulations were therefore conducted using each of the estimators, both with and without the component of variance due to estimation of the detection probabilities.

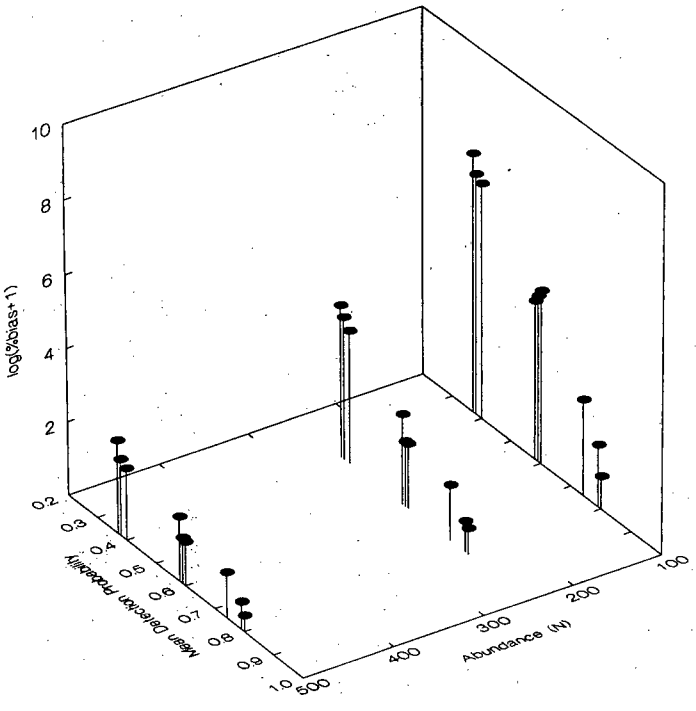
Bootstrap Interval Estimation

One of the motivations for using bootstrap CI estimation rather than analytic asymptotic methods is to avoid the strong independence assumption (that all detections are independent) implicit in the analytic methods. In the simulations conducted here, however, all detections are in fact independent so that these simulations cannot be used as an indicator of how well the bootstrap method copes with failure of the independence assumption. In the case of $\hat{N}_{\omega t}$, the simulations are used to compare (i) the performance of the bootstrap method in an ideal situation where the independence assumptions hold, but where sample sizes may be below those where the asymptotic variance of the estimators apply, and (ii) the analytic CI estimators which are based on the assumption that asymptotic variances apply. In the case of $\hat{N}_{\mu}$, for which no asymptotic variance estimator was developed in this thesis, the bootstrap method is the only proposed CI estimation method.

9.2 Point Estimation Results

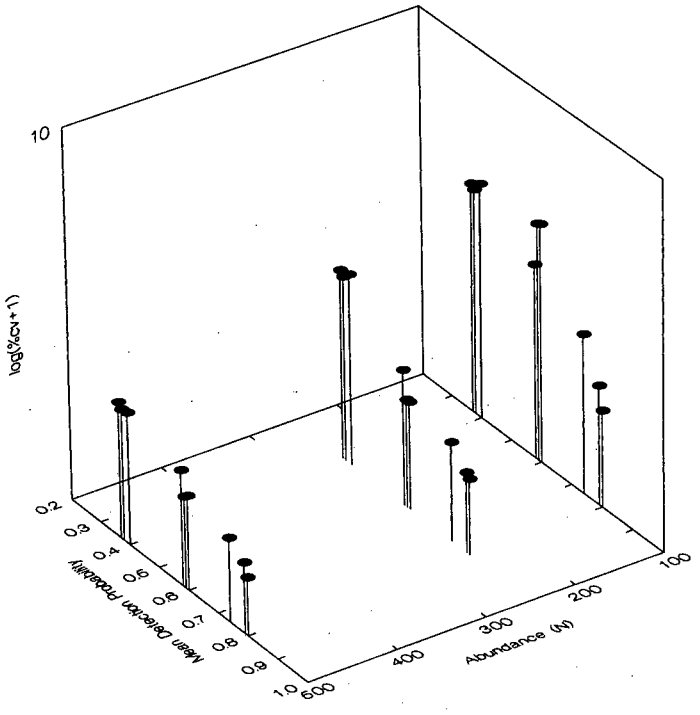
9.2.1 The %Bias and %cv of $\hat{N}_H$

Figure 9.1: $\hat{N}_H$: log(%Bias+1)



The %bias of the untruncated estimator $\hat{N}_H$ is shown in Table 9.1 and Figure 9.1. (The number of simulations carried out seems sufficient to justify reference to "bias" rather than "estimated bias".) The percentage coefficient of variation of the estimator is shown in Table 9.2 and Figure 9.2. Because variation

Figure 9.2: $\hat{N}_H$: log(%cv+1)



in %bias and in %cv over the various simulations is enormous, log(%bias+1) and log(%cv+1) rather than %bias and %cv are shown in the figures.

Table 9.1: $\hat{N}_H$: Percentage Bias

	Simulation Model								
	1			**2**			**3**		
N	a	b	c	a	b	c	a	b	c
100	**652**	**715**	**1174**	**77**	**94**	**115**	13	4.6	1.5
250	**38**	**49**	**64**	11	5.1	5.1	3.7	1.4	1.1
500	6.3	7.3	12	5.0	2.6	2.4	2.6	1.2	0.6

There are two striking features of the %bias results; they are always positive, and can be extremely large. In addition, the %cv is extremely large in the region that %bias is extremely large. Both decrease substantially from Model 1 to Model 3 (as mean detection probability for both observers increases from about 0.4 to about 0.8), and as N increases from 100 to 500. The effect of the level of heterogeneity is substantially smaller than the effect of either Model or N. The %bias approaches zero as total and duplicate sample sizes get larger. This is expected from the asymptotic theory. Except in the case of the model with highest detectabilities (Model 3), the %bias is "high" whenever the total sample size falls

below about 100. The %cv becomes "high" at similar sample sizes. This is worrying, as sample sizes this small are not uncommon in practice.

Table 9.2: $\hat{N}_H$: Percentage Coefficient of Variation

	Simulation Model								
	1			2			3		
N	a	b	c	a	b	c	a	b	c
100	**587**	**446**	**499**	**194**	**617**	**643**	**76**	25	13
250	**178**	**149**	**168**	**39**	18	18	14	8	7
500	**35**	**34**	**39**	22	11	12	9	6	4

To investigate the causes of the extreme biases and variances, x, z, the estimated detection probability ($\hat{p}_{..}(x,z)$), and the true (as simulated) detection probability ($p_{..}(x,z)$) for each detected animal were stored during simulation and estimation. Examination of these data reveal the mechanism causing the positive bias.

Figure 9.3: Example of the Small Sample Size Regression Problem. Column headings refer to the horizontal axis; row headings refer to the vertical axis.

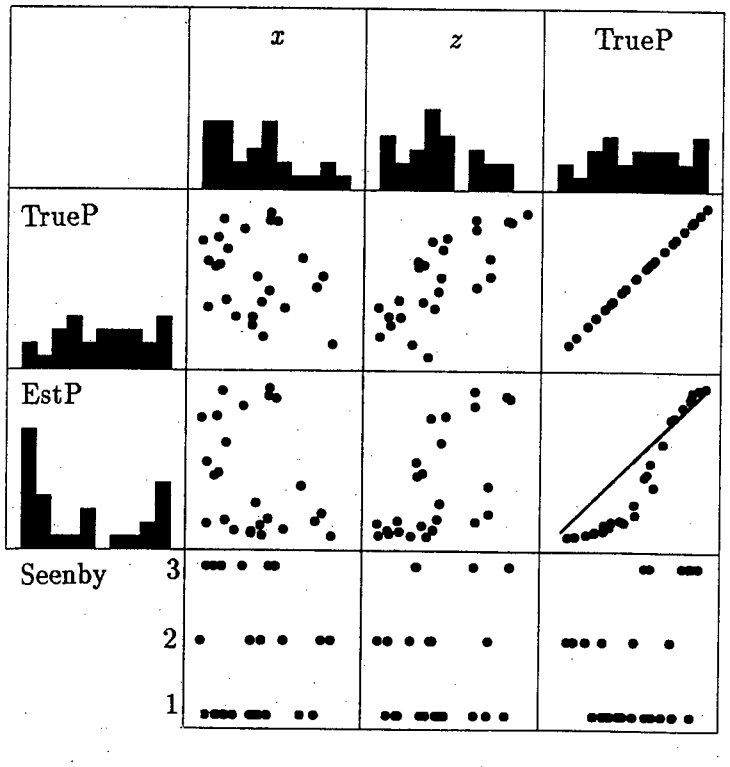


Figure 9.3 illustrates the problem and the mechanism. It is a fairly extreme example taken from one of the simulations of Model 1a with $N = 100$. The simulation resulted in a total sample size of 25 with only 6 duplicate detections. The columns correspond to observed x, observed z and the true detection probabilities of the detected animals ("TrueP"). The rows correspond to the true detection

probabilities of the detected animals ("TrueP"), the estimated detection probabilities of the detected animals ("EstP"), and the detection/capture type ("Seenby": 1=observer 1 only, 2=observer 2 only, 3=duplicate detections). The histograms are relative frequencies of detections with respect to each of these quantities.

Consider first the plot of "Seenby" *vs* "TrueP". All the duplicate detections are of animals with high true detection probabilities. Similarly, there is a dearth of duplicates at high x and at low z (see the plots of "Seenby" *vs* x and "Seenby" *vs* z). As a result there is very little information on how low the true detectabilities of the least detectable animals are, other than that they are low. With no duplicates in the low- detectability region, the estimated probability of detection tends towards zero here, under-estimating true detectability when it is small (see the plot of "EstP" *vs* "TrueP"; the diagonal line is "EstP"="TrueP") and over- estimating true detectability when it is large. As a result the estimator of N is positively biased. (Because estimated detection probability enters the abundance estimator as an inverse, a small negative bias in estimating detection probability when true detection probability is small, translates into a large positive bias in the contribution to the estimate of N from such an observation. Estimated detection probabilites in the region of high true detection probability are above the line "EstP"="TrueP". However, when detection probability is close to 1, positive bias in estimating this has relatively little effect on the abundance estimator.)

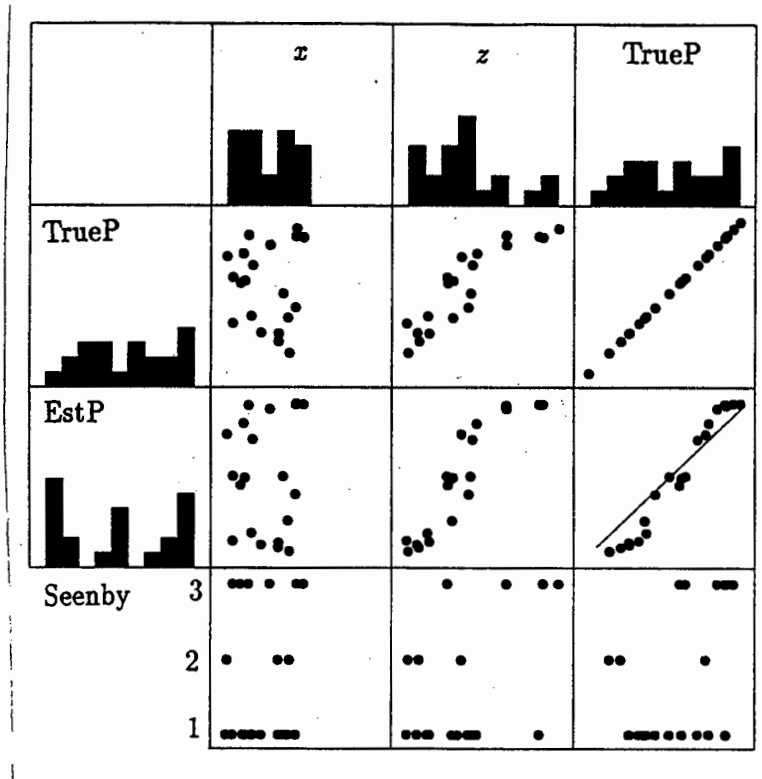
The behaviour of the estimator is a manifestation of a problem which is not uncommon in logistic regression. It is a consequence of the fact that for finite samples, the bias of the estimated slope parameters of the logistic regression model is infinite. (See Hastie and Tibshirani, 1990, p164, for example.) Consider a logistic regression with one explanatory variable, x. There is a non-zero probability that the "successes" and the "failures" will be completely separated in the x-dimension. When they are separated in this way, the estimated slope parameter for x ($\hat{\beta}_1$, say, where the detection function is $p(x) = \exp\{\beta_0 + \beta_1 x\}/(1 + \exp\{\beta_0 + \beta_1 x\})$ will tend to $-\infty$ (presuming the detection function is decreasing in x). The estimated detection probabilities at large x will tend to zero, and their inverses to ∞. The smaller the sample size, the greater the likelihood of this sort of separation of "successes" and "failures", and the more severe the problem of bias in practice.

One "fix" for the problem evidenced in Figure 9.3 is to truncate data beyond what in the current context is the duplicate observation with lowest $\hat{p}_{..}(x,z)$. As far as abundance estimation is concerned, this must necessarily lower $E[\hat{N}_H]$. It will result in a negatively biased estimator of N if $\hat{N}_H$ is unbiased. This sort of truncation is referred to as "T-truncation" here. An alternative, less severe sort of truncation is available in a LT context when the density of x is "known". This is to truncate the data at the largest observed x of duplicate detections, and to adjust for this fact in abundance estimation by shifting W to be equal to this x (call it x_{3max}). (This simply reduces the width of the covered area and does not introduce bias.) I have called this sort of truncation "t-truncation". With "t-truncation" and the assumption of a uniform distribution for x, N is estimated by $\tilde{N}_H \times (W/x_{3max})$, where W is the truncation distance before truncation and $\tilde{N}_H$ is the estimate obtained from the truncated data. (Note that when "T-truncation"

is used, the same correction is applied; but because this corrects only for truncation in the x-dimension and not in the z-dimension, the expectation of the estimator of N will be lower than it would be without truncation. If the distribution of z, was known a correction for truncation in the z- dimension could be applied.)

Figure 9.4 (which is the equivalent of Figure 9.3, but with "t-truncation" applied) illustrates the effect of "t-truncation" on the estimation of detection probabilities. The truncation is evident from the x-column of the Figure.

Figure 9.4: Example of The Effect of "t-truncation"



Comparing the plots of "TrueP" *vs* "EstP" in Figure 9.4 and Figure 9.3, it is clear that "t-truncation" has improved estimation of the detection probabilities; the points are on average substantially closer to the line "EstP"="TrueP". (Note that in the most extreme case in one dimension, where x is the only explanatory variable, the "successes" and "failures" are completely separated; in this case "t-truncation" will not yield an estimate of the slope parameter which approaches $-\infty$. Thus the data will be insufficient for estimation of the slope parameter and the estimated detection probability without the slope parameter will be unity.)

Incidentally, an interesting feature of the results in Figure 9.3 is contained in the plots of "TrueP" *vs* x and "TrueP" *vs* z. It is apparent from these that the correlation between true detection probability and z is higher than the correlation between true detection probability and x. That is, with Model 1a the "heterogeneity" explains more of the variation in detection probability than does perpendicular distance. In a line transect context, in which perpendicular distance is traditionally "**the**" explanatory variable, this corresponds to fairly severe heterogeneity. Similar plots for the "low-heterogeneity" models (Models

1c, 2c, 3c) reveal x to be substantially more highly correlated with true detection probability than z in these cases. The levels of heterogeneity for the simulation experiments were selected on the basis of what appeared "by eye" to span as wide a range of levels of heterogeneity as one might expect to encounter in practice. Although there are scant real data available to verify whether this has been achieved, the fact that the range includes both cases in which x is clearly the most significant explanatory variable, and also cases in which z explains more of the variation in detection probability than x, does suggest that the range is wide enough to cover most cases that might be encountered in practice. Of course, all the above is predicated on the range of x being sufficient for detection probabilites to vary as much as they might be expected to vary in practice; one can reduce the correlation between x and detection probability by restricting the range of x sufficiently. However, the choice of W was such that the range of x's and $p_{i.}(x)$'s used in the simulation experiment are fairly typical of real LT datasets. The other point to bear in mind is that the correlation evidenced in the plots is subject to sampling variation, so that one should be cautious not to over–interpret it. Nonetheless, it is reassuring that on the limited evidence produced by the plots referenced above for the various models, the simulated heterogeneity seems to span a reasonable range.

Truncated forms of $\hat{N}_H$

A comparison of the properties of the untruncated, "t"-truncated, and "T"- truncated "H-estimators" is given in Tables 9.3, 9.4 and 9.5. (Aside from the Figures and Tables in the preceeding sections, I have put all Figures and Tables in Appendices to this Chapter rather than in the body of the text. This is because the Tables and Figures are large and many, and would interfere with the continuity of the text if they were in the Chapter.)

Both $\hat{N}_H$ and $\hat{N}_{Ht}$ are positively biased (in all except three cases of negligible negative bias), extremely so when mean detection probability and sample size are small. The absolute value of the %bias of $\hat{N}_{Ht}$ is almost always less than that of $\hat{N}_H$ (in the two cases in which it is not, the difference is negligible). $\hat{N}_{HT}$ is consistently negatively biased, more so with higher levels of heterogeneity (i.e. when σ is larger). Unlike the other two estimators, however, it is never extremely biased (the highest absolute %bias is 25%). The absolute value of the %bias of $\hat{N}_{Ht}$ is less than that of $\hat{N}_{HT}$, except for the 5 cases with smallest total and duplicate sample sizes (and Models 3b and 3c with $N = 500$, but the difference there is negligible).

In terms of %cv, $\hat{N}_{HT}$ performs best; in only 4 of the 27 cases is there an estimator with lower %cv. However, because the means of different estimators vary substantially for the same simulated dataset, %cv alone is not a very useful measure of estimator performance. A more reasonable measure is %RMSE (which is equal to $\sqrt{bias^2 + se^2}/N$). $\hat{N}_{HT}$ still performs best overall in these terms; in only 4 of the 27 cases is there an estimator with lower %RMSE than $\hat{N}_{HT}$. In all cases this other estimator is $\hat{N}_{Ht}$, and when sample size and mean detection probabilities are not small, but there is substantial heterogeneity, $\hat{N}_{Ht}$ is arguably the best estimator because of the sometimes substantial negative bias of $\hat{N}_{HT}$ when

heterogeneity is large. In most other conditions, $\hat{N}_{HT}$ seems preferable. As an aside, note that $\hat{N}_{HT}$ is a conservative estimator of animal abundance in that it tends to be negatively biased; from a management point of view this is conservative because the lower the estimated abundance, the lower the planned mortality due to human intervention will be under harvesting strategies which fix the proportional level of the take.

The performance of $\hat{N}_H$ is worse than that of both $\hat{N}_{Ht}$ and $\hat{N}_{HT}$. This indicates that the performance of the Horvitz–Thompson–like estimator in a LT context is improved by truncating the data to a greater extent than one would normally truncate LT data from a single platform. (Recall that the simulations were constructed so that the truncation distance $W = 1$, called the "untruncated" case here, corresponds to a truncation approximately equal to that recommended for conventional LT data, in which $g(0) = 1$.)

9.2.2 The "ω-Estimators": $\hat{N}_\omega$, $\hat{N}_{\omega t}$, $\hat{N}_{\omega T}$

Given the results from the H-estimators in the preceding section, one would *a priori* expect both the variance and the bias of $\hat{N}_\omega$ to be lower than that of $\hat{N}_H$. (Because estimated detection probability is averaged with respect to an assumed $\pi(x)$, and extremely low estimated probabilities will occur less frequently after such averaging than in its absence.) These expectations are borne out by the simulation results presented in detail in Figures 9.5, 9.6 and 9.7 (showing the %bias), Figures 9.8, 9.9 and 9.10 (showing the %RMSE) and Tables 9.6, 9.7 and 9.8. (These Tables should be compared with Tables 9.3, 9.4 and 9.5, respectively.)

The conclusion from these sets of comparisons is clear. The use of $\pi(x) = W^{-1}$ in estimation (i.e. using ω–estimators instead of H–estimators) of N improves the performance of the each of the Horvitz–Thompson–like estimators in terms of both %bias and %RMSE, with one exception. This is that the absolute value of the %bias of $\hat{N}_{\omega T}$ is higher than that of $\hat{N}_{HT}$, although the %RMSE of $\hat{N}_{\omega T}$ is less than that of $\hat{N}_{HT}$. (T–truncation introduces negative bias since no correction is made for truncation in the z-dimension. The t–truncated H–estimator is more positively biased than the t-truncated ω–estimator, and the negative bias due to truncation in the z-dimension is counteracted to a greater degree in the case of the H- estimator, making the $\hat{N}_{HT}$ less biased than $\hat{N}_{\omega T}$.)

Comparisons of the performance of the various ω-estimators are presented in Tables 9.6, 9.7 and 9.8. Figures 9.11 and 9.12 compare the %bias and %RMSE of $\hat{N}_{\omega t}$ and $\hat{N}_{\omega T}$. Among the ω-estimators, $\hat{N}_{\omega T}$ is least biased when mean detection probability is small and $N = 100$. Over the remainder of the simulations $\hat{N}_{\omega T}$ is the most biased, while $\hat{N}_{\omega t}$ is slightly less biased than $\hat{N}_\omega$.

$\hat{N}_{\omega T}$ has lowest %RMSE when $N = 100$. (Actually $\hat{N}_{\omega t}$ has lower %RMSE in one case, namely Model 3c with $N = 100$, but the difference is negligible given that with 500 simulations the results will not be entirely free of Monte-Carlo variation.) As N and the mean detection probability of the simulation models increases, $\hat{N}_{\omega t}$ is increasingly the estimator with lowest %RMSE. The point at which $\hat{N}_{\omega t}$ becomes estimator with lowest %RMSE is higher (i.e. higher mean detection probability and higher N) when

heterogeneity (σ) is large than when it is small. That is, when heterogeneity is high, $\hat{N}_{\omega t}$ performs better in terms of %RMSE than $\hat{N}_{\omega T}$ only when mean detection probabilities and N are relatively high, whereas when heterogeneity is low $\hat{N}_{\omega t}$ performs better at relatively low mean detection probabilities and relatively low N. This behaviour can readily be explained in terms of the negative bias in estimated probabilities at low detection probabilities. When heterogeneity is high very small detection probabilities occur more frequently than when heterogeneity is low. Positive bias in abundance and variance estimation as a consequence of negatively biased detection probability estimates is more severe the smaller the estimated detection probabilities. Thus T-truncation, which removes the lowest estimated detection probabilities from the abundance estimator, has more effect when heterogeneity is high.

The untruncated estimator, $\hat{N}_{\omega}$, tends to perform better than either of the other two estimators when the expected duplicate sample size is biggest. This is not surprising because as expected duplicate sample size increases, the probability of any individual logistic regression estimating detection probability to be substantially smaller than it is in the region of low detectability, is reduced. This in turn reduces the need for truncation, and since truncation necessarily reduces sample size, an untruncated estimator would be expected to perform better than a truncated estimator as sample size approaches infinity.

Given the relative performance of $\hat{N}_{\omega}$, $\hat{N}_{\omega t}$ and $\hat{N}_{\omega T}$, it is difficult to select one as having the best overall performance. With large sample size and relatively large mean detection probability, $\hat{N}_{\omega}$ or $\hat{N}_{\omega t}$ would be the preferred estimator. With small sample size and relatively small mean detection probability $\hat{N}_{\omega T}$ would be preferred. The simulation results show $\hat{N}_{\omega}$ to be inferior to both $\hat{N}_{\omega t}$ and $\hat{N}_{\omega T}$, except when expected duplicate sample sizes are relatively large. This corroborates the results from the H-estimators regarding truncation to some extent. Except when expected duplicate sample sizes are relatively large, the performance of Horvitz–Thompson–like estimators in a LT context is improved by truncating the data to a greater extent than one would normally truncate LT data from a single platform. More specifically, it is improved by truncating in one or more dimensions with respect to the duplicate detection with lowest estimated detection probability. This is more the case with the H–estimators than the ω–estimators, and this difference can be explained in terms of the stabilising effect that averaging over x has on estimated detection probabilities. Although the ω–estimators clearly perform substantially betterthan the H–estimators overall, the results are inconclusive regarding which ω–estimator is best overall.

9.2.3 $\hat{N}_{\mu}$

Recall that this estimator assumes a Beta density for $\pi(\hat{p}_{..})$, and that the parameters of the Beta density must be estimated. Initial attempts to implement the estimator without constraining the first parameter of the Beta density (the parameter α of equation 8.2, which determines the intercept at 0) were unsuccessful. The parameter α frequently tended to 0 and the numerical minimization routine failed to converge for small sample sizes. This happened in association with extreme positive bias in estimating N and extremely large variances. This behaviour can be explained in terms of bias in estimating $p_{..}(x,z)$; the increasingly negative bias in $\hat{p}_{..}(x,z)$ as $p_{..}(x,z)$ decreases, translates into increasingly positive bias

in estimated density at $\hat{p}_{..}(x,z)$. This in turn tends to make the Beta density diverge as $\hat{p}_{..}(x,z) \to 0$. (See the histogram of estimated detection probabilities in Figure 9.3, for example.)

Constraining α so that the Beta density can't become infinite at 0 (i.e. requiring $\alpha \geq 1$) overcame this problem. The constraint is *ad hoc* to the extent that there is really no theoretical reason that the Beta density function for detection probabilities should not diverge as $\hat{p}_{..}(x,z) \to 0$. This does not mean that it is not a useful constraint, and it can be regarded as a smooth form of truncation. By comparison, "t"-truncation and "T"-truncation effectively introduce a discontinuity into the density of $p_{..}(x,z)$: the density jumps from zero to some non-zero value at the truncation point, giving observations with values less than the truncation point zero influence in determining the (empirical) density functions. The constraint above on α effectively downweights the influence of estimates of $p_{..}(x,z)$ close to zero, increasingly so as the proportion of estimates close to zero increases. As such, it is a potentially useful means of improving the robustness of the μ-estimator of abundance. Viewed as a mechanism for downweighting the influence of small estimated detection probabilities on the estimate of the Beta density, the constraint on α has the undesirable feature that the form and extent of the downweighting is not explicit. Thus the constraint remains somewhat *ad hoc*. .

As indicated in Chapter 7, there are two ways of implementing the estimator, namely on grouped or ungrouped $\hat{p}_{..}(x,z)$ data. Because of occasional problems with convergence for the ungrouped estimator, only the estimator using grouped data was implemented. Five groups of equal width (0.2) on the $\hat{p}_{..}(x,z)$-axis were used.

The results of the simulations are shown in Table 9.9. The %bias and %RMSE are also illustrated in comparison to that for $\hat{N}_{\omega t}$ in Figures 9.13 and 9.14. Three notable aspects of these results are as follows.

- $\hat{N}_{\mu}$ is relatively robust with respect to small sample sizes and small mean detection probabilities; $\hat{N}_{\mu}$ remains remarkably unbiased, with reasonably small %cv, in regions of the simulation space in which the ω-estimators and the H-estimators became extremely biased and highly variable.
- $\hat{N}_{\mu}$ tends to have greater negative bias the greater the heterogeneity of the simulation model.
- The %bias for Model 1a in particular is remarkably insensitive to sample size. The insensitivity of the estimator tends to be greater the greater the heterogeneity (sensitivity tends to decrease as one moves from "(c)" to "(a)" models) and the lower the mean detection probability (i.e. sensitivity tends to decrease as one moves from Model 3 to Model 1 for any given level of heterogeneity).

Taken together, these results indicate that the constraint on the Beta density parameter tends to make $\hat{N}_{\mu}$ an inconsistent estimator of N; the constraint is encountered more often when heterogeneity is high and/or mean detection probability is low, and in these cases the absolute value of the %bias of the estimator tends not to decrease as sample size increases (i.e. as one moves from simulated $N = 100$ to simulated $N = 500$). When mean detection probability is high and sample size is large, the constraint is encountered infrequently and in these cases the estimator appears to be consistent for N.

9.2.4 $\hat{N}_{\omega T}$ *vs* $\hat{N}_{\omega t}$ *vs* $\hat{N}_{\mu}$

From the comparisons above, the ω–estimators emerge as superior to the H–estimators. Except when expected duplicate sample sizes are relatively large, $\hat{N}_{\omega t}$, and $\hat{N}_{\omega T}$ perform better than $\hat{N}_{\omega}$. As it is primarily the small-sample properties of the estimators that is of interest here, I do not consider $\hat{N}_{\omega}$ further in the remaining comparisons between estimators.

A comparison of the results for $\hat{N}_{\mu}$ with those for $\hat{N}_{\omega T}$ suggests that the former performs better than the Horvitz–Thompson–like estimator $\hat{N}_{\omega T}$ when N is small and mean detection probability is low (Tables 9.10, 9.11, 9.12 and Figures 9.15 and 9.16).

A comparison between the %bias and %RMSE of $\hat{N}_{\omega t}$ and $\hat{N}_{\mu}$ is shown in Figures 9.13 and 9.14. When $N = 100$ (i.e. small sample sizes), $\hat{N}_{\mu}$ performs better than $\hat{N}_{\omega t}$ in terms of both %bias and %RMSE, except in the case of Model 3 with moderate to low heterogeneity. With $N = 250$ $\hat{N}_{\omega t}$ performs slightly better in terms of %bias. In terms of %RMSE, its performance improves relative to that of $\hat{N}_{\mu}$ as heterogeneity decreases. (It performs worse when heterogeneity is high and better when it is low.) The performance of $\hat{N}_{\omega t}$ improves in both absolute terms and relative to $\hat{N}_{\mu}$ when N is increased to 500. Interpretation of these results is easier in the light of the results obtained from estimating the 95% CI's for $\hat{N}_{\mu}$, so I defer discussion until these are presented (below).

9.3 Interval Estimation of N

9.3.1 Analytic Estimation and Neglect of Variance due to Estimating $p_{\cdot}(x,z)$

The actual coverage probabilities (as evaluated by simulation) for the 95% confidence interval estimator based on the analytic estimator of variance of Huggins (1989) and Alho (1990) for their estimator ($\hat{N}_H$) under the assumption of normality of $\hat{N}_H$ are shown in Figure 9.17 (dotted lines). Also shown are the coverage probabilites for the estimator when the variance due to estimation of the detection probabilities is neglected (solid lines).

The Figure shows clearly that estimation of $p_{\cdot\cdot}(x,z)$ contributes a substantial proportion to the variance of the estimator, and that confidence intervals obtained when this component of variance is neglected are substantially narrower than the nominal 95% in all cases (being as low as 50% in the most extreme case). As might be expected, the actual coverage when this component of variance is neglected, is closer to 95% when N is largest, but even in this case the nominal 95% confidence interval rarely has true coverage probability more than 80%. Neglect of variance due to estimation of detection probabilities leads to markedly over-optimistic confidence interval estimates. The same is true in the case of $\hat{N}_{\omega}$ (Figure 9.18).

The estimated 95% CI's have true coverage not far from 95% when the variance due to estimation of the detection probabilites is included, with the intervals being slightly too narrow when N is small and slightly too wide when N is large. These results are very much in line with the simulation results obtained by Alho (1990).

9.3.2 Estimated 95% Coverage of $\hat{N}_H$ *vs* $\hat{N}_\omega$: Analytic

The true coverage probabilities (as simulated) of the analytic 95% confidence interval estimators for $\hat{N}_H$ (dotted lines) and $\hat{N}_\omega$ (solid lines) are shown in Figure 9.19. The performance of the two estimators is very similar. Performance is worst at small N ($N = 100$), in which case coverage is less than 95% (around 88% in the worst case).

9.3.3 Estimated 95% Coverage of $\hat{N}_{\omega t}$: Analytic and Bootstrap

Figure 9.20 shows the coverage of the analytic and bootstrap percentile method 95% CI's for $\hat{N}_{\omega t}$. The bootstrap method incorporates variance due to the selection of a truncation point and the analytic method does not. The bootstrap estimator tends to perform better than the analytic method when sample size is small. This is the region in which the analytic estimator is expected to perform worst, since it is based on the asymptotic properties of the estimator. When sample size is large, the bootstrap and analytic estimators have similar coverage probabilities although the bootstrap estimator performs better in this respect when mean detection probability is low (Model 1). In view of these results and the fact that the analytic estimator incorporates stronger independence assumptions than are likely to apply in practice, the bootstrap estimator is the preferred method.

9.3.4 Estimated 95% Coverage of $\hat{N}_\mu$: Bootstrap

Figure 9.21 shows the coverage of the bootstrap percentile method 95% CI for $\hat{N}_\mu$. The coverage of estimated CI's is worse the lower the mean detection probability. Except in the case of Model 3 (with highest mean detection probability), coverage gets markedly worse as heterogeneity increases. Furthermore, coverage deteriorates as N (and hence sample size) increases.

These results are symptomatic of an inconsistent estimator for N. The problem is caused by the constraint placed on the α parameter of the Beta density. The more frequently the constraint is encountered (i.e the greater the heterogeneity and the lower the mean detection probability of the simulated model), the less consistent an estimator $\hat{N}_\mu$ becomes. Increasing the sample size (i.e. moving from simulations with $N = 100$ to those with $N = 500$ here) reduces the variance and the estimated variance of $\hat{N}_\mu$. When $\hat{N}_\mu$ is not consistent for N, this results in the estimated CI's shrinking about a biased estimate of N; hence the coverage of the estimated CI's deteriorates as sample size (and N) increases.

9.4 Conclusion

In drawing conclusions from the simulation study one must keep in mind its limited nature, particularly as regards the following points.

- Animals are assumed to be distributed completely independently of one another, so that the application of analytic CI estimators (which implicitly assume this independence) is defensible. In most

practical applications, this will not be the case.

- Model selection is not addressed and no heterogeneity goes unmodelled in the simulations. Even with appropriate variable selection procedures, it is quite possible that in practice some heterogeneity will be neglected in estimation. This could happen because a variable which is not chosen by whatever variable selection criteria are used, may nevertheless contribute substantial heterogeneity. It could also happen because a variable which contributes substantial heterogeneity is not considered as an explanatory variable for the model.

- Only the case in which there is positive correlation between the detection probabilities of the two observers is considered. In practice, there are various potential sources of negative correlation (see Buckland *et al.*, 1993a, for example).

Having said this, one must also bear in mind that the vast majority of analytic results regarding properties of estimators are obtained on the basis of similar idealizations of the process being modelled. The simulation study results are a useful starting point for understanding the properties of the estimators. The primary conclusions drawn from the study regarding these properties are as follows.

- **The untruncated Horvitz–Thompson–like estimators are extremely positively biased and have extremely high variance when mean detection probability is low and sample size is small.** A symptom of the problem is a concentration of duplicate detections in regions of the explanatory variable space associated with high detection probability, leaving very few duplicates in regions associated with low detection probability. The problem can be alleviated by "t-truncation" in the first instance. If the symptom persists after "t-truncation", it can be removed using "T-truncation", but at the expense of some negative bias (which is typically substantially less in absolute magnitude than the original positive bias).

  It is likely that the positive bias caused by the separation of duplicates in $p_{..}$–space will be less in the presence of heterogeneity which causes negative correlation between detection probabilities of different platforms. In this case duplicate detection probabilities will tend to be higher because animals which are difficult for one observer to detect will tend to be easy for the other observer to detect. As a result, duplicates will tend to be less clustered in $p_{..}$–space, and estimation of $p_{..}(\ )$ will tend to be less biased when $p_{..}(\ )$ is small.

  Since this problem is more severe the smaller the mean detection probabilities in the population (sampled and unsampled), one anticipates that it will be more severe with surveys of species like harbour porpoise, for which mean detection probability on the trackline can be in the region of 0.3 for a single observer team (Borchers *et al.*, 1995) than in any of the simulations presented here (in which the mean detection probability on the trackline is never less than 0.49).

  In applying the estimation methods to real data one should be wary of estimates based on small sample sizes, more so the smaller the estimated mean detection probability. A "small" sample size

is not well defined in this context, but the simulation results do give some very rough indication of what adequately large sample sizes might be for given mean detection probabilities in order for the estimator to be approximately unbiased. This is complicated by the fact that the bias can depend quite strongly on the degree of heterogeneity. Nevertheless, the plots and Tables in the Appendices to this Chapter, read together with the expected sample sizes and mean detection probabilites of Table 8.3, do provide some rough guidance on sample sizes.

- **Using the assumption that $\pi(x) = W^{-1}$ in estimation substantially reduces both the bias and the variance of the Horvitz-Thompson–like estimators.** (The conjecture of Chapter 7 that using information on $\pi(x)$ results in improved estimation precision is borne out by the simulation results.)

- **Except when mean detection probability is high and/or large samples are obtained, more severe truncation (and truncation of a different nature) than is usually advocated for conventional LT survey data appears necessary** to reduce bias in estimating N. At the lowest sample sizes and lowest mean detection probabilities, $\hat{N}_{\omega T}$ performs best, while in other circumstances $\hat{N}_{\omega}$ or $\hat{N}_{\omega t}$ performs best among the Horvitz–Thompson–like estimators.

- **$\hat{N}_{\mu}$ appears to be a robust but sometimes inconsistent estimator.** In terms of %RMSE, it performs better than any other estimator when sample size is small and mean detection probability is low. However, the coverage probabilities of the estimated 95% CI's are very poor except when heterogeneity is low. The constraint placed on one parameter of the Beta density function used to model the pdf of detection probabilities is somewhat *ad hoc*, and the results indicate that the estimator is consequently not consistent for N, except when heterogeneity is low and/or mean detection probability is high.

- **Neglecting the component of variance due to estimating the detection probabilities results in substantially biased estimates of variance** and in estimated CI's which are substantially too narrow.

- **While they are not exact, the CI estimators based on asymptotic theory have coverage probabilties which are not very far from their nominal levels**, even when sample sizes and mean detection probabilties are relatively small. (Note however, that this is not to say they are the most appropriate estimators to use in practice. Their strong implicit independence assumptions will usually not hold in practice.)

- **The percentile method bootstrap estimator of CI for $\hat{N}_{\omega t}$ performs well when detections are independent.** It has the additional advantage of being applicable (using transects as sampling units) without the strong independence assumptions implicit in the analytic estimation methods.

9.5 Appendices

9.5.1 Tables Comparing the Estimator Properties

Table 9.3: H-Estimators: Percentage Bias

		Simulation Model								
		1			**2**			**3**		
N	Estimator	a	b	c	a	b	c	a	b	c
100	$\hat{N}_H$	652	715	1174	77	94	115	13	4.6	1.5
	$\hat{N}_{Ht}$	225	148	73	67	33	**2.9**	**10**	**2.4**	**0.6**
	$\hat{N}_{HT}$	**-34**	**-20**	**-0.4**	**-25**	**-14**	-8.8	-13	-5.5	-2.3
250	$\hat{N}_H$	38	49	64	**11**	5.1	5.1	3.7	1.4	1.1
	$\hat{N}_{Ht}$	**12**	**8.8**	**-0.5**	**11**	**2.6**	**0.9**	**3.5**	**1.2**	**0.6**
	$\hat{N}_{HT}$	-32	-24	-16	-18	-7.2	-4.4	-8.2	-2.9	-0.8
500	$\hat{N}_H$	6.3	7.3	11.8	5.0	2.6	2.4	**2.6**	1.2	0.6
	$\hat{N}_{Ht}$	**5.7**	**-0.1**	**-1.8**	**4.7**	**1.7**	**0.9**	2.7	**1.1**	0.7
	$\hat{N}_{HT}$	-25	-15	-9.6	-13	-4.2	-2.0	-5.4	**-1.1**	**-0.3**

Table 9.4: H-Estimators: Percentage Coefficient of Variation

		Simulation Model								
		1			**2**			**3**		
N	Estimator	a	b	c	a	b	c	a	b	c
100	$\hat{N}_H$	587	446	499	194	617	643	76	25	13
	$\hat{N}_{Ht}$	694	454	502	322	336	27	45	14	**10**
	$\hat{N}_{HT}$	**584**	**309**	**485**	**66**	**23**	**21**	**17**	**13**	**10**
250	$\hat{N}_H$	178	149	168	39	18	18	14	8	7
	$\hat{N}_{Ht}$	80	112	42	52	17	14	13	**7**	**6**
	$\hat{N}_{HT}$	**32**	**26**	**28**	**17**	**16**	**13**	**10**	8	7
500	$\hat{N}_H$	35	34	39	22	11	12	9	6	**4**
	$\hat{N}_{Ht}$	42	**14**	**15**	26	**10**	**10**	10	**5**	**4**
	$\hat{N}_{HT}$	**20**	17	17	**13**	**10**	**10**	**8**	**5**	**4**

Table 9.5: H-Estimators: Percentage Root Mean Square Error

N	Estimator	Simulation Model								
		1			**2**			**3**		
		a	b	c	a	b	c	a	b	c
100	$\hat{N}_H$	4460	3701	6465	353	1200	1386	88	27	13
	$\hat{N}_{Ht}$	2266	1135	871	544	448	28	51	14	**10**
	$\hat{N}_{HT}$	**518**	**249**	**483**	**56**	**24**	**21**	**19**	**13**	**10**
250	$\hat{N}_H$	248	228	283	44	20	19	15	8	7
	$\hat{N}_{Ht}$	90	122	41	58	17	14	14	**8**	**6**
	$\hat{N}_{HT}$	**39**	**32**	**28**	**23**	**16**	**13**	**12**	8	**6**
500	$\hat{N}_H$	38	37	45	24	11	13	10	6	5
	$\hat{N}_{Ht}$	45	**14**	**15**	38	**11**	10	10	**5**	**4**
	$\hat{N}_{HT}$	**29**	21	18	**17**	11	**10**	**10**	**5**	**4**

Table 9.6: ω-Estimators: Percentage Bias

N	Estimator	Simulation Model 1 a	1 b	1 c	2 a	2 b	2 c	3 a	3 b	3 c
100	$\hat{N}_{\omega}$	131	97	150	38	14	4.6	8.2	1.6	**0.2**
	$\hat{N}_{\omega t}$	109	87	60	46	28	**0.6**	**8.0**	**1.2**	-0.3
	$\hat{N}_{\omega T}$	**-45**	**-25**	**-6.0**	**-27**	**-16**	-10	-13	-6.1	-3.4
250	$\hat{N}_{\omega}$	10	11	8.6	**7.6**	2.7	2.2	**2.6**	0.7	0.5
	$\hat{N}_{\omega t}$	**6.8**	**-0.2**	**-3.4**	8.0	**1.4**	**-0.3**	3.5	**0.6**	**0.1**
	$\hat{N}_{\omega T}$	-36	-24	-16	-18	-7.9	-4.9	-8.4	-3.0	-1.2
500	$\hat{N}_{\omega}$	3.7	1.7	**1.9**	3.1	1.3	1.0	**1.7**	0.7	**0.4**
	$\hat{N}_{\omega t}$	**2.9**	**-1.4**	-2.4	**0.8**	**0.7**	**0.1**	1.9	**0.6**	**0.4**
	$\hat{N}_{\omega T}$	-25	-15	-10	-13	-4.7	-2.6	-5.6	-1.3	-0.5

Table 9.7: ω-Estimators: Percentage Coefficient of Variation

N	Estimator	Simulation Model 1 a	1 b	1 c	2 a	2 b	2 c	3 a	3 b	3 c
100	$\hat{N}_{\omega}$	330	354	627	227	83	22	35	12	9
	$\hat{N}_{\omega t}$	435	425	519	261	341	19	31	11	**8**
	$\hat{N}_{\omega T}$	**36**	**296**	**452**	**21**	**19**	**17**	**14**	**10**	9
250	$\hat{N}_{\omega}$	55	58	29	31	14	11	11	**6**	6
	$\hat{N}_{\omega t}$	55	35	**22**	36	14	**10**	13	**6**	**5**
	$\hat{N}_{\omega T}$	**24**	**23**	**22**	**15**	**12**	**10**	**9**	7	6
500	$\hat{N}_{\omega}$	16	13	14	14	**8**	**8**	**7**	**4**	**4**
	$\hat{N}_{\omega t}$	**15**	**11**	**11**	14	**8**	**8**	8	**4**	**4**
	$\hat{N}_{\omega T}$	17	15	14	**11**	**8**	**8**	**7**	**4**	**4**

Table 9.8: ω-Estimators: Percentage Root Mean Square Error

N	Estimator	Simulation Model 1 a	1 b	1 c	2 a	2 b	2 c	3 a	3 b	3 c
100	$\hat{N}_{\omega}$	772	702	1573	314	96	23	39	12	9
	$\hat{N}_{\omega t}$	915	799	833	383	438	19	34	**11**	**8**
	$\hat{N}_{\omega T}$	**49**	**222**	**425**	**31**	**22**	**18**	**18**	**11**	9
250	$\hat{N}_{\omega}$	61	65	33	34	15	12	12	6	6
	$\hat{N}_{\omega t}$	59	35	**22**	40	14	**10**	14	**6**	**6**
	$\hat{N}_{\omega T}$	**36**	**30**	24	**22**	**14**	11	**12**	7	6
500	$\hat{N}_{\omega}$	17	13	14	15	8	**8**	**4**	**4**	**4**
	$\hat{N}_{\omega t}$	**16**	**11**	**11**	**15**	**8**	**8**	8	**4**	**4**
	$\hat{N}_{\omega T}$	28	20	16	16	9	8	9	**4**	**4**

Table 9.9: The $\hat{N}_\mu$: Percentage Bias, CV and RMSE

Property	N	Simulation Model								
		1			**2**			**3**		
		a	b	c	a	b	c	a	b	c
% Bias	100	-18	-0.3	11	-7.7	4.0	6.3	-2.8	2.5	2.4
	250	-17	0.5	7.6	-6.1	4.1	6.5	-1.8	1.9	2.1
	500	-17	1.0	5.5	-5.7	4.4	5.0	-1.4	1.6	1.4
% CV	100	24	**33**	**40**	18	20	23	13	13	11
	250	14	20	24	11	14	16	9	8	7
	500	9	12	16	8	10	12	6	6	5
%RMSE	100	27	33	46	19	21	25	13	13	12
	250	20	20	27	12	15	18	9	9	8
	500	19	13	18	9	11	13	6	6	5

Table 9.10: $\hat{N}_{\omega T}$ and $\hat{N}_{\omega t}$ *vs* $\hat{N}_\mu$: Percentage Bias

N	Estimator	Simulation Model								
		1			**2**			**3**		
		a	b	c	a	b	c	a	b	c
100	$\hat{N}_{\omega T}$	-45	-25	**-6.0**	-27	-16	-10	-13	-6.1	-3.4
	$\hat{N}_{\omega t}$	109	87	60	46	28	**0.6**	8.0	**1.2**	**-0.3**
	$\hat{N}_\mu$	**-18**	**-0.3**	11	**-7.7**	**4.0**	6.3	**-2.8**	2.5	2.4
250	$\hat{N}_{\omega T}$	-36	-24	-16	-18	-7.9	-4.9	-8.4	-3.0	-1.2
	$\hat{N}_{\omega t}$	**6.8**	**-0.2**	**-3.4**	8.0	**1.4**	**-0.3**	3.5	**0.6**	**0.1**
	$\hat{N}_\mu$	-17	0.5	7.6	**-6.1**	4.1	6.5	**-1.8**	1.9	2.1
500	$\hat{N}_{\omega T}$	-25	-15	-10	-13	-4.7	-2.6	-5.6	-1.3	-0.5
	$\hat{N}_{\omega t}$	**2.9**	-1.4	**-2.4**	**0.8**	**0.7**	**0.1**	1.9	**0.6**	**0.4**
	$\hat{N}_\mu$	-17	**1.0**	5.5	-5.7	4.4	5.0	**-1.4**	1.6	1.4

Table 9.11: $\hat{N}_{\omega T}$ and $\hat{N}_{\omega t}$ *vs* $\hat{N}_{\mu}$: Percentage CV

		Simulation Model								
		1			**2**			**3**		
N	Estimator	a	b	c	a	b	c	a	b	c
100	$\hat{N}_{\omega T}$	36	296	452	21	**19**	**17**	14	10	9
	$\hat{N}_{\omega t}$	330	354	519	227	83	22	35	12	**9**
	$\hat{N}_{\mu}$	**24**	**33**	**40**	**18**	20	23	**13**	13	11
250	$\hat{N}_{\omega T}$	24	23	**22**	15	**12**	**10**	**9**	7	**6**
	$\hat{N}_{\omega t}$	55	35	**22**	36	14	16	13	**6**	**6**
	$\hat{N}_{\mu}$	**14**	**20**	24	**11**	14	16	**9**	8	7
500	$\hat{N}_{\omega T}$	17	15	14	11	**8**	**8**	7	**4**	**4**
	$\hat{N}_{\omega t}$	15	**11**	**11**	14	**8**	**8**	8	**4**	**4**
	$\hat{N}_{\mu}$	**9**	12	16	**8**	10	12	**6**	6	5

Table 9.12: $\hat{N}_{\omega T}$ and $\hat{N}_{\omega t}$ *vs* $\hat{N}_{\mu}$: Percentage Root Mean Square Error

		Simulation Model								
		1			**2**			**3**		
N	Estimator	a	b	c	a	b	c	a	b	c
100	$\hat{N}_{\omega T}$	49	222	425	31	22	**18**	18	**11**	9
	$\hat{N}_{\omega t}$	915	799	833	383	438	19	34	**11**	**8**
	$\hat{N}_{\mu}$	**27**	**33**	**46**	19	**21**	25	**13**	13	12
250	$\hat{N}_{\omega T}$	36	30	24	22	**14**	11	12	7	6
	$\hat{N}_{\omega t}$	59	35	**22**	40	14	**10**	14	**2**	**1**
	$\hat{N}_{\mu}$	**20**	**20**	27	**12**	15	18	**9**	9	8
500	$\hat{N}_{\omega T}$	28	20	16	16	9	8	9	**4**	4
	$\hat{N}_{\omega t}$	**16**	**11**	**11**	15	**8**	**8**	8	**4**	**4**
	$\hat{N}_{\mu}$	19	13	18	**9**	11	13	**6**	6	5

9.5.2 Figures Comparing the Estimator Properties

The Figures in this Appendix contain graphical comparisons between the estimators two at a time. Each comparison consists of three Figures, with the comparison for $N = 100$ at the top, for $N = 250$ in the middle and for $N = 500$ at the bottom. Models 1a, 1b and 1c are those with mean detection probability (the x-axis) around 0.4, Models 2a, 2b and 2c those with mean detection probability around 0.6 and Models 3a, 3b and 3c those with mean detection probability around 0.8.

$+$ is used to label heterogeneity level (a), i.e. $\sigma_z = 1$. (An "(a)" also appears adjacent to the symbol in the legend at the foot of the Figure.)

$\bigcirc$ is used to label heterogeneity level (b), i.e. $\sigma_z = 0.5$. (A "(b)" also appears adjacent to the symbol in the legend at the foot of the Figure.)

$\clubsuit$ is used to label heterogeneity level (a), i.e. $\sigma_z = 0.25$. (A "(c)" also appears adjacent to the symbol in the legend at the foot of the Figure.)

w is used to label the estimator $\hat{N}_\omega$ in the legend at the foot of the Figure.

wt is used to label the estimator $\hat{N}_{\omega t}$ in the legend at the foot of the Figure.

wT is used to label the estimator $\hat{N}_{\omega T}$ in the legend at the foot of the Figure.

H is used to label the estimator $\hat{N}_H$ in the legend at the foot of the Figure.

Ht is used to label the estimator $\hat{N}_{Ht}$ in the legend at the foot of the Figure.

HT is used to label the estimator $\hat{N}_{HT}$ in the legend at the foot of the Figure.

mu is used to label the estimator $\hat{N}_\mu$ in the legend at the foot of the Figure.

H(no p) is used in the legend at the foot of the Figure to label the estimated coverage probability calculated for the estimator $\hat{N}_H$ when neglecting variance due to estimation of the detection probabilities.

w(no p) is used in the legend at the foot of the Figure to label the estimated coverage probability calculated for the estimator $\hat{N}_\omega$ when neglecting variance due to estimation of the detection probabilities.

Figure 9.5: Percentage Bias of $\hat{N}_\omega$ vs $\hat{N}_H$.

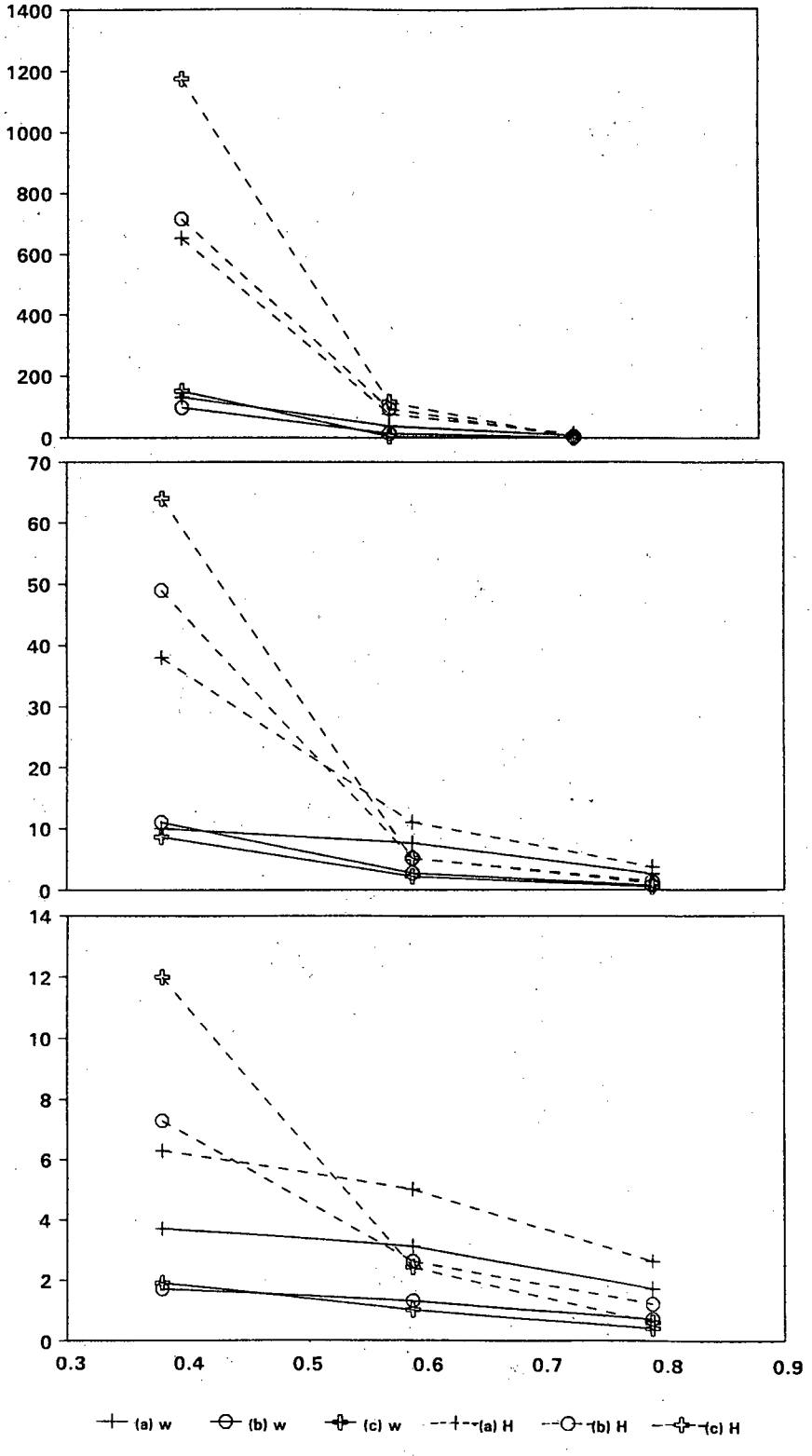


Figure 9.6: Percentage Bias of $\hat{N}_{\omega t}$ vs $\hat{N}_{Ht}$.

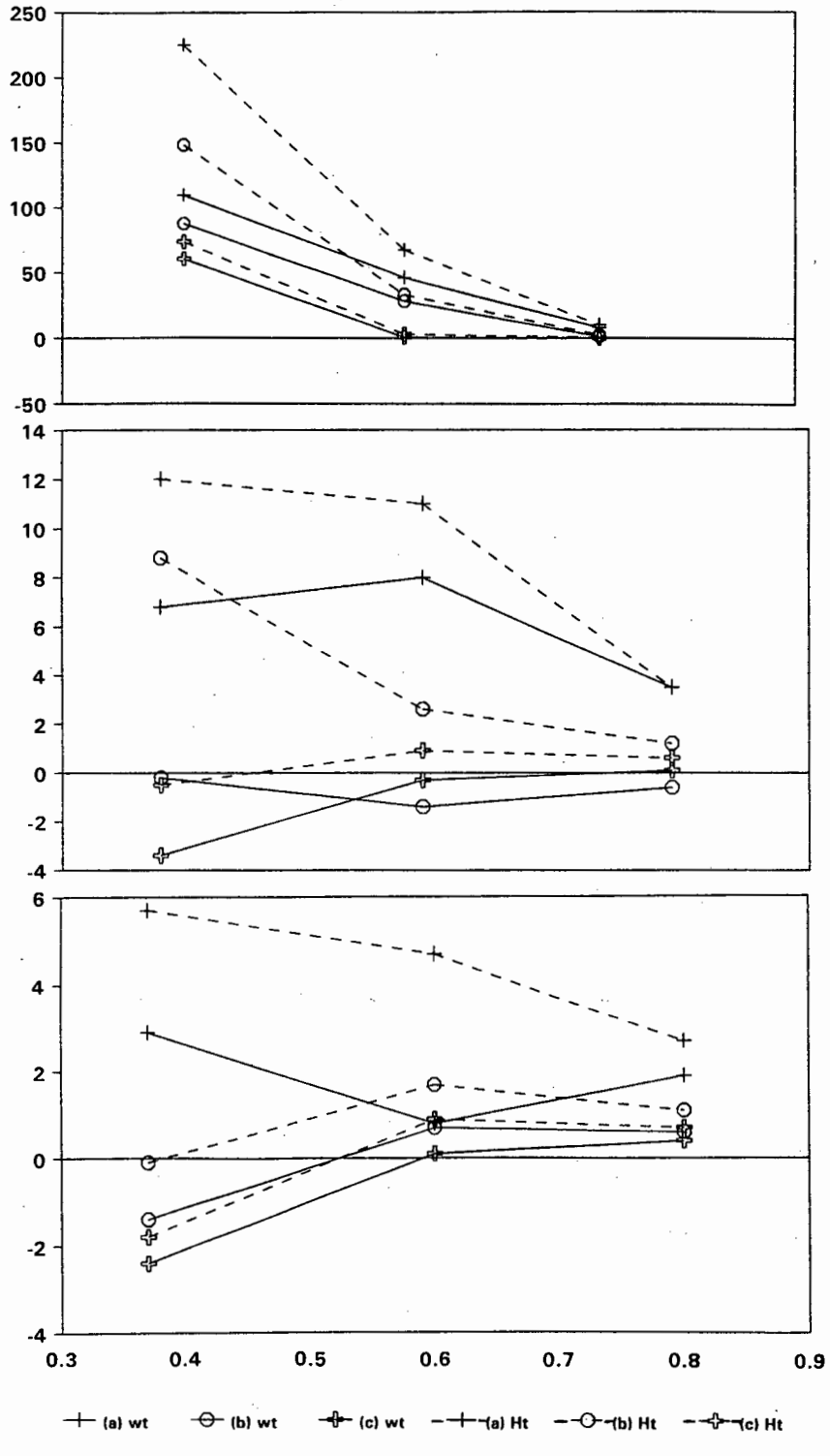


Figure 9.7: Percentage Bias of $\hat{N}_{\omega T}$ vs $\hat{N}_{HT}$.

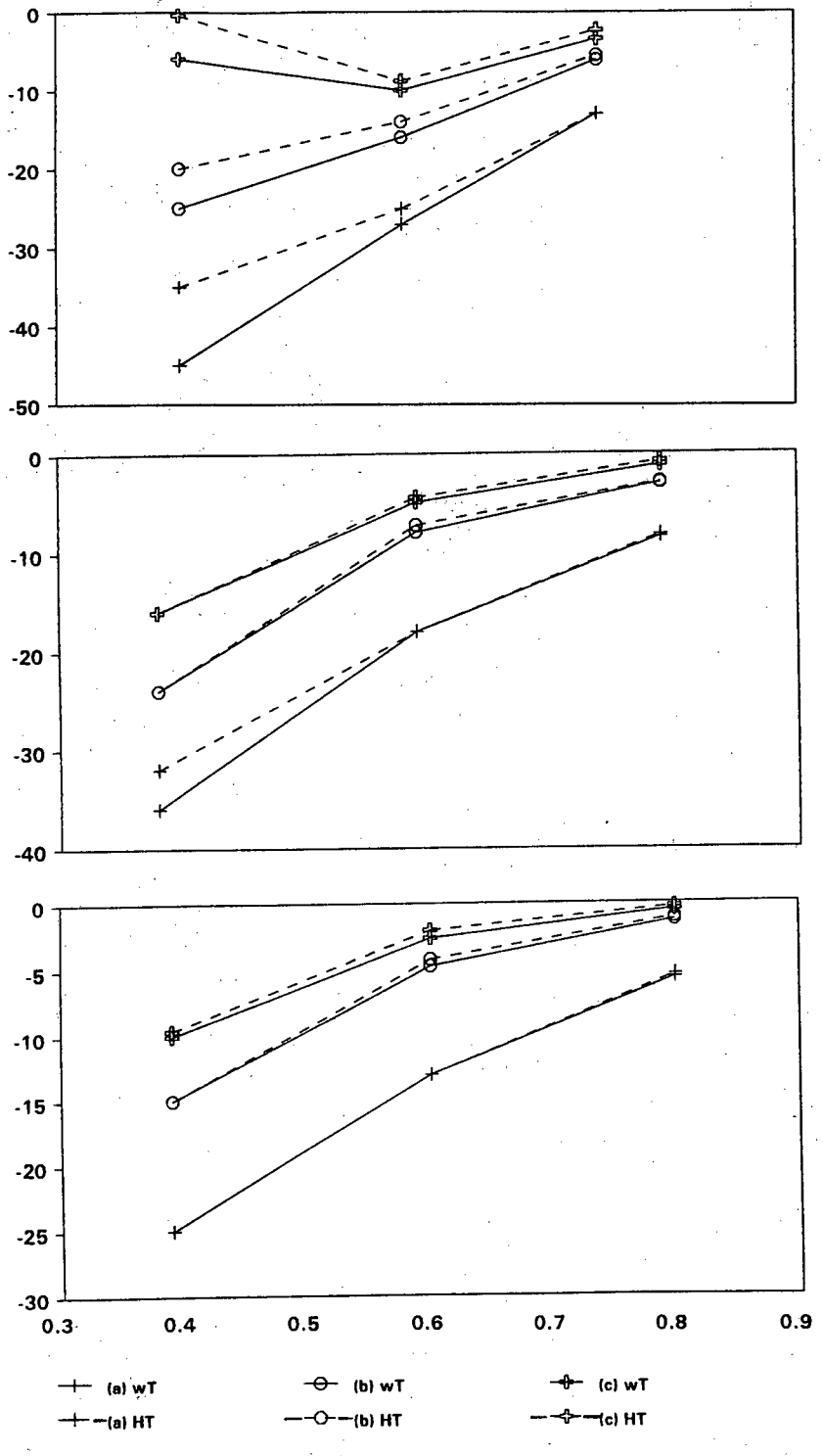


Figure 9.8: Percentage Root Mean Square Error of $\hat{N}_\omega$ vs $\hat{N}_H$.

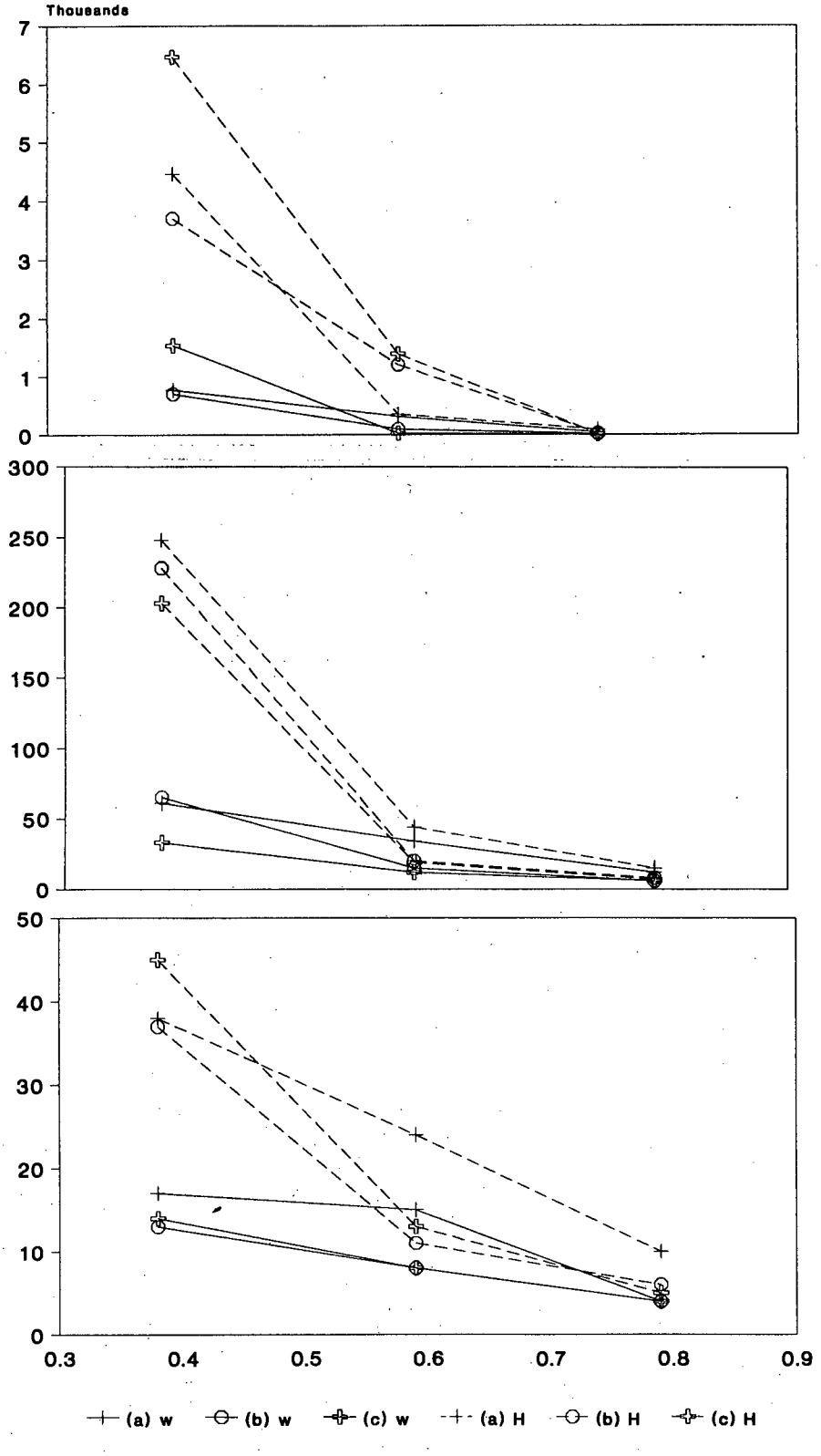


Figure 9.9: Percentage Root Mean Square Error of $\hat{N}_{\omega t}$ vs $\hat{N}_{Ht}$.

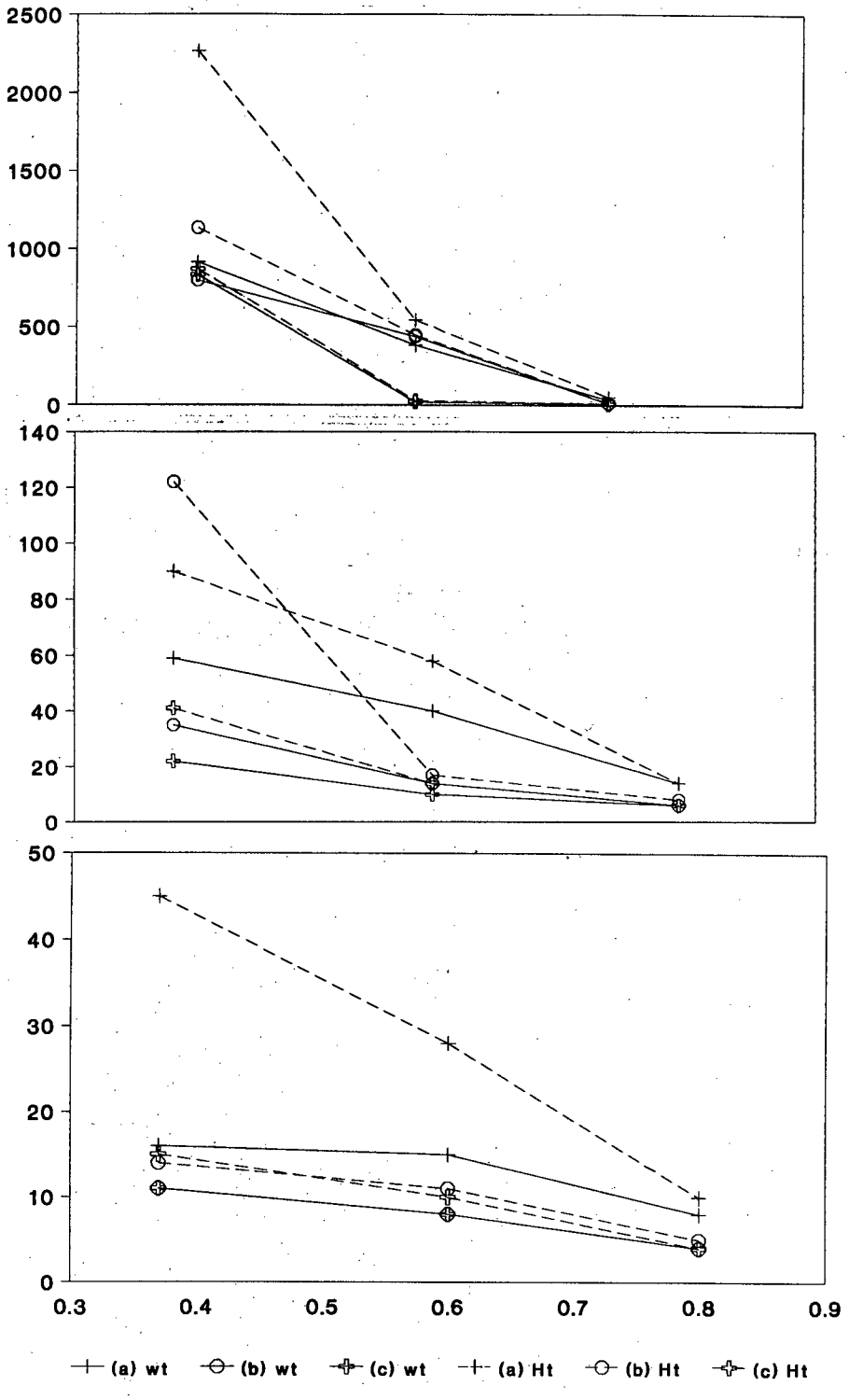


Figure 9.10: Percentage Root Mean Square Error of $\hat{N}_{\omega T}$ vs $\hat{N}_{HT}$.

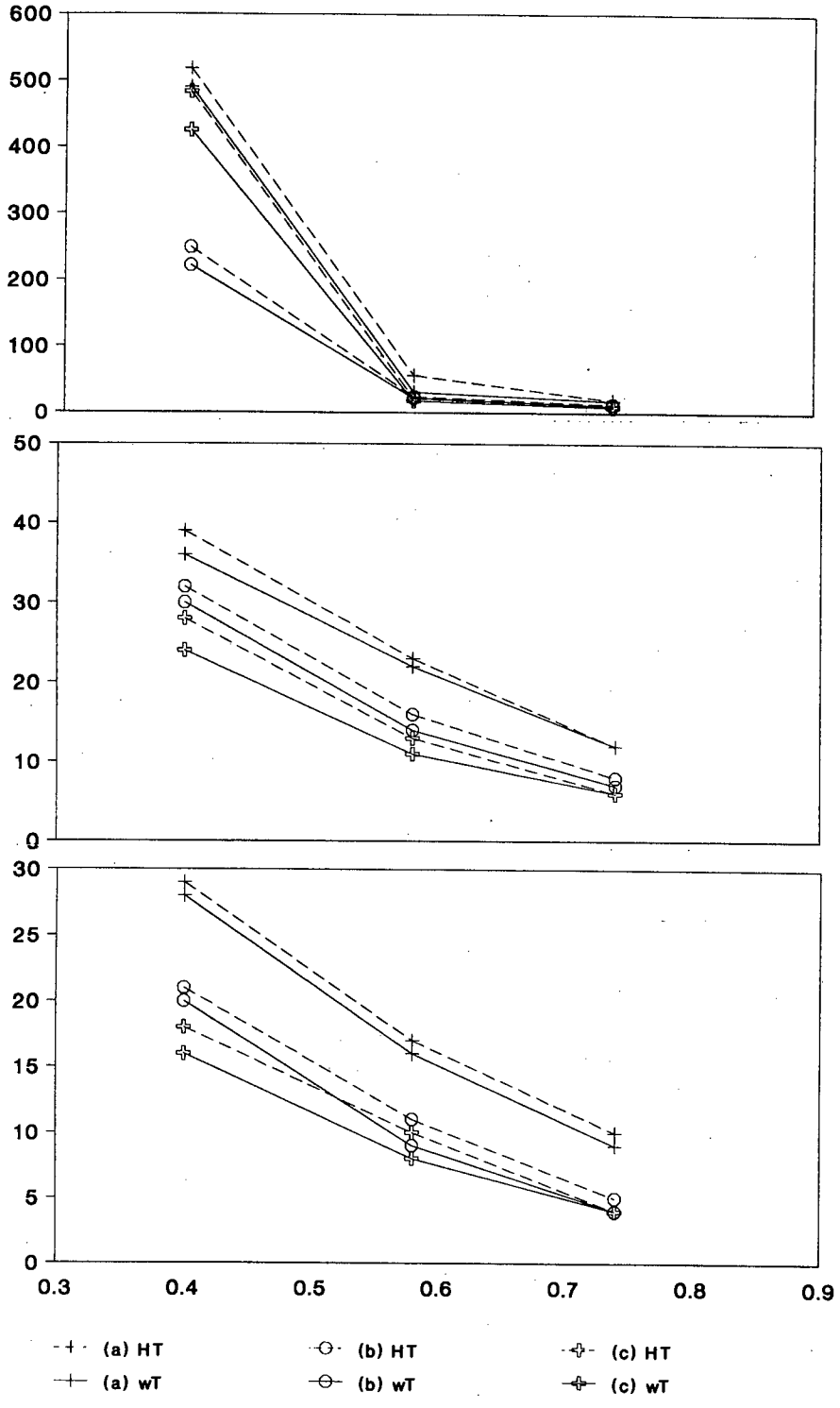


Figure 9.11: Percentage Bias of $\hat{N}_{\omega t}$ vs $\hat{N}_{\omega T}$.

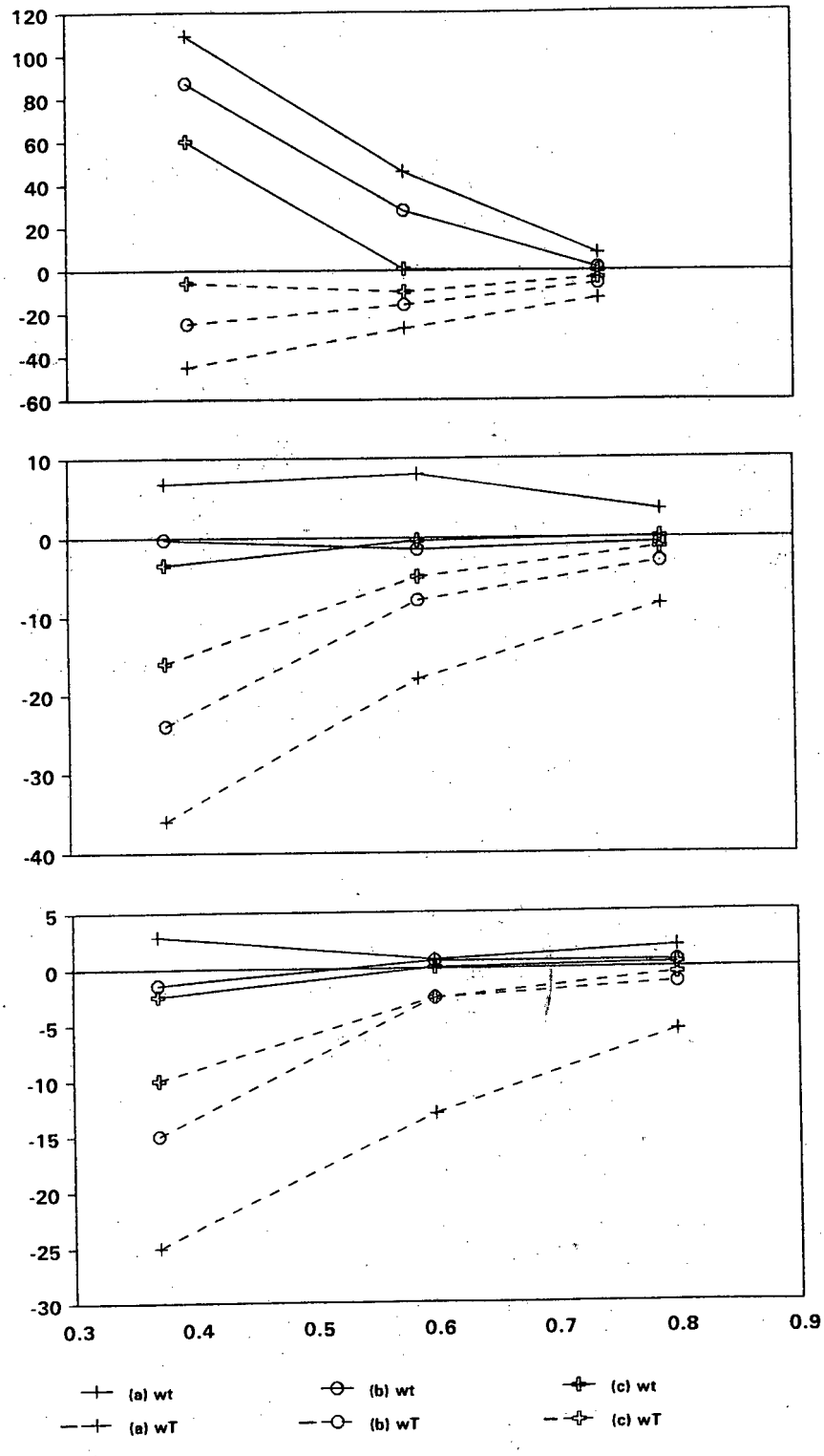


Figure 9.12: Percentage Root Mean Square Error of $\hat{N}_{\omega t}$ vs $\hat{N}_{\omega T}$.

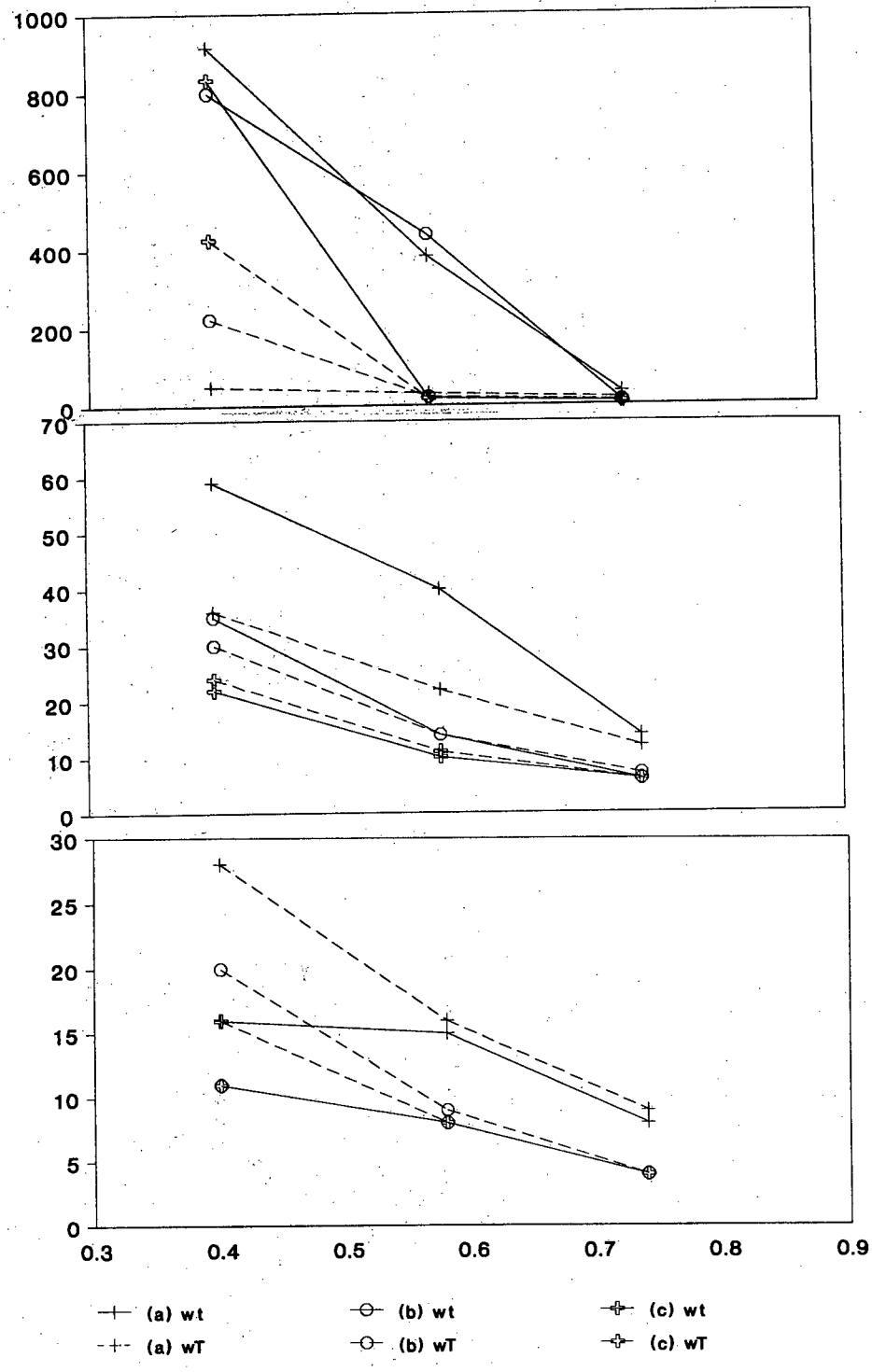


Figure 9.13: Percentage Bias of $\hat{N}_{\omega t}$ vs $\hat{N}_{\mu}$.

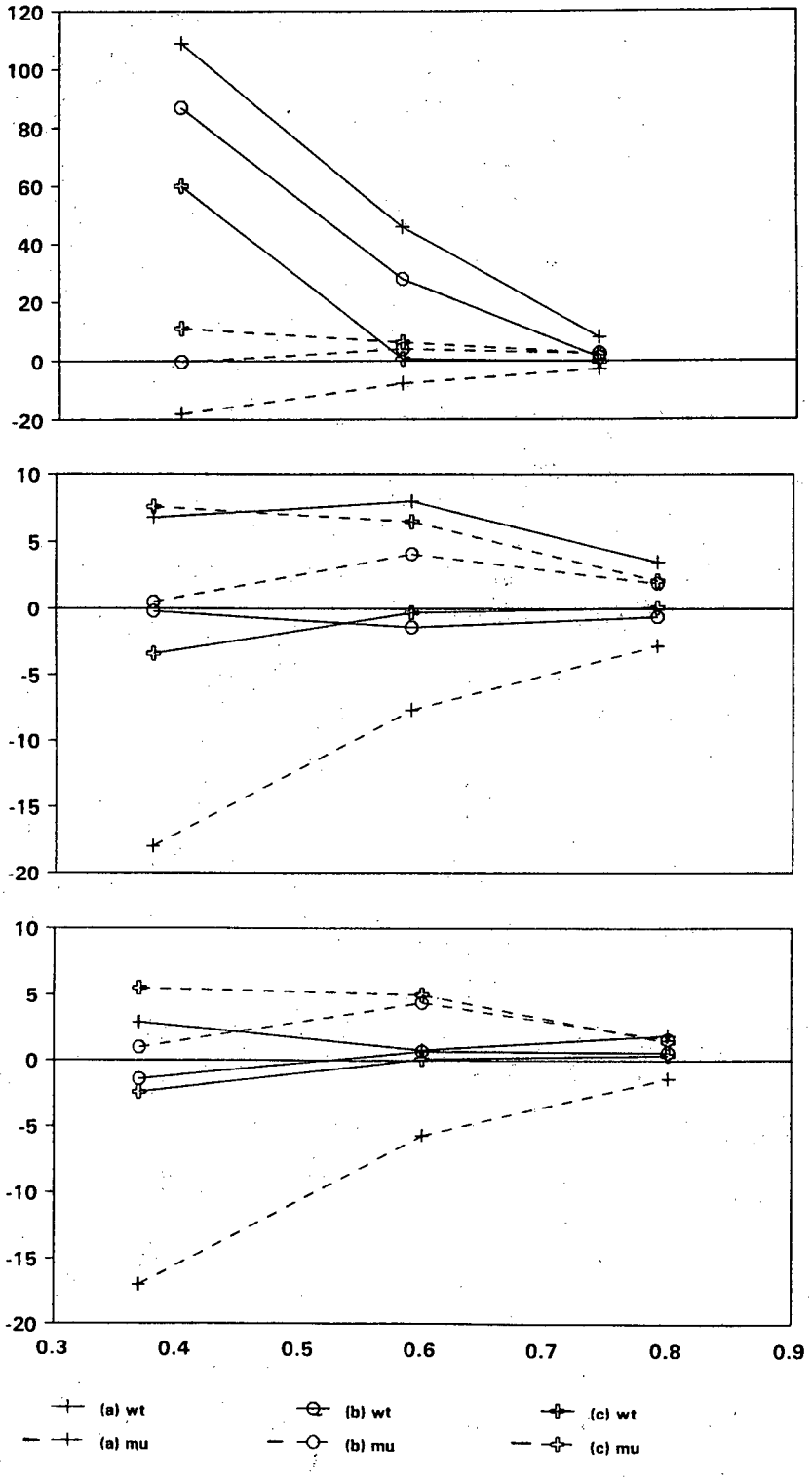


Figure 9.14: Percentage Root Mean Square Error of $\hat{N}_{\omega t}$ vs $\hat{N}_{\mu}$.

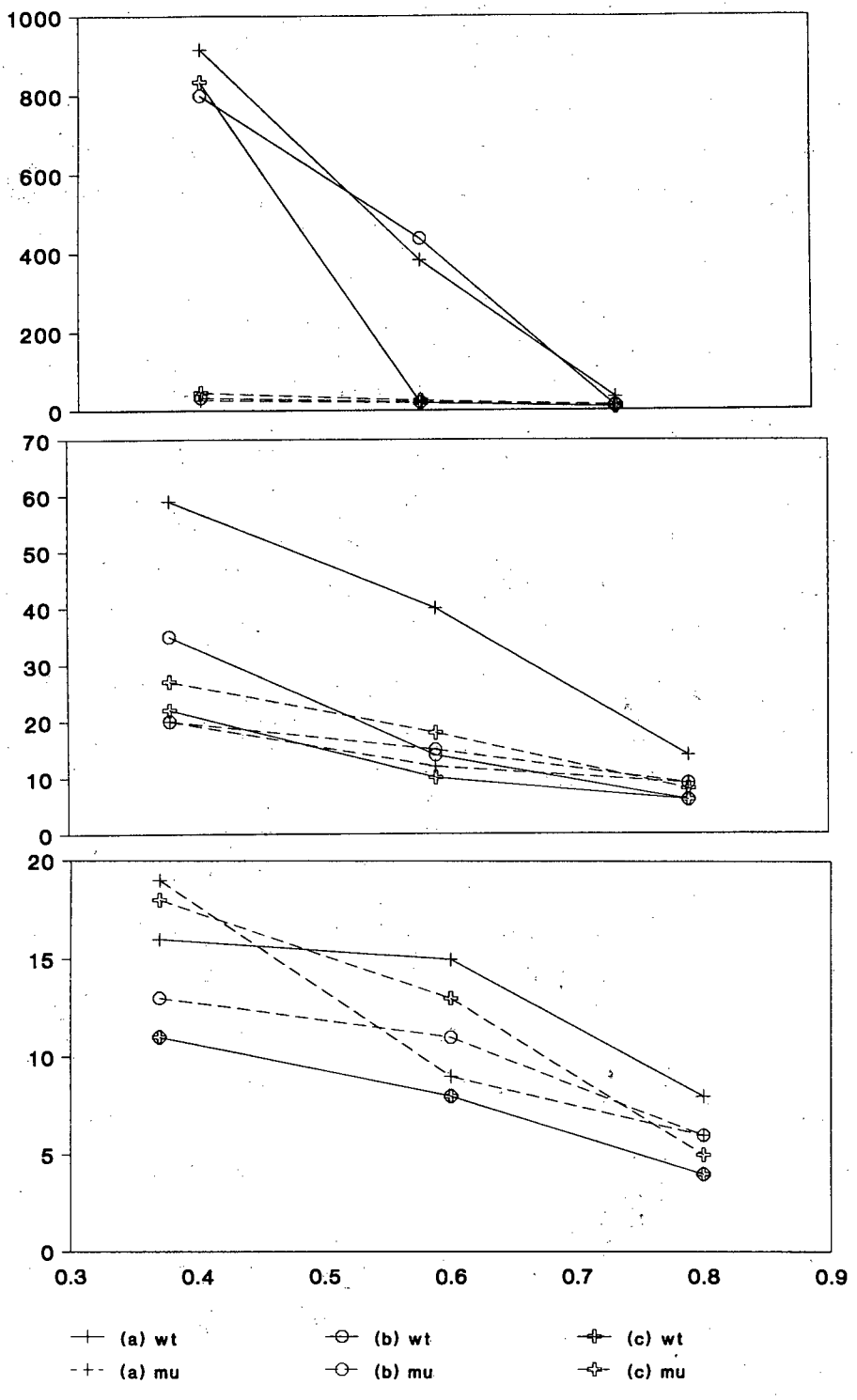


Figure 9.15: Percentage Bias of $\hat{N}_{\omega T}$ vs $\hat{N}_{\mu}$.

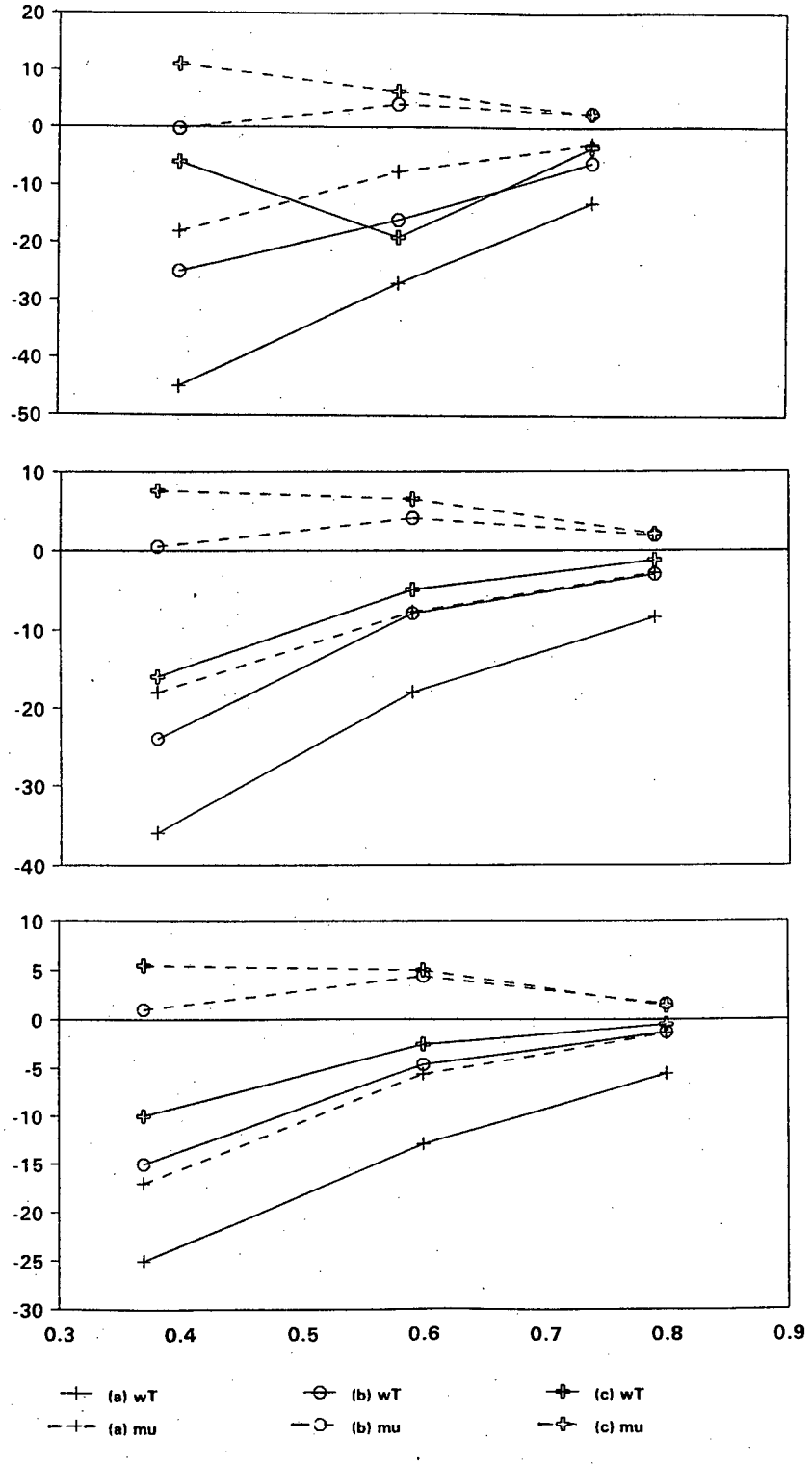


Figure 9.16: Percentage Root Mean Square Error of $\hat{N}_{\omega T}$ vs $\hat{N}_{\mu}$.

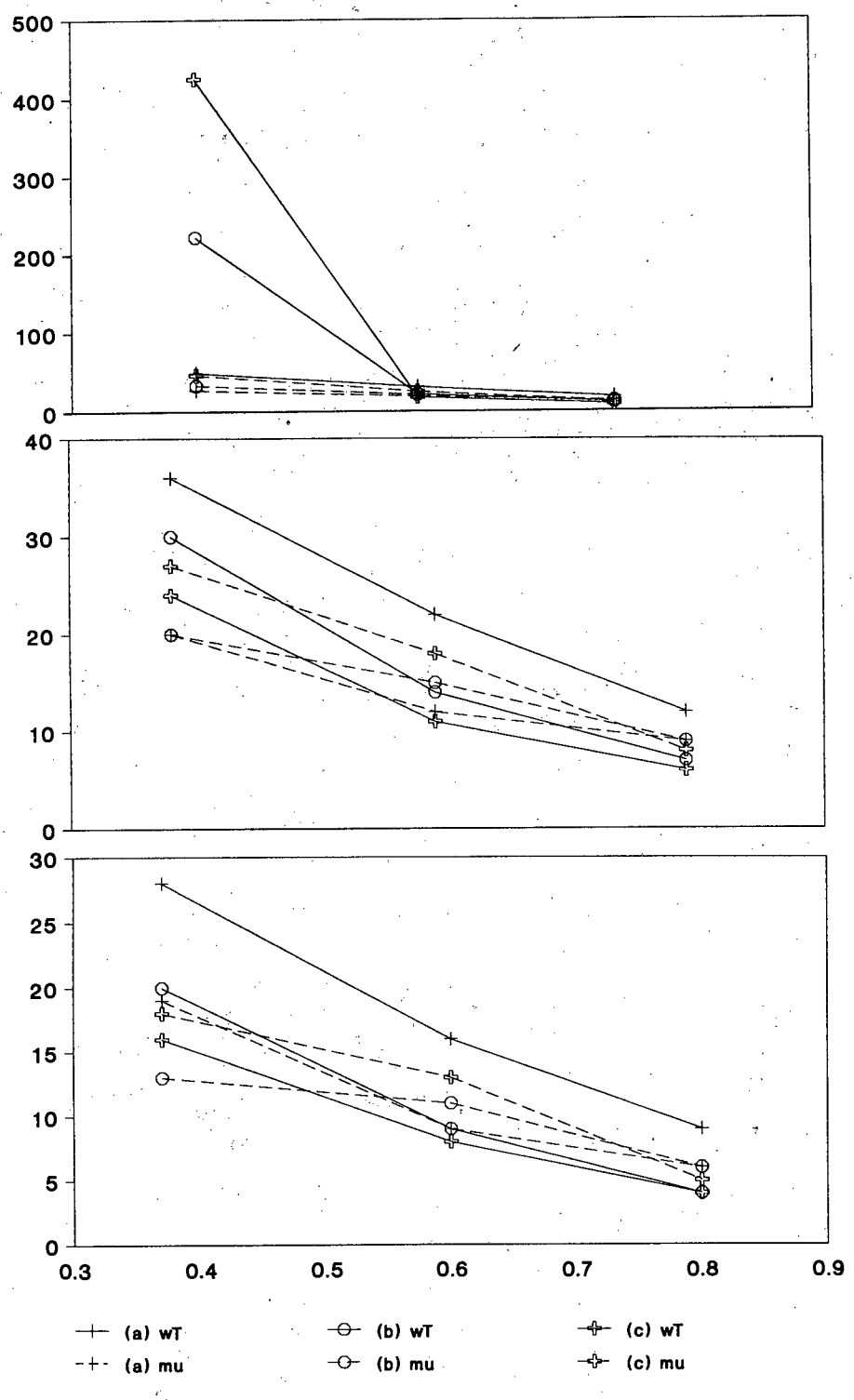


Figure 9.17: Percentage Coverage of N using the Analytic Estimator of Variance and assuming Normality for the Estimator $\hat{N}_H$, with (dotted lines) and without (solid lines) estimated variance due to estimation of detection probabilities

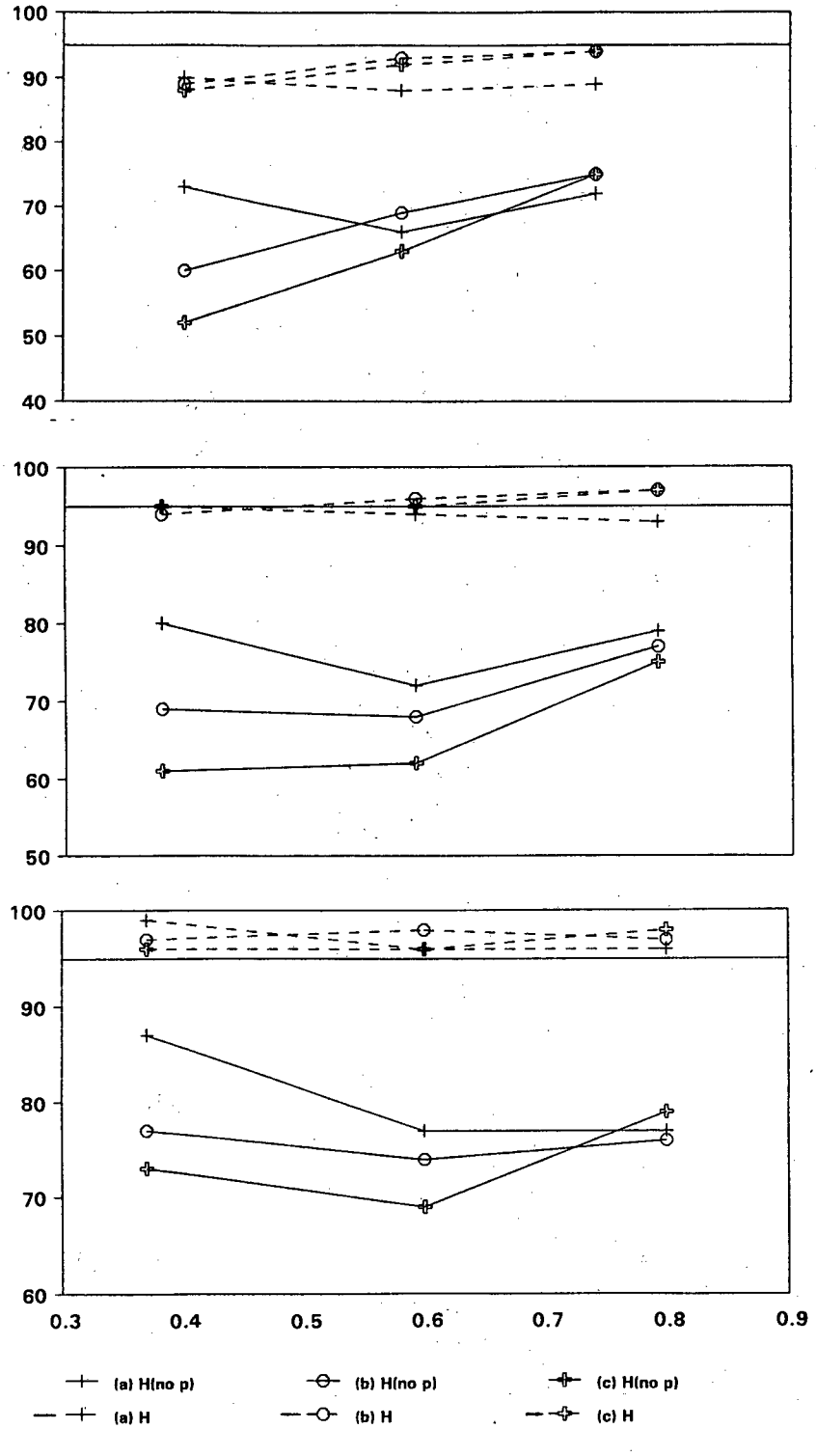


Figure 9.18: Percentage Coverage of N using the Analytic Estimator of Variance and assuming Normality for the Estimator $\hat{N}_\omega$, with (dotted lines) and without (solid lines) estimated variance due to estimation of detection probabilities

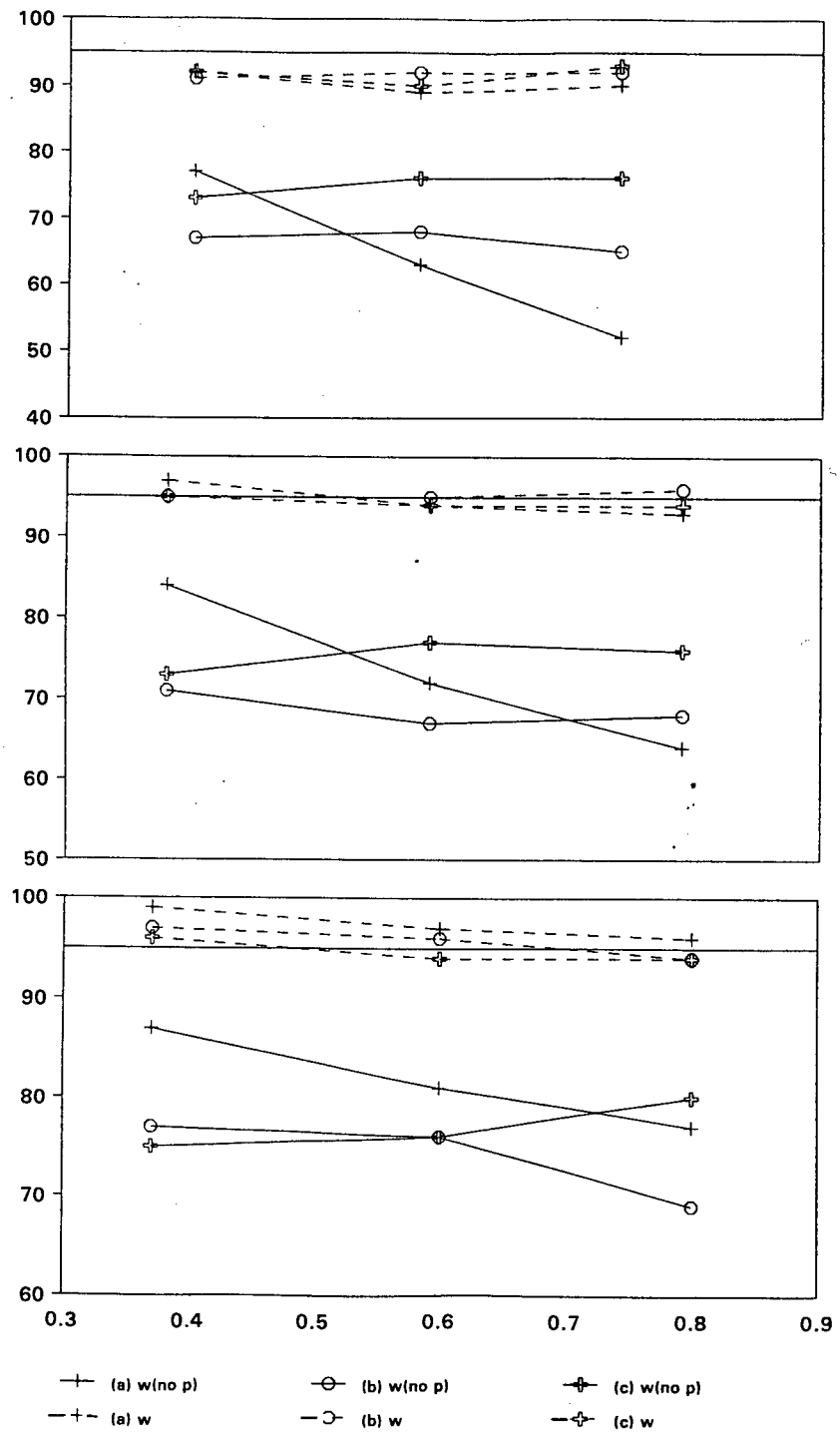


Figure 9.20: Percentage Coverage of N using the Analytic and Bootstrap Estimators of Variance: $\hat{N}_{\omega t}$.

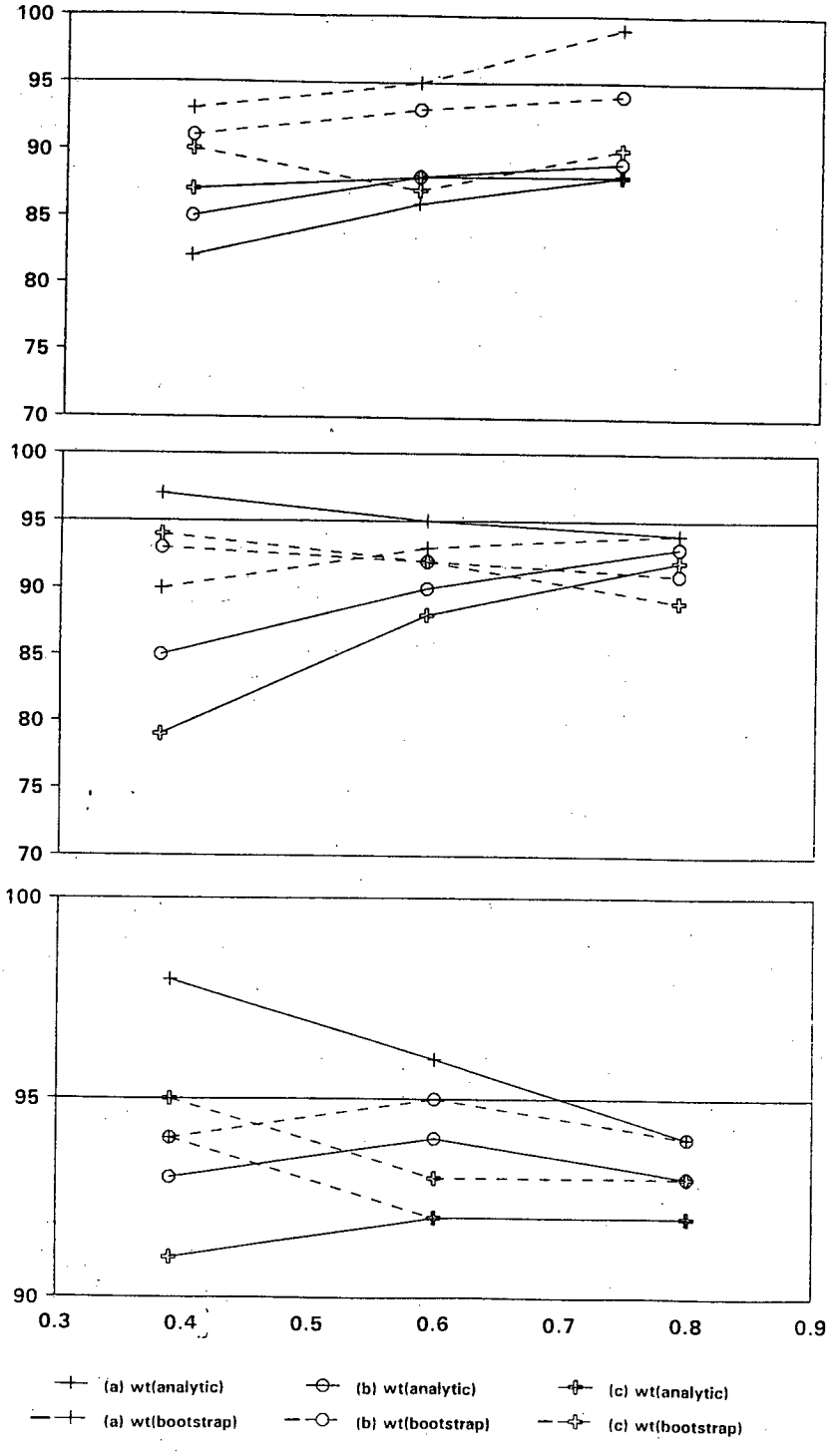
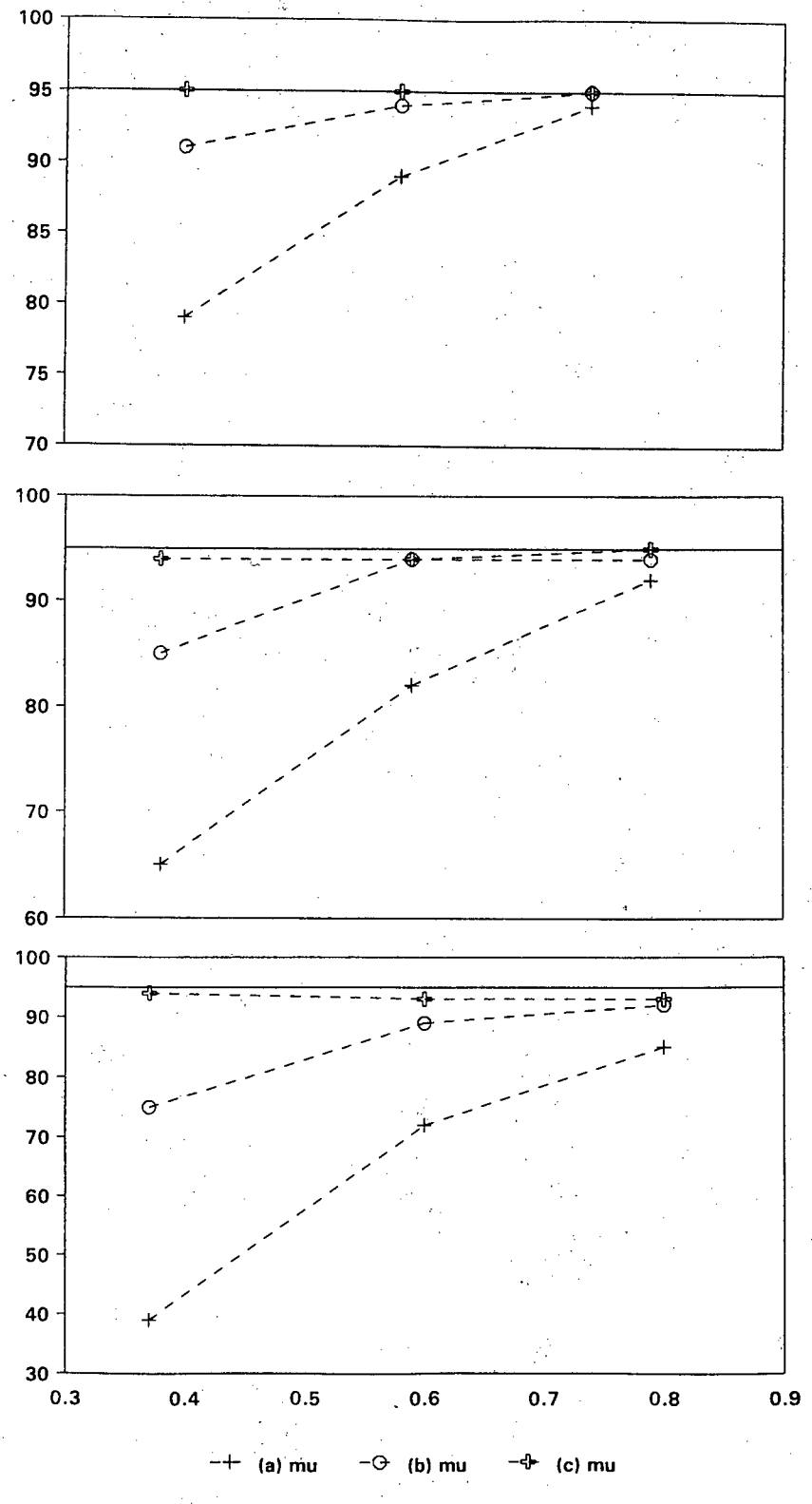


Figure 9.21: Percentage Coverage of N using the Bootstrap Estimator: $\hat{N}_{\mu}$.



Chapter 10

Conclusions

10.1 Overview

The success of estimation methods for conventional LT surveys (with certain detection on the trackline and detection functions which are modelled as functions of x alone) was built on the following.

(1) The development of a sound **statistical framework** for the estimation problem (specifically the derivation of likelihoods for grouped and ungrouped data).

(2) The development of **"model-robust" forms for the detection function** (i.e. forms for detection functions which yield approximately unbiased estimates of average detection probability for a wide range of true detection functions). Because in reality detection functions do depend on variables other than x, unbiased estimation relies on the theorem of Patil *et al.* (1993). This effectively states that in a conventional LT situation nothing is lost by ignoring variables other than x which affect detection probability, provided the form of the pooled detection function (pooled over all variables other than x) is modelled adequately. Model robust estimators allow for such adequate modelling of the pooled detection function; the theorem ensures unbiased inference given a detection function which is adequately modelled.

(3) The development of a **unifying model and model selection methods for detection functions** (i.e. the developments of Buckland, 1992a and 1992b). Together with (1) and (2) above, this provides a powerful tool for modelling detection functions and estimating abundance.

With a few exceptions (notably the work of Schweder on discrete LT models and of Hiby and Lovell in an aerial LT context), there has been little in the way of comparable developments for models in which detection on the trackline is not certain and detection depends on variables other than x. No general likelihood framework comparable to that for the conventional LT case has been available; multivariate detection function models have not been investigated in any depth (and when detection on the trackline is not certain, pooling over the additional variables is not an option because it results in biased estimation in general); there is no general framework for model selection.

The developments contained in this thesis go some way toward establishing statistical foundations for MRLT models comparable to those of conventional LT models ((1) to (3) above). A general statistical framework for MRLT models is developed. A new multivariate LT detection function model is presented. Combined with the the general model of (3) above, this has the potential to yield a general form of detection function (see below) in which detection probability can depend on both x and other variables, and in which detection on the trackline can be less than unity. This, together with the general MRLT likelihoods, in principle provides the means for general statistical model selection and estimation procedure comparable to those available for the conventional LT case.

In the following sections I attempt to draw together the main developments and central themes of this thesis, as well as to identify areas which would benefit from further research, and to suggest directions such research might take.

10.2 The Main Developments and Results of the Thesis

10.2.1 The Development of a General Theory for MRLT Models

Leaving aside the question of suitable forms for multivariate LT detection functions for the moment, there is a critical difference between MR contexts and the MRLT context. This is that in the latter, the distribution of one of the explanatory variables (namely x) in the population (detected and undetected) can reasonably be treated as known for the purposes of point estimation. Existing MR likelihoods which incorporate models for catchability/detectability as functions of explanatory variables do not incorporate the densities of explanatory variables, nor do the corresponding estimators use any assumptions about the distributions of explanatory variables. The density of x in the population (both detected and undetected), however, is central to conventional LT detection likelihoods and estimators.

A central part of this thesis is the development of likelihood functions for the MRLT context which incorporate the densities of the explanatory variables. Likelihoods are developed for both the case in which data are ungrouped, and the case in which they are grouped. The likelihoods are extensions of MR likelihoods to include the densities of explanatory variables. The MR likelihood of Huggins (1989) and Alho (1990) is one component of the MRLT likelihood for ungrouped data. The corresponding component of the MRLT likelihood for grouped data provides the equivalent of Huggins' (1989) and Alho's (1990) likelihood for grouped data. The MRLT likelihoods are also extensions of LT likelihoods to include a MR component and multivariate detection functions which have intercepts less than unity. Conventional LT likelihoods are readily obtained as special cases of these MRLT likelihoods (in both the grouped and the ungrouped data cases).

10.2.2 The Development of MRLT Detection Functions

A critical difference between MRLT models and conventional LT models is that the theorem of Patil *et al.* (1993) does not apply with MRLT models (when detection on the trackline is less than certain); pooling

data over all variables other than x in a MRLT context results in biased inference (Appendix 6.10.1). It is not in general possible to have "pooling–robust" estimators in a MRLT context. As a result, modelling multivariate detection probability functions is necessarily central to MRLT models. Little research has been carried out modelling such detection functions in the continuous availability case. All that has been done was carried out either in a conventional LT model context (in which detection on the trackline is certain) with only one variable other than x (usually group/animal size), or in a straight MR context (in which the considerable body of research on suitable forms for LT detection functions in the x-dimension is ignored).

In the special case in which detection on the trackline is constant (unaffected by any variables which affect probability of detection off the trackline), the conventional multivariate detection functions are easily adapted to the MRLT context by multiplying them by a "$g(0)$" parameter. (The parameter is the probability of detection on the trackline, called G in this thesis). In this case, pooling–robust estimation is possible so that modelling the dependence of the detection functions on variables other than x is not so critical. (Pooling–robust estimation can be achieved in the "constant-G" case when the pdf of observed x's of duplicate detections is modelled without assuming that the detection function for duplicates is the product of the detection functions of the two observers - see Appendix 6.10.1.) However, in most contexts it would not be plausible to assume that the probability of detection of animals on the trackline is the same for all animals and under all conditions. Other than Borchers *et al.* (1995), there are no multivariate LT detection function models in the literature which allow detection probability on the trackline to be less than unity and also allow it to depend on explanatory variables.

The logistic model has been used in straight MR contexts and in analysing observations from shore–based surveys. It accommodates both uncertain detection on the trackline and dependence of the detection function on explanatory variables both on and off the trackline. However, it was found to be inadequate in a MRLT context because its implicit shape constraints can prevent it modelling observed x distribution data adequately. A successful adaptation of this model was developed by transforming x using the hazard rate form developed for conventional LT surveys.

10.2.3 The Development and Testing of Estimators

The general MRLT likelihoods allow abundance in the covered area, N, to be estimated by simultaneously maximizing the appropriate likelihood with respect to the parameters of the detection function ($\boldsymbol{\beta}$, say), the parameters of the densities of the explanatory variables ($\boldsymbol{\phi}$, say) and N. A disadvantage of this approach is that it requires a functional form for the densities of the explanatory variables ($\pi(x, \boldsymbol{z})$) to be specified (as functions of the unknown parameters, $\boldsymbol{\phi}$). This is problematic for two reasons. Firstly it may be difficult *a priori* to specify suitable functional forms for $\pi(x, \boldsymbol{z})$. Secondly, if there are more than a few explanatory variables, the total number of parameters to be estimated is substantially increased by the addition of $\boldsymbol{\phi}$ to the other parameters which must necessarily be estimated, namely $\boldsymbol{\beta}$ and N.

It is possible to avoid estimation of $\boldsymbol{\phi}$ by adopting a conditional approach to estimation of N. In the

first step of such an approach, β is estimated from the MR component of the likelihood. This amounts to a form of binary regression. When the detection function is one of the standard link functions of Generalized Linear Models (GLM's), GLM software can be used directly to estimate β. (A logistic detection function is a case in point.) When the detection function is a non-standard link function (like the class of generalized MRLT detection functions outlined in Appendix 7.10.3) additional software will generally need to be developed to maximize the appropriate likelihood numerically with respect to β.

Conditional on the estimated β (call it $\hat{\beta}$), a Horvitz–Thompson estimator can be used to estimate N. The Horvitz–Thompson–like like estimator proposed by Huggins (1989) and Alho (1990) does not use the fact that in a LT context the density, $\pi(x)$, of x in the population (detected and undetected) can be treated as known for point estimation. However, their estimator can easily be adapted to make use of this information, and the resulting estimator is both less biased and more efficient than the Huggins/Alho estimator for finite samples in a LT context.

The simulation results for Horvitz–Thompson–like estimators strongly suggest that MRLT data should be truncated more severely than is generally recommended for conventional LT data. In particular, "t-truncated" or "T-truncated" versions of $\hat{N}_\omega$ performed better than the untruncated version in situations where the untruncated version provided substantially biased estimates of N.

The other form of conditional estimator investigated in the thesis is one which attempts to model the density of detection probabilites (by either or both observers) in the population by a Beta density function. Estimation of the parameters of this density and of N is performed conditional on $\hat{\beta}$. A form of the estimator which constrains the intercept of the Beta density at zero probability to be finite was found to be very successful in overcoming the very large bias of Horvitz–Thompson–like estimators when mean detection probability and sample size is small. However, the constraint on one of the estimated Beta density function parameters makes the estimator inconsistent for N in situations where the constraint is likely to be hit. One symptom of this is the fact that when heterogeneity is high and/or mean detection probability is low, the estimated 95% confidence intervals for N for this estimator have increasingly poor coverage probability as sample size increases.

Overall Horvitz–Thompson–like estimators which treat $\pi(x)$ as known are judged to perform best. (For brevity, I call this class of Horvitz–Thompson–like estimator "ω-estimators". The class includes $\hat{N}_\omega$, $\hat{N}_{\omega t}$ and $\hat{N}_{\omega T}$.) None of these estimators performs well when mean detection probability is low and sample size is small. They become extremely positively biased in this case (the ω-estimators less so than the various truncated versions of the Horvitz–Thompson–like estimator proposed by Huggins (1989) and Alho (1990)). $\hat{N}_\mu$ may be preferable in this case, despite the fact that it is not a consistent estimator of N (with small sample sizes consistency is not really relevant). Clustering of duplicates in the $p_{..}$-dimension is an indicator that the Horvitz–Thompson–like estimator may be severely biased.

In addition to performing well in terms of statistical criteria, ω- estimators have the following attractive properties.

- They provide a means for integrated estimation of abundance, as opposed to the separate estimation of $f(0)$, sighting rate and group size of conventional LT theory. Unlike the conventional approach, this integrated approach does not depend on assumptions of independence between what in the conventional LT estimation context are separate components of the estimator.
- They provide a statistically coherent means for estimating mean group size in the population in the presence of size-biased detections, even when detection on the trackline is not certain and probability of detection of animals on the trackline may itself depend on group size. (To date the only available methods for estimating mean school size in the presence of size-biased detection rely on the assumption that detection of animals on the trackline is either certain or independent of school size.)
- They provide a statistically coherent means for estimating the pdf in the population (detected and undetected) of any of the explanatory variables, the moments of these pdf's and the expectation of any function of the explanatory variables which may be of interest.
- They are readily combined with the survey method of Buckland and Turnock (1992) to provide a MRLT estimator which takes account of random or responsive animal movement.
- They provide a framework within which abundance estimation with variable coverage probability survey designs can readily be developed.
- Together with the general likelihoods developed in the thesis, they provide a unifying framework for estimation for all LT models (see below).

10.3 Future Developments

In this final section of the thesis, I briefly discuss some of the areas in which future research might usefully be directed. The topics fall into two classes. The first three are areas in which the developments and results of this thesis indicate fairly clearly the direction future research might usefully take. The last two are ones which have not been addressed in the thesis, but represent important unresolved issues for MRLT models.

10.3.1 Generalized MRLT Detection Functions

A logistic detection function was used in all the simulations of this thesis. This form for the detection function has been found to be inadequate for MRLT data in the single study in which it has been used (Borchers *et al*, 1995). Buckland (*pers. commn*) suggested a new form of MRLT detection function (Appendix 7.10.3). This was applied successfully to MRLT data in the analysis of Borchers *et al.* (1995). The method can be generalized. One such generalization would be to (a) use any one of the wide variety of conventional LT functional forms in place of the hazard rate form in transforming x, and (b) use other binary regression link functions in the GLM fitting step of the algorithm.

10.3.2 Unifying MRLT and Conventional LT Estimation Methods

In the ω-estimators ($\hat{N}_\omega$ or one of its truncated forms) and the corresponding estimators of abundance (denoted by $\hat{\aleph}_\omega$), one has a form of abundance estimator which enables a uniform approach to be taken to LT abundance estimation for all types of LT model (the types are illustrated in Figure 6.1). For example, the conventional LT estimator of abundance in the survey area (from Buckland *et al.*, 1993a; pp. 52-53) is $\hat{\aleph}_{conv} = An/\hat{P}_a$, where $\hat{P}_a$ is the estimated probability of detecting an animal within W of the trackline, unconditional on its distance from the trackline. This can be written in the notation of this thesis as follows.

$$\hat{\aleph}_{conv} = \frac{A}{a}\frac{n}{\hat{p}_{..}} = \frac{A}{a}\sum_{j=1}^{n}\frac{1}{\hat{p}_{..}} = \hat{\aleph}_\omega \tag{10.1}$$

Here $p_.(x,z) = p_.(x)$ (since detection probability is assumed to depend on x alone), $p_.(0) = 1$ (by conventional LT assumption), $\omega = \int p_.(x)\,dx$, $\pi(x) = W^{-1}$ and $p_{..} = \int p_.(x)\,\pi(x)\,dx$. $\hat{\aleph}_\omega$ is an ω-estimator because detection probabilities are averaged with respect to $\pi(x)$ in the denominator. This is something of a degenerate case since in the conventional model x is the only explanatory variable so that the probability associated with each of the detected animals in the estimator is identical for all animals.

Now consider a conventional LT estimator of individual abundance when animals occur in groups. If sampling is not size-biased, the jth group is of size s_j, and $\overline{s}$ is the mean observed group size, then a conventional LT estimator of abundance of individuals in this scenario is as follows.

$$\hat{\aleph}_{(groups)} \times \overline{s} = A\frac{n\overline{s}}{2L}\hat{f}(0) = \frac{A}{a}\sum_{j=1}^{n}\frac{s_j}{\hat{p}_{..}} = \hat{\aleph}_{\omega(indiv)} \tag{10.2}$$

Here $\hat{\aleph}_{(groups)}$ is the conventional estimator applied to groups. The conventional LT estimator ($\hat{N}_{groups} \times \overline{s}$) is equivalent to an ω-estimator for individuals, $\hat{\aleph}_{\omega(indiv)}$ (see section 7.8 for more on the estimation of abundance of individuals using $\hat{N}_\omega$). In the bivariate case, where detection probability depends on x and s, the ω-estimator for individual abundance is

$$\hat{\aleph}_\omega = \frac{A}{a}\sum_{j=1}^{n}\frac{s_j}{\hat{p}_.(s_j)} \tag{10.3}$$

where

$$\hat{p}_.(s_j) = \frac{1}{W}\int_0^W \hat{p}_.(x,s_j)\,dx \tag{10.4}$$

Together with a bootstrap variance and confidence interval estimation procedure, this form of estimator provides a general means of estimating individual abundance when sampling is size-baised. Unlike conventional LT estimation methods, this method does not assume mean group size to be independent of

group density. It is quite plausible for mean group size to be negatively correlated with group density. This could occur if, for example, individual density remained constant but animals ocurred in smaller groups in some regions, perhaps because of local environmental conditions. Confidence interval estimation under the conventional assumption of independence of group encounter rate and group size would be biased in this instance. A transect-based bootstrap estimation procedure implemented with the estimator of individual abundance, $\hat{\aleph}_\omega$ above, could incorporate the dependence, and if mean group size is negatively correlated with encounter rate, individual abundance could in fact be estimated with greater precision than group abundance.

What if sampling is size-biased? The parallels between LT estimators in the literature and $\hat{N}_\omega$ persist in this case and other conventional bivariate/multivariate situations. Drummer and McDonald's (1987) estimator and Quang's (1991) estimator, for example, correspond to a special case of $\hat{\aleph}_\omega$ for the conventional multivariate LT scenario (see section 4.7).

The bivariate and multivariate conventional detection function forms in the literature (namely those of Drummer and McDonald, 1987 and Quang, 1991) are "Constant-Shape" models, i.e. models in which variables other than x affect only the scale parameter of the detection function. The complimentary log-log form for the detection function (Chapter 4) has the potential to provide a general form for multivariate conventional detection functions which need not be "Constant-Shape" models. The complimentary log-log form could in principle be used with a wide variety of transformations of x and of the other explanatory variables to provide a wide range of multivariate detection functions. Conventional LT detection functions which are not "Constant-Shape" forms can in principle be obtained by including interaction terms in the "linear predictor" (Chapter 4). Together with $\hat{N}_\omega$, this could provide a general method for estimation in the conventional multivariate case.

The use of $\hat{N}_\omega$ (and its truncated forms) in a MRLT context has been illustrated at various points in this thesis. These ω-estimators have been found to perform well in a MRLT scenario; with its parallels in non- MRLT LT contexts, they provide the core of a unified estimation framework for all varieties of LT model.

10.3.3 Variable Coverage Probability Designs

With very few exceptions (Cooke, 1984; Hiby and Lovell, 1996) LT estimators of abundance are based on an assumption of equal coverage probabilities for all animals within predefined strata, if not all animals in the survey area. The assumption of equal coverage probability has been used throughout this thesis. The Horvitz–Thompson–like estimators covered in this thesis (including the ω-estimators) do, however, provide a means for estimating abundance when coverage probabilities vary within strata. (Both Cooke, 1984, and Hiby and Lovell, 1996, use Horvitz–Thompson estimators to estimate abundance in an unequal coverage probability scenario.)

In the equal coverage probability scenario, the coverage probability for the jth animal is $C_j = a/A$ (recall

that $a = 2LW$). In general this coverage probability will be different for each animal and will depend only on the position of the jth animal and the survey design. The estimator $\hat{N}_\omega$ is readily applied using the C_j's to give an estimator of animal abundance in the survey area, $\aleph$.

$$\hat{\aleph}_\omega = \sum_{j=1}^{n_{..}} \frac{W}{C_j \hat{\omega}_{.}(\boldsymbol{z}_j)} \tag{10.5}$$

With equal coverage probabilities, $C_j = a/A$ is factored out of the sum to give the following more familiar form.

$$\hat{\aleph}_\omega = \frac{A}{a} \sum_{j=1}^{n_{..}} \frac{W}{\hat{\omega}_{.}(\boldsymbol{z}_j)} \tag{10.6}$$

A substantial amount of research has been done in an aerial survey context on obtaining analytic estimators of variance of this form of estimator. With a few exceptions, this work has assumed the detection probability to be constant with respect to x (so that perpendicular distance detection functions are not used). In cases where detection probability is modelled as a function of some other explanatory variables $\boldsymbol{z}$, this is usually achieved by stratifying with respect to the explanatory variables rather than modelling the form of the dependence of the detection function on these variables. Two exceptions are the work of Hiby and Lovell (1996) and Evans *et al.* (1994).

The model of Hiby and Lovell (1996) is effectively a sophisticated MRLT model which also includes a likelihood component relating to duplicate identification. The model of Evans *et al.* (1994) is not framed in terms of detection functions at all, but in terms of a log-linear model for observed cell frequencies (each cell corresponds a particular {capture history; explanatory variable value} pair) with constraints on probabilites between cells. Were detection functions to be modelled in this framework, it would be as particular forms of constraints on the probabilities between cells. While this is not an obvious form in which to parameterize LT models, this is not to say it is not a useful one (although the MRLT formulation is more natural). Research on the use of Evans *et al.* (1994) estimation scheme in a LT context has yet to be pursued.

With the inclusion of coverage probabilites in LT estimators and the development of the models of Evans *et al.* (1994) and Wong (1996), theory for aerial surveys and non-aerial LT surveys (for want of a better term) is clearly converging. A closer comparison of the aerial estimation research and the MRLT theory and estimators developed here seems to be worth pursuing. The aerial work is framed in terms different from terrestrial and marine MRLT theory but, in principle the two problems are the same so that aerial survey and MRLT theory are special cases of a more general underlying theory. Aerial research seems to have concentrated on coverage probabilities to a greater extent than terrestrial and marine LT work, while the latter has concentrated on developing estimators of mean detection probabiltity but neglected the coverage probability component. Thompson and Seber (1994) present an estimation framework which

goes some way to unifying the two estimation approaches, as well as extending it to adaptive sampling schemes. (However, they do not address the estimation of detection probabilities and do not develop or use likelihood functions in estimation.)

10.3.4 Duplicate Identification

The MRLT theory and methods developed in this thesis assume that duplicates/"recaptures" can be identified without error. In practice this is not the case, and the uncertainty in duplicate identification can contribute substantially to the variance and bias of MRLT estimators.

The innovative work by Hiby and Lovell (1996) represents a substantial advance in this area and constitutes the only successful integrated statistical treatment of this problem in the LT literature. Their method uses independent information on errors in estimating animal positions and on the speeds of animal movement to incorporate uncertainty due to duplicate identification. I describe it in rough outline here. (See Hiby and Lovell (1996) for more details.) Under the assumption that animals are distributed independently in the plane, and given N, then the likelihood of any pair of sightings being duplicates (given their times and positions of detection) can be evaluated in terms of the stochastic models for estimated positions and animal movement. The likelihood of obtaining the observed positions at the observed times of detections made by a single observer team can be evaluated similarly. This then allows calculation of the likelihood of observing any particular set of duplicate and non-duplicate sightings, given N and the parameters of the detection functions. The likelihood of there being $n_{3.}$ duplicates is then the sum of such likelihoods over all such sets in which there are $n_{3.}$ duplicates. It is this likelihood which Hiby and Lovell (1996) maximize with respect to N, the parameters of the detection functions and $n_{3.}$.

There is an incompatability between their approach and the approaches to abundance estimation developed in this thesis. The regression methods for estimating detection probabilities which are used in this thesis require individual duplicate pairs to be identified. The method of Hiby and Lovell (1996) simultaneously estimates abundance and the detection function parameters without identifying which particular pairs of detections were duplicates (although in calculating the likelihood that they use for estimation, they do calculate the likelihood of individual sets of duplicate and non-duplicate observations).

10.3.5 Discrete Animal Availability

The models and estimators covered in this thesis do not explicitly accommodate discrete animal availability, except to the extent that this can be summarized in a single statistic. It is not clear to what extent this may bias the resulting estimators. Clearly this will depend on the particular survey methods employed and the type of heterogeneity in animal availability. The degree to which neglect of discrete availability may be a problem will vary from application to application, and the effects of such neglect will therefore need to be investigated on a case-specific basis. If estimation is based on the MRLT models and the estimators developed in this thesis, then it seems prudent to investigate the effects of likely forms

of discrete availability for the particular applications in which the estimators are applied. This could be done by simulation, using simulated data which reflect the true situation as far as possible.

Chapter 11

References

Alpizar–Jara,R. 1994. A combination line transect and capture–recapture sampling model for multiple observers in aerial surveys. *MSc. Dissertation.* North Carolina State University. 96pp.

Akaike,H. 1973. Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory*, 2nd edn (eds B.N. Petran and F. Csaaki), Akadeemiai Kiadi, Budapest, Hungary: 267–281.

Alho,J.M. 1990. Logistic regression in capture–recapture models. *Biometrics* **46**: 623–635.

Aldredge,J.R. and Gates,C.E. 1985. Line transect estimators for left–truncated distributions. *Biometrics* **41**: 273–280.

Anderson,D.R., Burnham,K.P. and Crain,B.R. 1980. Some comments on Anderson and Pospahala's correction of bias in line transect sampling. *Biometrical Journal* **22**: 513–524.

Anderson,D.R. and Pospahala,R.S. 1970. Correction of bias in belt transects of immotile objects. *Journal of Wildlife Management* **34**: 141–146.

Barlow,J. 1988. Harbour porpoise, *phocoena phocoena*, abundance estimation for California, Oregon and Washington: I. Ship surveys. *Fishery Bulletin* **86**: 417–432.

Barlow,J. 1993 (unpublished). The abundance of cetaceans in California waters estimated from shipboard surveys in summer/fall 1991. *NMFS Administrative report LJ–93–09.* 39pp.

Barlow,J., Oliver,C.W., Jackson,T.D. and Taylor,B.L. 1988. Harbour porpoise, *phocoena phocoena*, abundance estimation for California, Oregon and Washington: II. Aerial surveys. *Fishery Bulletin* **86**: 433–444.

Borchers,D.L. 1989 (unpublished). Estimates of $g(0)$ for minke schools from IO, IOT and CIO survey modes on the 1987/88 IWC/IDCR Antarctic sightings cruise. *IWC Document SC/41/SHMi5.* 19pp.

Borchers,D.L. 1991 (unpublished). A short overview of $g(0)$ estimation from parallel ship experiments. *IWC Document SC/43/O13.* 18pp.

Borchers,D.L., Buckland,S.T., Clarke,E.D. and Cumberworth,S.L. 1995 (unpublished). Estimation of

cetacean abundance from the SCANS shipboard survey in summer 1994. *IWC Document SC/47/SM36.* 41pp.

Borchers,D.L. and Haw,M.D. 1989 (unpublished). An investigation of duplicate identification criteria for the parallel ship and independent observer experiments. *IWC Document SC/41/SHMi9.* 18pp.

Borchers,D.L. and Haw,M.D. 1990. Determination of minke whale response to a transiting survey vessel from visual tracking of sightings. *Rep. int. Whal. Commn* **40**: 257–269.

Buckland,S.T. 1985. Perpendicular distance models for line transect sampling. *Biometrics* **41**: 117–195.

Buckland,S.T. 1987a. An assessment of the performance of line transect models for fitting IWC/IDCR cruise data, 1978/79 to 1984/85. *Rep. int. Whal. Commn* **37**: 277–279.

Buckland,S.T. 1987b. Estimation of minke whale numbers from the 1984/85 IWC/IDCR Antarctic sightings data. *Rep. int. Whal. Commn* **37**: 263– 268.

Buckland,S.T. 1988. Reliability of hazard–rate line transect estimates from IWC/IDCR Antarctic minke whale sightings data. *Rep. int. Whal. Commn* **38**: 82.

Buckland,S.T. 1992a. Fitting density functions using polynomials. *Applied Statistics* **41**: 63–76.

Buckland,S.T. 1992b. Maximum likelihood fitting of Hermite and simple polynomial densities. *Applied Statistics* **41**: 241–266.

Buckland,S.T. 1992c. Effects of heterogeneity on estimation of probability of detection on the trackline. *Rep. int. Whal. Commn* **49**: 569–573.

Buckland,S.T., Anderson,D.R., Burnham,K.P. and Laake,J.L. 1993a. *Distance Sampling: Estimating abundance of biological populations.* Chapman & Hall, London. 446pp.

Buckland,S.T. and Anganuzzi,A.A. 1988. Comparison of smearing methods in the analysis of minke sightings data from IWC/IDCR Antarctic cruises. *Rep. int. Whal. Commn* **38**: 257–263.

Buckland,S.T., Breiwick,J.M, Cattanach,K.L. and Laake,J.L. 1993b. Estimating population size of the California gray whale. *Mar. Mamm. Sci.* **9**: 235–249.

Buckland,S.T. and Turnock,B.J. 1992. A robust line transect method. *Biometrics* **48**: 901–909.

Burnham,K.P. 1972. Estimation of population size in multinomial capture recapture studies when capture probabilities vary among animals. *PhD. Dissertation.* Oregon State University, Corvallis, Oregan, USA.

Burnham,K.P. 1979. A parametric generalization of the Hayne estimator for line transect sampling. *Biometrics* **35**: 587–595.

Burnham,K.P. and Anderson,D.R. 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics* **32**: 325–336.

Burnham,K.P., Anderson,D.R. and Laake,J.L. 1979. Robust estimation from line transect data. *Journal of Wildlife Management* **43**: 992–996.

Burnham,K.P., Anderson,D.R. and Laake,J.L. 1980. Estimation of density from line transect sampling of animal populations. *Wildlife Monograph Number* **72**.

Burnham,K.P. and Overton,W.S. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrics* **65**: 625–633.

Burnham,K.P., Anderson,D.R., White,G.C., Brownie,C. and Pollock,K.H. 1987. Design and analysis methods for fish survival experiments based on release– recapture. *American Fisheries Society*, Monograph No. **5**.

Butterworth,D.S. 1982a. On the functional form used for $g(y)$ for minke whale sightings, and bias in its estimation due to measurement inaccuracies. *Rep. int. Whal. Commn* **32**: 883–888.

Butterworth,D.S. 1982b. A possible basis for choosing a functional form for the distribution of sightings with right angle distance: some preliminary ideas. *Rep. int. Whal. Commn* **32**: 555–558.

Butterworth,D.S. 1986. A note on the analysis of the 1980/81 variable speed experiment. *Rep. int. Whal. Commn* **36**: 485–489.

Butterworth,D.S. 1991. The use of $g(0) = 1$ for Southern Hemisphere minke whale sighting surveys. *Rep. int. Whal. Commn* **41**: 126

Butterworth,D.S., Best,P.B. and Basson,M. 1982. Results of analysis of sightings experiments carried out during the 1980/81 southern hemisphere minke whale assessment cruise. *Rep. int. Whal. Commn* **32**: 819–834.

Butterworth,D.S. and Rickett,L.H. and Best,P.B. 1984a (unpublished). Preliminary analysis of results from the 1983/84 IWC/IDCR minke whale assessment cruise in Antarctic Area VI. *IWC Document SC/36/Mi13*. 26pp.

Butterworth,D.S., Best,P.B. and Hembree,D. 1984b. Analysis of experiments carried out during the 1981/82 IWC/IDCR Antarctic minke whale assessment cruise in Area II. *Rep. int. Whal. Commn* **34**: 365–392.

Butterworth,D.S. and Hiby,A.R. 1984. Estimates of $g(0)$ from variable speed and parallel ship experiments. Report of the sub–committee on Southern Hemisphere minke whales, Appendix 6. *Rep. int. Whal. Commn* **34**: 91.

Butterworth,D.S. and McQuaid,L.H. 1985 (unpublished). Preliminary analysis of certain sighting experiments carried out on the 1984/85 IWC/IDCR Antarctic minke whale assessment cruise. *IWC Document SC/J85/S6*. 33pp.

Butterworth,D.S. and Borchers,D.L. 1988. Estimates of g(0) for minke schools from the results of the independent observer experiment on the 1985/86 and 1986/87 IWC/IDCR Antarctic assessment cruises. *Rep. int. Whal. Commn* **38**: 301–313.

Butterworth,D.S. and Brown,M.R. 1995 (unpublished). A comparison between barrel sightings of minke

whales made in the NASS and IO mode IWC/IDCR surveys, with implications for the value of $g(0)$ for the former. *IWC Document SC/47/O21.*

Caughley,G. 1974. Bias in aerial survey. *Journal of Wildlife Management* **38**: 921–933.

Chen,S.X. 1996. Studying school size effects in line transect sampling using the kernel method. *Applied Statistics* **45**: 135-150.

Cook,D.R. and Jacobson,J.O. 1979. A design for estimating visibility bias in aerial surveys. *Biometrics* **35**: 735–742.

Cooke,J.G. 1984. The estimation of whale populations of non–uniform density from sightings, with implications for survey design. *IWC Document SC/36/Mi29.* 12pp.

Cooke,J.G. 1985. Notes on the estimation of whale density from line transects. *Rep. int. Whal. Commn* **35**: 319–323.

Cooke,J.G. 1987a (unpublished). Further analyses of parallel ship experiments conducted during the 1984/85 IDCR cruise. *IWC Document SC/39/Mi22.* 5pp.

Cooke.J.G. 1987b (unpublished). Some notes on independent observer experiments. *IWC Document IWC/IDCR/10thMi/SM1.* 6pp.

Cooke,J.G. 1987c. Estimation of the population of minke whales in Antarctic Area IVW in 1984/85. *Rep. int. Whal. Commn* **37**: 273–276.

Cooke,J.G. 1989 (unpublished). More about $g(0)$. *IWC Document SC/41/SHMi22.* 6pp.

Cooke,J.G. 1995. Further analyses for the estimation of $g(0)$ and effective strip width for minke whale sightings. *IWC Document SC/47/NA14.* 15pp.

Cormack,R.M. 1989. Log–linear models for capture–recapture. *Biometrics* **45**: 395–413.

Cox,D.R. and Oakes,D. 1984. *Analysis of survival data.* Chapman & Hall. London. 201pp.

Crain,B.R., Burnham,K.P., Anderson,D.R. and Laake,J.L. 1978. *A Fourier series estimator of population density for line transect sampling.* Utah State University Press, Logan, UT, USA.

Crain,B.R., Burnham,K.P., Anderson,D.R. and Laake,J.L. 1979. Nonparametric estimation of population density for line transect sampling using Fourier series. *Biometrical Journal* **21**: 731–748.

deVries,P.G. 1979. A generalization of the Hayne–type estimator as an example of line intercept sampling. *Biometrics* **35**: 743–748.

Drummer,T.D. and McDonald,L.L. 1987. Size bias in line transect sampling. *Biometrics* **43**: 13–21.

Eberhardt,L.L. 1978. Transect methods for population studies. *Journal of Wildlife Managemant* **42**: 1–31.

Evans,M.A. and Bonnet,D.G. 1993. A constrained Cook–Jacobson model of visibility bias. *Biometrics* **49**: 853–859.

Evans,M.A., Bonnet,D.G. and McDonald,L.L. 1994. A general theory for modelling capture–recapture data from a closed population. *Biometrics* **50**: 396–405.

Folklow,L.P. and Blix,A.S. 1993. Daily changes in surfacing rates of minke whales in Norwegian waters. *Rep. int. Whal. Commn* **43**: 311–314.

Gates,C.E. 1969. Simulation study of estimators for the line transect sampling method. *Biometrics* **25**: 317–328.

Gates,C.E. and Smith,P.W. 1980. An implementation of the Burnham–Anderson distribution free method of estimating wildlife densities from line transect data. *Biometrics* **36**: 155–160.

Gunlaugsson,T. 1990 (unpublished). A minke whale sightings experiment on board *Hvalur 9* in Faxafloi, southwest Iceland, July–August 1988. *IWC Document SC/41/NHMi14.* 18pp.

Gunlaugsson,T. 1991a (unpublished). Effect of Beaufort on minke whale sightings rate in Icelandic whale observation surveys 1982–1986. *IWC Document SC/42/NHMi25.* 10pp.

Gunlaugsson,T. 1991b (unpublished). Reanalysis of a minke whale sightings experiment on board *Hvalur 9* in Faxafloi, southwest Iceland, July– August 1988. *IWC Document SC/42/NHMi1* 3pp.

Gunlaugsson,T. and Sigurjonsson,J. 1990. NASS–87: estimation of whale abundance based on observations made onboard Icelandic and Faroese survey vessels. *Rep. int. Whal. Commn* **40**: 571–580.

Hammond,P.S. 1984. An investigation into the effects of different techniques of smearing the IWC/IDCR minke whale sightings data and the use of different models to estimate density of schools. *Rep. int. Whal. Commn* **34**: 301–307.

Hastie,T.J. and Tibshirani,R.J. 1990. *Generalized additive models.* Chapman & Hall, London.

Hayes,R.J. and Buckland,S.T. 1983. Radial distance models for the line transect method. *Biometrics* **39**: 29–42.

Hayne,D.W. 1949. An examination of the strip census method for estimating animal populations. *Journal of Wildlife Management* **13**: 145–157.

Hiby.A.R. 1986. Results of a hazard rate model relevant to experiments on the 1984/85 IDCR minke whale assessment cruise. *Rep. int. Whal. Commn* **36**: 497–498.

Hiby,A.R. and Hammond,P.S. 1989. Survey techniques for estimating current abundance and monitoring trends in abundance of cetaceans. In *The Comprehensive Assessment of Whale Stocks: the early years* (ed. G.P. Donovan). *Rep. int. Whal. Commn Special Issue 11*:: 47–80.

Hiby,A.R. and Lovell,P. 1996. Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Submitted to Biometrics.* 22pp.

Holt,S.R. 1987. Estimating density of dolphin schools in the eastern tropical Pacific ocean by line transect methods. *Fishery Bulletin* **85**: 419– 434.

Holt,S.R. and Cologne,J. 1987. Factors affecting line transect estimates of dolphin school density. *Journal*

of Wildlife Management **51**: 836– 843.

Horvitz,D.G. and Thompson,D.J. 1952. A generalization of sampling without replacement from a finite universe. *J.A.S.A.* **47**: 663–685.

Huggins,R.M. 1989. On the statistical analysis of capture experiments. *Biometrika* **76**: 133–140.

Johnson,E.G. and Routledge,R.D. 1985. The line transect method: a nonparametric estimator based on shape restrictions. *Biometrics* **41**: 669–679.

Joyce,G. 1982. Blow patterns as sighting cues for censusing minke whales in Antarctic waters. *Rep. int. Whal. Commn* **32**: 787–790.

Kelker,G.H. 1945. Measurement and interpretation of forces that determine populations of managed deer. *PhD Disseration*, University of Michigan, Ann Arbor, MI, USA.

Kishino,H. 1986. On parallel ship experiments and the line transect method. *Rep. int. Whal. Commn* **36**: 491–495.

Kishino,H., Kasamatsu,F. and Miyashita,T. 1986. Difference of detection functions between two ships in PS experiment and *eh* factor. *Rep. int. Whal. Commn* **36**: 467–468.

Laake,J.L., Rugh,D.J., Lerczak,J.A. and Buckland,S.T. 1994 (unpublished). Preliminary estimates of population size of gray whales from the 1992/93 and 1993/94 shore–based surveys. *IWC Document SC/46/AS7.* 13pp.

Lawson,C.L. and Hanson,R.J. 1974. *Solving least squares problems.* Englewood Cliffs, New Jersey, Prentice–Hall.

Leatherwood,S., Aubrey,F.T. and Thomas,J.A. 1982. Minke whale response to a transiting survey vessel. *Rep. int. Whal. Commn* **32**: 795–803.

Oien, N.1990. Estimates of $g(0)$ for minke whales based on an independent observer experiment during the Norwegian sightings surveys in July 1988. *Rep. int. Whal. Commn* **40**: 331–335.

Oien, N.1993 (unpublished). Estimates of $g(0)$ for harbour porpoise based on a shipboard survey in the North Sea in 1990. *IWC Document SC/44/SM7.* 4pp.

Otis,D.L., Burnham,K,P., White,G.C. and Anderson,D.R. 1978. Statistical inference for capture data on closed animal populations. *Wildlife Monograph* **62**. 135pp.

Otten,A. and deVries,P.G. 1984. On line transect estimates for population density based on elliptic flushing curves. *Biometrics* **40**: 1145– 1150.

Otto, M.C. and Pollock, K.H. 1990. Size bias in line transect sampling: a field test. *Biometrics* **46**: 239–245.

Palka,D.L. 1992 (unpublished). Abundance estimate of the gulf of Maine harbour porpoise. *IWC Document SC/44/SM24.* 24pp.

Palka,D.L. 1993. Estimating density of animals when assumptions of line transect survey are violated. *PhD Dissertation*, University of California, San Diego. 169pp.

Parzen,E. 1960. *Modern probability theory and its applications.* John Wiley & Sons, New York.

Patil, G.P., Taille, C. and Talwalker, S. 1993. Encounter sampling and modelling in ecological and environmental studies using weighted distribution methods. In *Statistics for the environment* (eds V.Barnett and K.F.Turkman), John Wiley & Sons, New York: 45–69.

Polacheck,T. 1990. The swimming direction of harbour porpoise in relationship to a survey vessel. *Rep. int. Whal. Commn* **40**: 463–470.

Polacheck,T. 1991a (unpublished). Simulation results on the effects of dive time and movements on line transect estimates. *IWC Document SC/42/NHMi31.* 13pp.

Polacheck,T. 1991b (unpublished). Results of a field test of line transect methods for shipboard sighting surveys for harbour porpoise. *IWC Document SC/42/SM38.* 24pp.

Polacheck,T. 1992 (unpublished). The effect of increasing observer trackline effort in shipboard line transect surveys for harbour porpoise. *IWC Document SC/43/SM10.* 24pp.

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

Pollock,K.H. and Otto,M.C. 1983. Robust estimation of population size in closed animal populations from capture–recapture experiments. *Biometrics* **39**: 1035–1049.

Pollock,K.H., Hines,J.E. and Nichols,D. 1984. The use of auxiliary variables in capture–recapture and removal experiments. *Biometrics* **40**: 329- -340.

Pollock,K.H. and Kendall,W.L. 1987. Visibility bias in aerial surveys: a review of estimation procedures. *Journal of Wildlife Management* **51**: 502–510.

Quang,P.X. 1991. A nonparametric approach to size–biased line transect sampling. *Biometrics* **47**: 269–79.

Quang,P.X. and Lanctot,R.B. 1991. A line transect model for aerial surveys. *Biometrics* **47**: 1089–1102.

Quinn.T.J. 1977. The effects of aggregation on line transect estimators of population abundance with application to marine mammal populations. *MS thesis.* University of Washington, WA, USA.

Quinn,T.J. and Galluci,V.F. 1980. Parametric models for line transect estimators of abundance. *Ecology* **61**: 293–302.

Ramsay,F.L. 1979. Parametric models for line transect surveys. *Biometrika* **66**: 505–512.

Ramsey,F.L, Wildman,V. and Engbring,J. 1987. Covariate adjustments to effective area in variable–area wildlife surveys. *Biometrics* **43**: 1–11.

Robinette,W.L., Loveless,C.M. and Jones,D.A. 1974. Field tests of strip census methods. *Journal of Wildlife Management* **38**: 81–96.

Rivest,L., Potvin,F., Crepeau,H. and Daigle,G. 1995. Statistical methods for aerial surveys using the double–count technique to correct visibility bias. *Biometrics* **51**: 461–470.

Sandland,R.L. and Cormack,R.M. 1984. Statistical inference for Poisson and multinomial models for capture–recapture experiments. *Biometrika* **71**: 27–33.

Schweder,T. 1974 (unpublished). Transformations of point processes: applications to animal sighting and catch problems, with special emphasis on whales. *PhD thesis*, University of California, Berkeley, CA, USA.

Schweder,T. 1977. Point process models for line transect experiments. In *Recent developments in statistics* (eds J.R. Barba, F. Brodeau, G. Romier and B. van Custem), North–Holland Publishing Company, New York, USA: 221–242.

Schweder,T. 1990. Independent observer experiments to estimate the detection function in line transect surveys of whales. *Rep. int. Whal. Commn* **40**: 349–356.

Schweder,T. Host,G. and Oien,N. 1991. A note on the bias in capture–recapture type estimates due to the fact that whales are diving. *Rep. int. Whal. Commn* **41**: 417–432.

Schweder,T. and Host,G. 1992. Integrating experimental data and survey data to estimate $g(0)$: a first approach. *Rep. int. Whal. Commn* **42**: 575– 582.

Seber,G.A.F. 1982. *The estimation of animal abundance and related parameters.* Charles Griffin & Company,London.

Sen,A.R, Tourigny,J. and Smith,G.E.J. 1974. On the line transect sampling method. *Biometrics* **30**: 329–341.

Steinhorst,R.K. and Samuel,M.D. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. *Biometrics* **45**: 415–425.

Thompson,S.K. 1992. *Sampling.* John Wiley & Sons. New York. 343pp.

Thompson,S.K. and Seber,G.A.F. 1994. Detectability in conventional and adaptive sampling. *Biometrics* **50**: 712–724.

Ward,A.J. 1988. Accounting for the effect of pod size in the estimation of blow rates: analysis of data from the 1984/85, 85/86 and 86/87 IDCR minke whale assessment cruises. *Rep. int. Whal. Commn* **38**: 265–268.

Ward,A.J., Hiby,A.R. and Thompson,D. 1986. Photographic estimation of angles to sightings and distribution of sighting effort with angle – results obtained during the 1984/85 IWC/IDCR minke whale assessment cruise. *Rep. int. Whal. Commn* **326**: 477–479.

Wong, C-N. 1996. Population size estimation using the modified Horvitz–Thompson estimator with estimated sighting probability. *PhD. Dissertation.* Colorado State University, Fort Collins, Co 80523, USA.

Zahl,S. 1989a. Line transect samping with unknown probability of detection along the transect. *Biometrics* **45**: 453–470.

Zahl,S. 1989b (unpublished). Analysis of the 1984/85 Antarctic variable speed experiment. *IWC Document SC/41/SHMi8.*