

INFERENCE FOR PLANT-CAPTURE

Jonathan Ashbridge

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1998

Full metadata for this item is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/13741>

This item is protected by original copyright



University of St. Andrews, Scotland

Inference for Plant-Capture

A thesis submitted to the University of St. Andrews for the degree of
Doctor of Philosophy

by

Jonathan Ashbridge

September 1997



ProQuest Number: 10167340

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10167340

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Th C 421

I, Jonathan Ashbridge, hereby certify that this thesis, which is approximately 27000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 4/12/97 signature of candidate

I was admitted as a research student in October 1994 and as a candidate for the degree of Doctor of Philosophy in October 1995 ; the higher study for which this is a record was carried out in the University of St. Andrews between 1994 and 1997.

date 4/12/97 signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date signature of supervisor

In Submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date 4/12/97 signature of candidate

I would like to thank my research supervisor Dr. I. B. J. Goudie for his advice, help and encouragement during the work which led to the production of this thesis.

I would like to acknowledge the support of the Engineering and Physical Sciences Research Council.

Abstract

When investigating the dynamics of an animal population, a primary objective is to obtain reasonable estimates of abundance or population size. This thesis concentrates on the problem of obtaining point estimates of abundance from capture-recapture data and on how such estimation can be improved by using the method of plant-capture.

Plant-capture constitutes a natural generalisation of capture-recapture. In a plant-capture study a pre-marked population of known size is added to the target population of unknown size. The capture-recapture experiment is then carried out on the augmented population.

Chapter 1 considers the addition of planted individuals to target populations which behave according to the standard capture-recapture model M_0 . Chapter 2 investigates an analogous model based on sampling in continuous time. In each of these chapters, distributional results are derived under the assumption that the behaviour of the plants is indistinguishable from that of members of the target population. Maximum likelihood estimators and other new estimators are proposed for each model. The results suggest that the use of plants is beneficial, and furthermore that the new estimators perform more satisfactorily than the maximum likelihood estimators.

Chapter 3 introduces, initially in the absence of plants, a new class of estimators, described as coverage adjusted estimators, for the standard capture-recapture model M_n . These new estimators are shown, through simulation and real life data, to compare favourably with estimators that have previously been proposed. Plant-capture versions of these new estimators are then derived and the usefulness of the plants is demonstrated through simulation.

Chapter 4 describes how the approach taken in chapter 3 can be modified to produce a new estimator for the analogous continuous time model. This estimator is then shown through simulation to be preferable to estimators that have previously been proposed.

Table of Contents

Signature page	ii
Acknowledgements	iii
Abstract	iv
Table of contents	v
Introduction	viii
Chapter 1 : Plant-Capture Applied to the Model M_0 :	
Discrete Time Sampling Procedure	1
§ 1.1 : Introduction	1
§ 1.2 : Sampling Procedure and Assumptions	1
§ 1.3 : The Sufficient Statistics	2
§ 1.4 : The Distribution Function of the Sufficient Statistics	4
§ 1.5 : The δ - Numbers	5
§ 1.6 : The Maximum Likelihood Estimator	8
§ 1.7 : A Peterson-Type Estimator	9
§ 1.8 : A Conditionally Unbiased Estimator	10
§ 1.8a : A Note on the Evaluation of the CUE	12
§ 1.9 : A Comparison of All Three Estimators	15
§ 1.9a : Discussion	17
§ 1.10 : The Unconditional Performance of the CUE and the Peterson-Type Estimators	31
§ 1.10a : Discussion	31
§ 1.11 : The Performance of Plant-Capture When Applied to the Model M_0 : Under Discrete Time Sampling	39
Chapter 2 : A Plant-Capture Approach for Sequential Tagging	42
§ 2.1 : Introduction	42
§ 2.2 : Sampling Procedure and Assumptions	42
§ 2.3 : A Note on Software Reliability	43
§ 2.4 : The Sufficient Statistics	43
§ 2.5 : The Distribution Function of the Sufficient Statistics	46
§ 2.6 : The Q - Numbers	48
§ 2.7 : The Maximum Likelihood Estimator	50
§ 2.8 : The Harmonic Mean Estimator	51
§ 2.9 : A Peterson-Type Estimator	52

§ 2.9a : A New Estimator for Homogeneous Populations	52
§ 2.10 : A Conditionally Unbiased Estimator	54
§ 2.10a: A Note on the Evaluation of the CUE	55
§ 2.11 : A Comparison of All Four Estimators	58
§ 2.11a: The c Constant	61
§ 2.11b: Discussion of Relative Performance of Estimators	62
§ 2.12 : The Unconditional Performance of the CUE, HME and Peterson-Type Estimators	73
§ 2.12a: Discussion	74
§ 2.13 : The Performance of Plant-Capture When Applied to the Model M_0 : Under Continuous Time Sampling	85
Chapter 3 : Estimation Under the Capture-Recapture Model M_h : Discrete Time Sampling Procedure	88
§ 3.1 : Introduction	88
§ 3.2 : Sampling Procedure and Assumptions	88
§ 3.3 : A Class of Coverage Adjusted Estimators for the Model M_h	89
§ 3.4 : Other Estimators for the Model M_h	93
§ 3.5 : Simulation Study	94
§ 3.6 : Some Standard Data Sets	119
§ 3.7 : Plant-Capture Applied to the Model M_h	125
§ 3.7.1 : Introduction	125
§ 3.7.2 : Sampling Procedure, Assumptions and Some Additional Notation	125
§ 3.7.3 : Some Distribution Theory	126
§ 3.7.4 : A Peterson-Type Estimator	127
§ 3.7.5 : Plant-Capture Versions of the Overton and Coverage Adjusted Estimators	127
§ 3.7.6 : Plant-Capture Simulation Study	129
Chapter 4 : Estimation Under the Capture-Recapture Model M_h : Continuous Time Sampling Procedure	143
§ 4.1 : Introduction	143

§ 4.2	: Sampling Procedure, Assumptions, Some Notation and the Sufficient Statistics	143
§ 4.3	: A New Coverage Adjusted Estimator for the Model M_h	144
§ 4.4	: Simulation Study	146
Chapter 5	: Conclusions	155
§ 5.1	: Initial Objectives and Results Obtained	155
Appendix 1	: The Classical Occupancy Distribution	158
Appendix 2	: The Stirling Distribution of the Second Kind	160
Appendix 3	: The Distribution of A Sum of Zero Truncated Binomial Random Variables	162
Appendix 4	: Estimation of Sample Coverage : Model M_h : Discrete Time Sampling Procedure	164
Appendix 5	: Estimation of Sample Coverage : Model M_h : Continuous Time Sampling Procedure	166
References		168

Introduction

Capture-recapture methods can be used to estimate population size and other fundamental demographic variables. The most popular class of models that describe the behaviour of closed populations were first introduced as a set by Pollock(1974, 1976) and later more fully described in a wildlife monograph by Otis et al. (1978). Another important reference for this class of models is White et al. (1982). Each model within the class requires a sequence of t samples to be taken from the population. After each sample is taken animals within the sample not previously caught each receive a unique tag so that they can be recognised if recaptured in a later sample. After each sample is taken all animals are released. In each of their models Otis et al. (1978) allow the capture probabilities to vary due to time(t), due to heterogeneity (h) between the capture probabilities of the different animals, and due to a behavioural (b) response to the traps used. A total of eight possible models within this class results from the fact that each of these three factors can be present or absent. The sampling scheme considered within the Otis et al. (1978) monograph is referred to as discrete time sampling.

There is a continuous time sampling analogue of each of the models described by Otis et al. (1978), in which the population is under continuous observation for some period of time, with the animals being seen according to independent Poisson processes. For this continuous time sampling procedure one animal is seen at a time and animals seen for the first time receive a unique tag so that they may be subsequently recognised.

Plant-capture constitutes a natural generalisation of capture-recapture. In a plant-capture study a pre-marked population of known size is first added to the target population of unknown size. The capture-recapture experiment is then carried out on the augmented population. Under the assumption that members of the planted population behave in an identical manner to those of the target population, one obtains, through sightings of the plants, additional information which can improve estimation of target population size.

This thesis concentrates on closed populations which behave according to two of the eight closed capture-recapture models described by Otis et al. (1978). The most basic model M_0 is considered in chapter 1, and the important heterogeneity model M_h is considered within chapter 3. While chapter 1 is largely concerned with the case where plants are present, the emphasis in chapter 3 is on proposing an improved class of estimators for the standard model M_h .

The work contained within chapters 2 and 4, which considers the continuous time analogues of the models considered in chapters 1 and 3 respectively, could also be applied in a software reliability context where the population in question is one of errors or 'bugs' in a computer program.

The assumption that members of the planted and target populations behave in an identical manner is central to most plant-capture methodology. Indeed one should only apply plant-capture methods when there are adequate grounds for believing that this assumption is a reasonable approximation to reality.

The use of plants to assist population size estimation has been considered in a number of quite different situations. Change-in-ratio methods have been widely used to estimate the abundance of animal populations, see Seber(1982, chapter 9). It was Kelker(1940) who first introduced the idea that the size of a wildlife population could be estimated from a knowledge of sex ratios before and after a differential kill of the sexes. Rupp(1966) recognised that the theory was still valid when the ratios are changed by the insertion of planted individuals. In a software reliability context the idea of introducing plants into the population prior to sampling has been considered by Mills(1972), and by Duran & Wiorowski(1981) who speak of 'deliberately seeding errors into the software' prior to testing. Laska & Meisner(1993) have described how the U.S. Census Bureau used a plant-capture method in an attempt to estimate the size of a selected component population of homeless people. Laska & Meisner(1993) state that 'there are many potential applications' of their methodology. Martin et al. (1995) have also investigated using a plant-capture method for estimating the size of the street dwelling population. Goudie(1995) has considered the use of plants in order to improve stopping rules for determining , within a specified error probability, when all members of a target population have been seen. Yip(1996) describes a martingale based approach for estimating population size from plant-capture data. Norris and Pollock(1996b) have considered the use of plants in connection with the heterogeneity model M_h .

N.B. Although every effort has been made to ensure that the notation used throughout this thesis remains consistent, the notation used to denote estimators must be viewed as being specific to each chapter.

Chapter 1 : Plant-Capture Applied to the Model M_0 :

Discrete Time Sampling Procedure

§ 1.1 : Introduction

This chapter considers how the method of plant-capture may be used to aid the problem of estimating population size in a multiple capture-recapture experiment when the population in question behaves according to the standard capture-recapture model known as M_0 . The model M_0 is one of the set of models described by Otis et al. (1978) for capture-recapture data in closed populations.

The discrete time sampling procedure considered within this chapter essentially constitutes what is known in the literature as a Schnabel Census with random sample sizes, see Schnabel(1938) or, for a more comprehensive review, Seber(1982). In the absence of plants the most commonly used estimator for the model M_0 , under discrete time sampling, is the maximum likelihood estimator, which was first considered by Darroch(1958) and later by Otis et al. (1978).

§ 1.2 : Sampling Procedure and Assumptions

The sampling procedure may be described as follows. Prior to the commencement of the experiment it is assumed that the target population, whose size N we wish to estimate, is augmented by the insertion of a known number R of planted individuals. Each planted individual is assumed to have received a unique tag prior to its release. A sequence of t sampling experiments is then carried out on the augmented population which is assumed to be closed and of size $N+R$. Independently of other animals and independently of its previous capture history animal i ($i=1,2,\dots,N+R$) is captured in sample j ($j=1,2,\dots,t$) with probability p . After each sample is taken every animal within that sample not previously marked receives a unique tag before its release so that it may be recognised on subsequent trapping occasions. The experiment generates an $(N+R)$ by t matrix A where

$$a_{ij} = \begin{cases} 1 & \text{if animal } i \text{ is caught on sampling occasion } j \\ 0 & \text{if animal } i \text{ is not caught on sampling occasion } j \end{cases}$$
$$i = 1, 2, \dots, N+R.$$
$$j = 1, 2, \dots, t.$$

The sample space is the set of such matrices.

In the absence of plants, the sampling procedure considered here is the one most commonly used in practice.

Please note that the above implies that the behaviour of the planted individuals is assumed to be indistinguishable from the behaviour of members of the original population.

§ 1.3 : The Sufficient Statistics

In order to obtain the sufficient statistics some notation is needed :

Let $X_w^{(1)}$ \equiv the number of animals from the target population with capture history w . For example, for $t = 3$, $X_{101}^{(1)}$ is the number of animals seen on the first and third but not the second capture periods.

$X_w^{(2)}$ \equiv the number of animals from the planted population with capture history w .

$X_1 = \sum_w X_w^{(1)}$ \equiv the number of distinct animals seen from the target population.

$X_2 = \sum_w X_w^{(2)}$ \equiv the number of distinct animals seen from the planted population.

N.B. \sum_w is used to represent the summation over all w except $w = \underline{0}$.

$X = X_1 + X_2$ \equiv the number of distinct animals seen from the augmented population.

i_w \equiv the number of ones in w .

$Z_1 = \sum_w i_w X_w^{(1)}$ \equiv the total number of captures from the target population.

$Z_2 = \sum_w i_w X_w^{(2)}$ \equiv the total number of captures from the planted population.

$Z = Z_1 + Z_2$ \equiv the total number of captures from the augmented population.

And let $\{X_w^{(i)}\}$ denote the vector of the $X_w^{(i)}$'s, except for the unobservable $X_0^{(i)}$, $i=1, 2$.

For the moment if we consider only the target population it is seen that the distribution of $\{X_w^{(1)}, X_0^{(1)}\}$ is multinomial, with 2^t cells and N trials. Given $\{X_w^{(1)}\}$ and N one can easily deduce the unobservable value of $X_0^{(1)}$ so we consider $\text{Prob}(\{X_w^{(1)}\})$ in place of $\text{Prob}(\{X_w^{(1)}, X_0^{(1)}\})$.

$$\text{Prob}(\{X_w^{(1)}\}) = \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right)} (P_0)^{N-x_1} \prod_w (P_w)^{X_w^{(1)}},$$

where P_w denotes the cell probability of capture history w , and \prod_w is used to represent the product over all w except

$$w = 0.$$

$$= \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right)} [(1-p)^t]^{N-x_1} \prod_w [p^{i_w} (1-p)^{t-i_w}]^{X_w^{(1)}}$$

$$= \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right)} [(1-p)^t]^{N-x_1} p^{\sum_w i_w X_w^{(1)}} (1-p)^{\sum_w (t-i_w) X_w^{(1)}}$$

$$= \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right)} [(1-p)^t]^{N-x_1} p^{z_1} (1-p)^{t x_1 - z_1}$$

$$= \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right)} p^{z_1} (1-p)^{t(N-x_1)}.$$

This result was first obtained by Darroch(1958).

In a similar way it may be shown that

$$\text{Prob}(\{X_w^{(2)}\}) = \frac{R!}{(R-x_2)! \left(\prod_w X_w^{(2)}! \right)} p^{z_2} (1-p)^{tR-z_2}.$$

In view of the independence between target and planted populations we may write

$$\begin{aligned} \text{Prob}(\{X_w^{(1)}, X_w^{(2)}\}) &= \text{Prob}(\{X_w^{(1)}\}) \cdot \text{Prob}(\{X_w^{(2)}\}) \\ &= \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right)} p^{z_1} (1-p)^{t(N-x_1)} \frac{R!}{(R-x_2)! \left(\prod_w X_w^{(2)}! \right)} p^{z_2} (1-p)^{tR-z_2} \\ &= \frac{N!}{(N-x_1)! \left(\prod_w X_w^{(1)}! \right) (R-x_2)! \left(\prod_w X_w^{(2)}! \right)} p^z (1-p)^{t(N+R)-z} \end{aligned}$$

which implies that

$$L(N, p | X_w^{(1)}, X_w^{(2)}) \propto \frac{N!}{(N-x_1)!} p^z (1-p)^{t(N+R)-z}$$

where $L(N, p | X_w^{(1)}, X_w^{(2)})$ represents the likelihood function for N and p . Hence by the Neyman-Pearson factorisation theorem it is seen that the sufficient statistics for N and p are in fact x_1 and $z = z_1 + z_2$.

§ 1.4 : The Distribution Function of the Sufficient Statistics

The joint distribution function of X_1 and Z may be obtained as follows :

$$\begin{aligned}
 \text{Prob}(X_1 = x_1, Z = z) &= \text{Prob}(Z = z)\text{Prob}(X_1 = x_1|Z = z) \\
 &= \text{Prob}(Z = z)\sum_{z_1} \text{Prob}(X_1 = x_1|Z_1 = z_1, Z = z)\text{Prob}(Z_1 = z_1|Z = z) \\
 &= \text{Prob}(Z = z)\sum_{z_1} \text{Prob}(X_1 = x_1|Z_1 = z_1)\text{Prob}(Z_1 = z_1|Z = z) \\
 &= \text{Prob}(Z = z)\sum_{z_1} \frac{\text{Prob}(Z_1 = z_1|X_1 = x_1)\text{Prob}(X_1 = x_1)}{\text{Prob}(Z_1 = z_1)}\text{Prob}(Z_1 = z_1|Z = z). \quad (1.1)
 \end{aligned}$$

From the above assumptions it is known that

$$X_1 \sim \text{Bin}(N, 1 - (1 - p)^t),$$

$$Z_1 \sim \text{Bin}(Nt, p)$$

and that $Z \sim \text{Bin}((N + R)t, p)$.

The distribution of $Z_1|Z$ is hypergeometric. And we may observe that $Z_1|X_1 = x_1$ is the sum of x_1 zero truncated Binomial random variables, the probability function of which is derived in appendix 3.

Now from (1.1) it follows that

$$\begin{aligned}
 \text{Prob}(X_1 = x_1, Z = z) &= \\
 &= \frac{\binom{(N+R)t}{z} p^z (1-p)^{(N+R)t-z} \sum_{z_1} \frac{p^{z_1} (1-p)^{t x_1 - z_1} \sum_{j=0}^{x_1} \binom{x_1}{j} \binom{tj}{z_1} (-1)^{x_1-j} \binom{N}{x_1} [1 - (1-p)^t]^{x_1} [(1-p)^t]^{N-x_1} \binom{Nt}{z_1} \binom{Rt}{z-z_1}}{\binom{Nt}{z_1} p^{z_1} (1-p)^{Nt-z_1}}}{\binom{(N+R)t}{z}} \\
 &= \binom{N}{x_1} p^z (1-p)^{(N+R)t-z} \sum_{j=0}^{x_1} \binom{x_1}{j} \left[\sum_{z_1} \binom{tj}{z_1} \binom{Rt}{z-z_1} \right] (-1)^{x_1-j} \\
 &= \binom{N}{x_1} p^z (1-p)^{(N+R)t-z} \sum_{j=0}^{x_1} \binom{x_1}{j} \binom{tj + Rt}{z} (-1)^{x_1-j}, \quad (1.2)
 \end{aligned}$$

$$x_1 = 0, 1, 2, \dots, N.$$

$$z = x_1, x_1+1, x_1+2, \dots, tx_1 + tR.$$

The joint probability function, given by (1.2), of the sufficient statistics X_1 and Z has not previously appeared in the literature for values of R greater than or equal to zero. However, the conditional distribution of X_1 given Z has previously been considered in an urn model context. For $R = 0$ the conditional distribution of X_1 given Z is identical to the distribution of the number of occupied urns where there is a limit r on the

capacity of each urn, Romanovsky(1934). Charalambides(1981) generalised this result of Romanovsky(1934) by introducing a control urn of capacity s , where s is not necessarily equal to r . When $\frac{s}{r}$ is integer valued, the situation considered by Charalambides(1981), with $R = \frac{s}{r}$, is probabilistically equivalent to the one considered here, and so led Charalambides(1981) to derive the conditional distribution of X_1 given Z : for values of R greater than or equal to zero.

N.B. The summation which appears in (1.2) is of importance within this chapter and is considered in more detail in the following section.

§ 1.5 : The δ - Numbers

The δ -numbers are defined as follows

$$\delta(x_1, z; t, R) = \sum_{j=0}^{x_1} \binom{x_1}{j} \binom{tj + tR}{z} (-1)^{x_1-j},$$

$$x_1 = 0, 1, 2, \dots$$

$$z = x_1, x_1 + 1, x_1 + 2, \dots, tx_1 + tR.$$

Within this chapter these δ -numbers are of importance since they appear in the joint probability function of X_1 and Z , as given by equation (1.2).

The δ -numbers are multiples of a subset of the Gould-Hopper numbers, see Gould and Hopper(1962), Charalambides(1979) and Charalambides and Singh(1988). Explicitly the Gould-Hopper number is defined as $G(z, x_1, t, s) = \frac{1}{x_1!} [\Delta^{x_1} (ty + s)_z]_{y=0}$.

When $\frac{s}{t}$ is integer valued, the Gould-Hopper number $G(z, x_1, t, s) = \frac{z!}{x_1!} \delta\left(x_1, z; t, \frac{s}{t}\right)$.

The Gould-Hopper numbers are a generalisation of the C-numbers. The C-numbers have been extensively studied, see Charalambides and Singh(1988), and are defined as $C(z, x_1, t) = \frac{1}{x_1!} [\Delta^{x_1} (ty)_z]_{y=0}$. The relationship between the δ -numbers and

the C-numbers is given by

$$C(z, x_1, t) = \frac{z!}{x_1!} \delta(x_1, z; t, R = 0).$$

In order to investigate the distributional properties of the estimators which are considered further on in this chapter it is necessary to evaluate the δ -numbers over some particular range of parameter values. This can lead to computational problems

since the form of the δ -numbers is not desirable from a computational point of view. That is the alternating sign within the summation means that, for large values of N , R and t , very large numbers are repeatedly being added to and in particular subtracted from one another, and this is a major source of rounding error. To help avoid this, and other significant computational problems, one may consider the following 'triangular' recurrence relation of the δ -numbers.

$$z\delta(x_1, z; t, R) = (tx_1 + tR - z + 1)\delta(x_1, z - 1; t, R) + tx_1\delta(x_1 - 1, z - 1; t, R). \quad (1.3)$$

A direct proof of this is as follows

$$\begin{aligned} & (tx_1 + tR - z + 1)\delta(x_1, z - 1; t, R) + tx_1\delta(x_1 - 1, z - 1; t, R) \\ &= (tx_1 + tR - z + 1)\sum_{j=0}^{x_1} \binom{x_1}{j} \binom{tj + tR}{z-1} (-1)^{x_1-j} + tx_1 \sum_{j=0}^{x_1-1} \binom{x_1-1}{j} \binom{tj + tR}{z-1} (-1)^{x_1-1-j} \\ &= (tx_1 + tR - z + 1) \binom{tx_1 + tR}{z-1} + (tx_1 + tR - z + 1) \sum_{j=0}^{x_1-1} \binom{x_1}{j} \binom{tj + tR}{z-1} (-1)^{x_1-j} \\ & \quad + tx_1 \sum_{j=0}^{x_1-1} \binom{x_1-1}{j} \binom{tj + tR}{z-1} (-1)^{x_1-1-j} \\ &= (tx_1 + tR - z + 1) \frac{z}{(tx_1 + tR - z + 1)} \binom{tx_1 + tR}{z} \\ & \quad + \sum_{j=0}^{x_1-1} (-1)^{x_1-j} \binom{tj + tR}{z-1} \left[(tx_1 + tR - z + 1) \binom{x_1}{j} - tx_1 \binom{x_1-1}{j} \right] \\ &= z \binom{tx_1 + tR}{z} \\ & \quad + \sum_{j=0}^{x_1-1} (-1)^{x_1-j} \binom{x_1}{j} \binom{tj + tR}{z} \frac{z}{(tj + tR - z + 1)} \left[(tx_1 + tR - z + 1) - tx_1 \frac{(x_1 - j)}{x_1} \right] \\ &= z \binom{tx_1 + tR}{z} + \sum_{j=0}^{x_1-1} (-1)^{x_1-j} \binom{x_1}{j} \binom{tj + tR}{z} \frac{z}{(tj + tR - z + 1)} [tj + tR - z + 1] \\ &= z \binom{tx_1 + tR}{z} + z \sum_{j=0}^{x_1-1} (-1)^{x_1-j} \binom{x_1}{j} \binom{tj + tR}{z} \\ &= z \sum_{j=0}^{x_1} \binom{x_1}{j} \binom{tj + tR}{z} (-1)^{x_1-j} \\ &= z\delta(x_1, z; t, R). \end{aligned}$$

Equation (1.3), when $R=0$, essentially reduces to the recurrence relation of the C -numbers as given by equation (3.25) in Charalambides and Singh(1988).

The triangular recurrence relation (1.3) along with the initial conditions

$$\delta(0, z; t, R) = \binom{tR}{z}, \quad \delta(x_1, x_1; t, R) = t^{x_1} \quad \text{and} \quad \delta(x_1, tx_1 + tR; t, R) = 1 \quad (1.3a)$$

enables one to evaluate the required δ -numbers without having to perform any subtraction operations whatsoever, and hence one can more easily avoid computational rounding error.

N.B. The first and third initial conditions are easy to show directly. The second can be shown to hold as follows. Firstly substituting $z = x_1$ into (1.3) implies that

$\delta(x_1, x_1; t, R) = t\delta(x_1 - 1, x_1 - 1; t, R)$, then after observing that $\delta(0, 0; t, R) = 1$ it is easy to see that $\delta(x_1, x_1; t, R) = t^{x_1}$ for all $x_1 \geq 0$.

Comments

Using (1.3), one can show that a similar 'triangular' recurrence relation exists between the probabilities of the joint distribution of X_1 and Z , given by equation (1.2).

It can be shown that

$$P_{x_1, z} = \frac{p}{z(1-p)} \left[(tx_1 + tR - z + 1)P_{x_1, z-1} + t(N - x_1 + 1)P_{x_1-1, z-1} \right], \quad (1.4)$$

$$\text{where} \quad P_{x_1, z} = \text{Prob}(X_1 = x_1, Z = z).$$

It is also straightforward to show that (1.4) is subject to the initial conditions

$$P_{0, z} = p^z (1-p)^{tN+tR-z} \binom{tR}{z}, \quad z = 0, 1, 2, \dots, tR. \quad (1.4a)$$

$$P_{x_1, x_1} = \binom{N}{x_1} p^{x_1} (1-p)^{tN+tR-x_1} t^{x_1}, \quad x_1 = 0, 1, 2, \dots, N. \quad (1.4b)$$

$$\text{and} \quad P_{x_1, tx_1+tR} = \binom{N}{x_1} p^{tx_1+tR} (1-p)^{tN-tx_1}, \quad x_1 = 0, 1, 2, \dots, N. \quad (1.4c)$$

Again in an attempt to avoid numerical computational problems, one can determine the initial conditions (1.4a) and (1.4b) using the following recurrence relations :

$$(i) \quad P_{0, z} = \frac{p}{(1-p)} \left[\frac{tR - z + 1}{z} \right] P_{0, z-1}, \quad z = 1, 2, \dots, tR.$$

$$(ii) \quad P_{x_1, x_1} = \frac{tp}{(1-p)} \left[\frac{N - x_1 + 1}{x_1} \right] P_{x_1-1, x_1-1}, \quad x_1 = 1, 2, \dots, N.$$

Where the appropriate initial condition for both (i) and (ii) is $P_{0,0} = (1-p)^{t(N+R)}$.

(Technically (1.4c) is not an 'initial condition', since the P_{x_1, tx_1+tR} can be generated using (1.4) along with (1.4a) and (1.4b). (1.4c) is included for completeness.)

§ 1.6 : The Maximum Likelihood Estimator

From equation 1.2 it follows that the joint likelihood for N and p is given by

$$L(N, p) \propto \frac{N!}{(N - x_1)!} p^z (1 - p)^{(N+R)t-z}. \quad (1.5)$$

This is maximised over p as follows :

$$\frac{\partial L}{\partial p} \propto -p^z (t(N+R) - z)(1 - p)^{t(N+R)-z-1} + zp^{z-1} (1 - p)^{t(N+R)-z}$$

equate to zero to obtain \hat{p} :

$$\begin{aligned} p^z (t(N+R) - z)(1 - p)^{t(N+R)-z-1} &= zp^{z-1} (1 - p)^{t(N+R)-z} \\ p(t(N+R) - z) &= z(1 - p) \\ \Rightarrow \hat{p} &= \frac{z}{t(N+R)}. \end{aligned}$$

\hat{p} is now substituted into (1.5) to obtain the profile likelihood for N :

$$L(N) \propto \frac{N!}{(N - x_1)!} \left[\frac{z}{t(N+R)} \right]^z \left[1 - \frac{z}{t(N+R)} \right]^{(N+R)t-z}.$$

It is more convenient to consider the log-profile-likelihood from this point. It is easily shown that the log-profile-likelihood may be written as

$$l(N) \propto \ln \left[\frac{N!}{(N - x_1)!} \right] - t(N+R) \ln[t(N+R)] + [t(N+R) - z] \ln[t(N+R) - z]. \quad (1.6)$$

Due to numerical complications, which can occur for larger values of \hat{N} , it was found that the most satisfactory way of calculating the value of the maximum likelihood estimator is as follows :

After observing that the likelihood function is uni-modal it is seen that $\hat{N} = k$, where k is the smallest integer in the set $\{x_1, x_1 + 1, x_1 + 2, \dots\}$ to satisfy the condition

$$\begin{aligned} &L(k) > L(k+1) \\ \Leftrightarrow &l(k) > l(k+1) \\ \Leftrightarrow &\ln \left[\frac{k!}{(k - x_1)!} \right] - t(k+R) \ln[t(k+R)] + [t(k+R) - z] \ln[t(k+R) - z] \\ &> \ln \left[\frac{(k+1)!}{(k+1 - x_1)!} \right] - t(k+1+R) \ln[t(k+1+R)] + [t(k+1+R) - z] \ln[t(k+1+R) - z] \end{aligned}$$

from (1.6)

$$\begin{aligned} \Leftrightarrow &\ln \left[\frac{k!}{(k - x_1)!} \right] - \ln \left[\frac{(k+1)!}{(k+1 - x_1)!} \right] \\ &> t(k+R) \ln[t(k+R)] - [t(k+R) - z] \ln[t(k+R) - z] \\ &\quad - t(k+1+R) \ln[t(k+1+R)] + [t(k+1+R) - z] \ln[t(k+1+R) - z] \end{aligned}$$

$$\Leftrightarrow \ln \left[\frac{(k+1-x_1)}{k+1} \right] + [t(k+R)-z] \ln [t(k+R)-z] - t(k+R) \ln [t(k+R)]$$

$$> [t(k+1+R)-z] \ln [t(k+1+R)-z] - t(k+1+R) \ln [t(k+1+R)].$$

N.B. Once \hat{N} has been determined, this value may then be used in the calculation of the maximum likelihood estimate of p : $\hat{p} = \frac{z}{t(\hat{N}+R)}$.

§ 1.7 : A Peterson-Type Estimator

This section introduces an estimator of population size which is only dependent upon the observed numbers of distinct animals seen from the target and planted populations. The estimator is derived from the conditional distribution of X_1 given X .

From the assumptions stated above one may deduce that

$$X_1 \sim \text{Bin}(N, 1 - (1-p)^t),$$

$$X_2 \sim \text{Bin}(R, 1 - (1-p)^t)$$

and that $X \sim \text{Bin}(N+R, 1 - (1-p)^t)$.

It is then easy to show that the distribution of $X_1|X$ is in fact hypergeometric with probability function

$$\text{Prob}(X_1 = x_1 | X = x) = \frac{\binom{N}{x_1} \binom{R}{x-x_1}}{\binom{N+R}{x}}, \quad \max(0, x-R) \leq x_1 \leq \min(N, x).$$

The likelihood function for N based on this probability function is maximised by the Peterson-type estimator $\tilde{N}_p = RX_1/X_2$. To avoid introducing an estimator which becomes infinite when $X_2=0$, the estimator \tilde{N}_p is now slightly modified. That is from

this point consideration is given to the estimator $\hat{N}_p = \left[0.5 + \frac{(R+1)X_1}{(X_2+1)} \right]$, where $[.]$

denotes the integer part of.

§ 1.8 : A Conditionally Unbiased Estimator

This section introduces the Conditionally Unbiased Estimator \tilde{N}_u , which is an estimator of population size N defined by

$$\tilde{N}_u = \left(\frac{z+1}{t} \right) \frac{\delta(x_1, z+1; t, R)}{\delta(x_1, z; t, R)} - \left(\frac{Rt-z}{t} \right), \quad (1.7)$$

$$\text{where } \delta(x_1, z; t, R) = \sum_{j=0}^{x_1} \binom{x_1}{j} \binom{tj+tR}{z} (-1)^{x_1-j},$$

as defined in section 1.5.

This Conditionally Unbiased Estimator (CUE) was derived from the conditional distribution of X_1 given Z . As previously mentioned in section 1.4, the conditional distribution of X_1 given Z has appeared in the urn model literature. Charalambides(1981) considered a situation which in some respects may be described as a generalisation of the one discussed here - this being the reason why the conditional distribution of X_1 given Z can be obtained from his work. In addition, Charalambides(1981) introduced an estimator which is essentially equivalent to \tilde{N}_u : for values of R greater than or equal to zero. He shows it is a minimum variance unbiased estimator with respect to the conditional distribution of X_1 given Z , provided that $Z \geq N$.

In the absence of plants an estimator very similar to \tilde{N}_u has previously been considered in a capture-recapture context : Pathak(1964) derived an estimator in terms of X_1 and $\underline{n} = \{n_1, n_2, \dots, n_t\}$, where the n_i are the number of animals seen on the i th sampling occasion. Pathak(1964) assumed the n_i to be known constants. A special case of Pathak's estimator is obtained when all the n_i are equal to one : in this situation Berg(1974) showed that Pathak's estimator reduces to a ratio of Stirling numbers of the second kind. This latter result is consistent with the work of Harris(1968). Berg continued his work on Pathak's estimator ; a problem associated with the estimator of Pathak(1964) is that it can be very difficult to compute : being a ratio of two rapidly growing summations. To overcome this problem, in the situation where all the n_i are equal to one, Berg(1975) derived a recurrence relation which enables one to more easily evaluate the estimate produced by Pathak's estimator. In Berg(1976) the result for this latter special case was extended to include the general multiple-capture census. These recurrence relations for Pathak's estimator were given as functions of X_1 and $\underline{n} = \{n_1, n_2, \dots, n_t\}$. The work of Berg(1974, 1975, 1976) provided the motivation for much of the work presented within section 1.8a of this chapter and of section 2.10a in chapter 2.

In order to prove that, provided that the condition $Z \geq N$ holds, \tilde{N}_u is in fact unbiased over the conditional distribution of X_1 given Z , as a first step and to make the

following proof more straightforward, the probability function of X_1 given Z is obtained explicitly :

$$\begin{aligned} \text{Prob}(X_1 = x_1 | Z = z) &= \frac{\text{Prob}(X_1 = x_1, Z = z)}{\text{Prob}(Z = z)} \\ &= \frac{\binom{N}{x_1} p^z (1-p)^{Nt+Rt-z} \delta(x_1, z; t, R)}{\binom{Nt+Rt}{z} p^z (1-p)^{Nt+Rt-z}}. \end{aligned}$$

This follows from equation (1.2) and the fact that $Z \sim \text{Bin}(Nt + Rt, p)$. Hence the probability distribution function of X_1 given Z can be written as

$$\text{Prob}(X_1 = x_1 | Z = z) = \frac{\binom{N}{x_1} \delta(x_1, z; t, R)}{\binom{Nt+Rt}{z}}. \quad (1.8)$$

This is essentially identical to the probability function (2.8), on page 604 of Charalambides(1981).

Now the expectation of \tilde{N}_u taken over the conditional distribution of X_1 given Z is given by

$$\begin{aligned} E(\tilde{N}_u) &= \sum_{x_1} \tilde{N}_u \text{Prob}(X_1 = x_1 | Z = z) \\ &= \sum_{x_1} \left(\left(\frac{z+1}{t} \right) \frac{\delta(x_1, z+1; t, R)}{\delta(x_1, z; t, R)} - \left(\frac{Rt-z}{t} \right) \right) \text{Prob}(X_1 = x_1 | Z = z) \\ &= \left(\sum_{x_1} \left(\frac{z+1}{t} \right) \frac{\delta(x_1, z+1; t, R)}{\delta(x_1, z; t, R)} \text{Prob}(X_1 | Z) \right) - \left(\frac{Rt-z}{t} \right) \\ &= \left(\sum_{x_1} \left(\frac{z+1}{t} \right) \frac{\delta(x_1, z+1; t, R)}{\delta(x_1, z; t, R)} \frac{\binom{N}{x_1} \delta(x_1, z; t, R)}{\binom{Nt+Rt}{z}} \right) - \left(\frac{Rt-z}{t} \right) \quad \text{using (1.8)} \\ &= \left(\left(\frac{z+1}{t} \right) \sum_{x_1} \frac{\binom{N}{x_1} \delta(x_1, z+1; t, R)}{\binom{Nt+Rt}{z}} \right) - \left(\frac{Rt-z}{t} \right) \\ &= \left(\left(\frac{z+1}{t} \right) \frac{\binom{Nt+Rt}{z+1}}{\binom{Nt+Rt}{z}} \sum_{x_1} \frac{\binom{N}{x_1} \delta(x_1, z+1; t, R)}{\binom{Nt+Rt}{z+1}} \right) - \left(\frac{Rt-z}{t} \right) \\ &= \left(\left(\frac{z+1}{t} \right) \left(\frac{Nt+Rt-z}{z+1} \right) \cdot 1 \right) - \left(\frac{Rt-z}{t} \right) \quad \text{if } Z \geq N \\ &= N. \end{aligned}$$

This shows that \tilde{N}_u is unbiased over the conditional distribution of X_1 given Z , provided that the condition $Z \geq N$ holds. Furthermore, again provided that $Z \geq N$, since X_1 and Z are sufficient, it follows that \tilde{N}_u is the minimum variance unbiased estimator, Rao(1952).

In view of the fact that population size N is integer valued, in later sections consideration is given to the following slightly modified version of \tilde{N}_u :

$$\hat{N}_u = \left[0.5 + \left(\frac{z+1}{t} \right) \frac{\delta(x_1, z+1; t, R)}{\delta(x_1, z; t, R)} - \left(\frac{Rt-z}{t} \right) \right],$$

where the square brackets have been used to denote the integer part.

§ 1.8a : A Note on the Evaluation of the CUE

Direct use of equation (1.7) to evaluate the estimates produced by the estimator \tilde{N}_u can often be difficult, and involve very cumbersome computation. This is due to the fact that the δ -numbers, present within (1.7), grow rapidly with increasing arguments. To overcome this computational problem, a recurrence relation linking the \tilde{N}_u is stated and proved. To make the following proof more easily read some shorthand notation is necessary.

Let
$$N_{x_1, z} = \tilde{N}_u = \left(\frac{z+1}{t} \right) \frac{\delta(x_1, z+1; t, R)}{\delta(x_1, z; t, R)} - \left(\frac{Rt-z}{t} \right)$$

and let
$$\delta_{x_1, z} = \delta(x_1, z; t, R).$$

The $N_{x_1, z}$ are then subject to the following recurrence relation

$$N_{x_1, z} = x_1 + \left(\frac{tN_{x_1-1, z-1} + Rt - z + 1}{tN_{x_1, z-1} + Rt - z + 1} \right) (N_{x_1, z-1} - x_1), \quad (1.9)$$

with initial conditions
$$N_{0, z} = 0 \quad \text{for } z = 0, 1, 2, \dots, tR, \quad (1.10)$$

and
$$N_{x_1, x_1} = \frac{x_1}{2t} [2Rt + tx_1 - x_1 + t + 1] \quad \text{for } x_1 \geq 0. \quad (1.11)$$

Proof of (1.9) :

$$\begin{aligned}
& x_1 + \left(\frac{tN_{x_1-1,z-1} + Rt - z + 1}{tN_{x_1,z-1} + Rt - z + 1} \right) (N_{x_1,z-1} - x_1) \\
&= x_1 + \left(\frac{z \frac{\delta_{x_1-1,z}}{\delta_{x_1-1,z-1}} - Rt + z - 1 + Rt - z + 1}{z \frac{\delta_{x_1,z}}{\delta_{x_1,z-1}} - Rt + z - 1 + Rt - z + 1} \right) \left(\frac{z}{t} \frac{\delta_{x_1,z}}{\delta_{x_1,z-1}} - \frac{(Rt - z + 1)}{t} - x_1 \right) \\
&= x_1 + \frac{\delta_{x_1-1,z}}{\delta_{x_1-1,z-1}} \frac{\delta_{x_1,z-1}}{\delta_{x_1,z}} \frac{(z\delta_{x_1,z} - (tx_1 + Rt - z + 1)\delta_{x_1,z-1})}{t\delta_{x_1,z-1}} \\
&= x_1 + \frac{\delta_{x_1-1,z}}{t\delta_{x_1,z}} \frac{(z\delta_{x_1,z} - (tx_1 + Rt - z + 1)\delta_{x_1,z-1})}{\delta_{x_1-1,z-1}} \\
&= x_1 + \frac{\delta_{x_1-1,z}}{t\delta_{x_1,z}} \frac{(tx_1\delta_{x_1-1,z-1})}{\delta_{x_1-1,z-1}} \quad \text{using (1.3)} \\
&= x_1 + \frac{tx_1\delta_{x_1-1,z}}{t\delta_{x_1,z}} \\
&= x_1 + \frac{((z+1)\delta_{x_1,z+1} - (tx_1 + Rt - z)\delta_{x_1,z})}{t\delta_{x_1,z}} \quad \text{using (1.3) with } z \text{ replaced by } z+1 \\
&= x_1 + \frac{(z+1)\delta_{x_1,z+1}}{t\delta_{x_1,z}} - \frac{(tx_1 + Rt - z)}{t} \\
&= x_1 + \frac{(z+1)\delta_{x_1,z+1}}{t\delta_{x_1,z}} - \frac{(Rt - z)}{t} - x_1 \\
&= \left(\frac{z+1}{t} \right) \frac{\delta_{x_1,z+1}}{\delta_{x_1,z}} - \left(\frac{Rt - z}{t} \right) \\
&= N_{x_1,z}.
\end{aligned}$$

Proof of (1.10) :

$$\begin{aligned}
N_{0,z} &= \left(\frac{z+1}{t} \right) \frac{\delta_{0,z+1}}{\delta_{0,z}} - \left(\frac{Rt - z}{t} \right) \\
&= \left(\frac{z+1}{t} \right) \frac{\binom{Rt}{z+1}}{\binom{Rt}{z}} - \left(\frac{Rt - z}{t} \right) \\
&= \left(\frac{z+1}{t} \right) \frac{\binom{Rt-z}{z+1} \binom{Rt}{z}}{\binom{Rt}{z}} - \left(\frac{Rt - z}{t} \right) \\
&= 0.
\end{aligned}$$

Proof of (1.11) :

As a first step in this proof it is necessary to prove the identity

$$\delta_{x_1, x_1+1} = \frac{t^{x_1}}{2} [2Rt + tx_1 - x_1]. \quad (1.12)$$

The identity (1.12) may be proved by induction :

Anchor : (1.12) is clearly true for $x_1 = 0$, since $\delta_{0,1} = Rt$.

Assume true for $x_1 = k$, i.e. assume $\delta_{k, k+1} = \frac{t^k}{2} [2Rt + tk - k]$.

Then

$$\begin{aligned} \delta_{k+1, k+2} &= \frac{1}{k+2} \left[(t(k+1) + Rt - (k+2) + 1) \delta_{k+1, k+1} + t(k+1) \delta_{k, k+1} \right] && \text{using (1.3)} \\ &= \frac{1}{k+2} \left[(tk + t + Rt - k - 1) t^{k+1} + t(k+1) \frac{t^k}{2} [2Rt + tk - k] \right] && \text{using assumption} \\ &= \frac{t^{k+1}}{2(k+2)} \left[2(tk + t + Rt - k - 1) + (k+1)(2Rt + tk - k) \right] \\ &= \frac{t^{k+1}}{2(k+2)} \left[3tk + 2t - 3k - 2 + 4Rt + 2kRt + tk^2 - k^2 \right] \\ &= \frac{t^{k+1}}{2(k+2)} \left[(k+2)(2Rt + t(k+1) - (k+1)) \right] \\ &= \frac{t^{k+1}}{2} [2Rt + t(k+1) - (k+1)]. \end{aligned}$$

This shows that, if (1.12) is true for $x_1 = k$, then it must also be true for $x_1 = k+1$. Since it has been shown that (1.12) is true for $x_1 = 0$, it follows by induction that (1.12) holds for all $x_1 \geq 0$.

The proof of (1.11) may now be completed :

$$\begin{aligned} N_{x_1, x_1} &= \left(\frac{x_1 + 1}{t} \right) \frac{\delta_{x_1, x_1+1}}{\delta_{x_1, x_1}} - \left(\frac{Rt - x_1}{t} \right) \\ &= \left(\frac{x_1 + 1}{t} \right) \frac{\frac{t^{x_1}}{2} [2Rt + tx_1 - x_1]}{t^{x_1}} - \left(\frac{Rt - x_1}{t} \right) \\ &= \frac{1}{2t} \left[(x_1 + 1)(2Rt + tx_1 - x_1) - 2(Rt - x_1) \right] \\ &= \frac{1}{2t} \left[2x_1Rt + tx_1^2 - x_1^2 + tx_1 + x_1 \right] \\ &= \frac{x_1}{2t} [2Rt + tx_1 - x_1 + t + 1]. \end{aligned}$$

§ 1.9 : A Comparison of All Three Estimators

In order to compare the performance of the three estimators which have so far been discussed we consider their mean, standard deviation and root mean square error conditional on the event $C = \{Z > X_1\}$. This conditioning is necessary since the maximum likelihood estimator \hat{N} yields infinite estimates when $Z = X_1$. It is important to note however that both the Peterson-type estimator \hat{N}_p and the CUE \hat{N}_u produce finite estimates with probability one. The unconditional performance of \hat{N}_p and \hat{N}_u is considered later on in section 1.10.

Conditional on the event $C = \{Z > X_1\}$, the mean, standard deviation and root mean square error of each estimator are presented in tables 1.1a,b,c, 1.2a,b,c, 1.3a,b,c and 1.4a,b,c. These tables summarise the performance of the estimators for each combination from the following factorial design :

			0
	10	5	5
			0.05
$N =$	25	10	10
\times	50	15	25
			0.20
	100	20	50
			100

Note however that, for each value of population size N , only values of R up to and including N are considered; this is done for obvious practical reasons.

The notation used within each table is as follows :

Statistics

exp.	\equiv	mean or expectation.
s.d.	\equiv	standard deviation.
rmse	\equiv	root mean square error.
$P(\text{inf mle})$	\equiv	$1 - \text{Prob}(C) = \text{Prob}(\bar{C}) = \text{Prob}(Z = X_1)$, which is the probability of the maximum likelihood estimator producing an infinite estimate.

Estimators

X_1	\equiv	X_1 , the number of distinct individuals seen from the target population.
P	\equiv	\hat{N}_p , the Peterson-type estimator of section 1.7.
CUE	\equiv	\hat{N}_u , the conditionally unbiased estimator of section 1.8.
MLE	\equiv	\hat{N} , the maximum likelihood estimator of section 1.6.

It is straightforward to obtain the distributions of both \hat{N} and \hat{N}_u given C. In order to obtain the conditional distribution of the Peterson-type estimator \hat{N}_p given C we need to derive the conditional distribution of X_1 and X_2 given C. This may be done as follows :

C is defined as being the event $\{Z > X_1\}$.

Let \bar{C} be the complementary event $\{Z = X_1\}$.

\bar{C} occurs $\Leftrightarrow X_2 = 0$ and each animal in target population is seen at most once.

$$\text{Now } \text{Prob}(X_2 = 0) = \left[(1-p)^t \right]^R. \quad (1.13)$$

(This follows from the fact that $X_2 \sim \text{Bin}(R, 1 - (1-p)^t)$.)

Let Y_i = the number of sightings of animal i, it follows that $Y_i \sim \text{Bin}(t, p)$.

It may then be observed that

Prob(each animal in target population is seen at most once)

$$\begin{aligned} &= \prod_{i=1}^N \text{Prob}(Y_i \leq 1) \\ &= \left[(1-p + tp)(1-p)^{t-1} \right]^N. \end{aligned} \quad (1.14)$$

Use of (1.13) and (1.14) implies that

$$\begin{aligned} \text{Prob}(C) &= 1 - \text{Prob}(\bar{C}) \\ &= 1 - \left[(1-p)^t \right]^R \left[(1-p + tp)(1-p)^{t-1} \right]^N. \end{aligned}$$

Now

$$\begin{aligned} \text{Prob}(X_1 = x_1, X_2 = x_2 | Z > X_1) &= \frac{\text{Prob}(X_1 = x_1, X_2 = x_2, Z > X_1)}{\text{Prob}(Z > X_1)} \\ &= \frac{\text{Prob}(X_1 = x_1, X_2 = x_2) \text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2)}{\text{Prob}(Z > X_1)} \\ &= \frac{\text{Prob}(X_1 = x_1) \text{Prob}(X_2 = x_2) \text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2)}{\text{Prob}(Z > X_1)}. \end{aligned}$$

It is clear that $\text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2) = 1$ if $X_2 > 0$.

When $X_2 = 0$ it may be observed that $Z | X_1, X_2 \equiv Z_1 | X_1$. It is known that the distribution of $Z_1 | X_1$ may be characterised as being the sum of X_1 zero truncated Binomial random variables, the distribution of which is derived in appendix 3.

Explicitly the probability function of $Z_1 | X_1$ is given by

$$\text{Prob}(Z_1 = z_1 | X_1 = x_1) = \frac{p^{z_1} (1-p)^{x_1 - z_1}}{\left[1 - (1-p)^t \right]^{x_1}} \delta(x_1, z_1; t, 0).$$

It follows that

$$\begin{aligned}
 \text{Prob}(Z > x_1 | X_1 = x_1, X_2 = x_2) &= \text{Prob}(Z_1 > x_1 | X_1 = x_1) \\
 &= 1 - \text{Prob}(Z_1 = x_1 | X_1 = x_1) \\
 &= 1 - \frac{p^{x_1} (1-p)^{t^{x_1-x_1}}}{[1-(1-p)^t]^{x_1}} \delta(x_1, x_1; t, 0) \\
 &= 1 - \frac{t^{x_1} p^{x_1} (1-p)^{t^{x_1-x_1}}}{[1-(1-p)^t]^{x_1}}, \quad \text{using 1.3a.}
 \end{aligned}$$

Using the notation $\tilde{P}(C) = \text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2)$ then allows one to write :

$$\text{Prob}(X_1 = x_1, X_2 = x_2 | Z > X_1) = \frac{\binom{N}{x_1} [1-(1-p)^t]^{x_1} [(1-p)^t]^{N-x_1} \binom{R}{x_2} [1-(1-p)^t]^{x_2} [(1-p)^t]^{R-x_2} \tilde{P}(C)}{P(C)},$$

$$x_1 = \begin{cases} 0, 1, 2, \dots, N & \text{for } R > 0 \\ 1, 2, \dots, N & \text{for } R = 0 \end{cases}$$

$$x_2 = \begin{cases} 1, 2, \dots, R & \text{for } R > 0, x_1 = 0 \\ 0, 1, 2, \dots, R & \text{for } R > 0, x_1 > 0, \\ 0 & \text{for } R = 0 \end{cases}$$

$$\text{where } \tilde{P}(C) = \begin{cases} 1 & \text{for } X_2 > 0 \\ 1 - \frac{t^{x_1} p^{x_1} (1-p)^{t^{x_1-x_1}}}{[1-(1-p)^t]^{x_1}} & \text{for } X_2 = 0 \end{cases}$$

$$\text{and } P(C) = \text{Prob}(Z > X_1) = 1 - [(1-p)^t]^{R+1} [(1-p+t p)(1-p)^{t-1}]^N.$$

§ 1.9a : Discussion

Let us firstly compare the performance of the Peterson-type estimator \hat{N}_p to that of the CUE \hat{N}_u ; the comparison between these two estimators is straightforward in situations with or without plants. In the absence of plants , i.e. when $R = 0$, since \hat{N}_p reduces to X_1 , the number of distinct individuals seen from the target population, one would expect \hat{N}_u to clearly outperform \hat{N}_p . This is broadly true, in that for the great majority of situations considered, when $R = 0$, the CUE generally possesses both a better mean and root mean square error. In the remaining three situations where the root mean square error of \hat{N}_u is marginally greater than that of \hat{N}_p , the CUE is less biased. When sampling with plants, i.e. when $R > 0$, the CUE is again seen to be clearly a better alternative to \hat{N}_p . When $R > 0$ both estimators have very small bias. However the estimator \hat{N}_u is in almost all situations less biased than \hat{N}_p , and where its bias is worse

the difference is minimal. The standard deviation of \hat{N}_u is always less than that of \hat{N}_p , the difference between these two statistics being appreciable when the number of plants is small relative to population size. In terms of root mean square error, when $R > 0$, \hat{N}_u is uniformly better than \hat{N}_p ; the root mean square error of \hat{N}_u is more appreciably better than that of \hat{N}_p when R is small relative to N .

In situations where only a very small amount of information is available the maximum likelihood estimator \hat{N} has a tendency to be positively biased, sometimes extremely so. This is most noticeable when considering the larger population sizes. In contrast to this the CUE tends to be negatively biased when only a very small amount of information is present. As more information becomes available both \hat{N} and \hat{N}_u each perform extremely well in terms of mean, with \hat{N}_u on all but a few occasions being the less biased of the two. With regard to bias, it appears that, of the two estimators \hat{N} and \hat{N}_u , the CUE behaves in a far more desirable way. To illustrate this consider table 1.4a, in which $N = 100$ and $p = 0.05$. Consider the situation where R is equal to zero: for $t = 5, 10, 15$ and 20 the mean values taken by \hat{N} are respectively 115, 108, 102 and 101; whereas the corresponding mean values taken by \hat{N}_u are respectively 80.6, 99.9, 100 and 100. This shows how the bias of the estimators can alter as more information is gained through additional sampling occasions. To show that the estimators respond in a similar way as information is gained through the planted individuals consider the column giving results for $t = 5$: for $R = 0, 5, 10, 25$ and 50 the mean values taken by \hat{N} are respectively 115, 123, 122, 111 and 105; whereas the corresponding mean values taken by \hat{N}_u are respectively 80.6, 92.0, 97.1, 99.9 and 100. These examples highlight in particular the general feature that the mean of the CUE improves uniformly with more information whereas that of the MLE is less predictable.

In all but a few situations, the CUE exhibits a smaller standard deviation than the MLE. In particular, when the number of sampling occasions is small the standard deviation of \hat{N}_u tends to be significantly smaller than that of \hat{N} . The MLE only has a smaller standard deviation than that of \hat{N}_u in a few situations, wherein $p = 0.20$, and notably in these situations one would expect to see on average at least 96% of the target population.

Since, on the whole, the CUE tends to possess both a smaller absolute bias and standard deviation, it necessarily follows that \hat{N}_u usually also has the smaller root mean square error. When only a small amount of information is available \hat{N}_u , in terms of root mean square error, is seen to significantly outperform \hat{N} . Whereas, as more information is gained the two estimators are seen to behave more closely in terms of root mean square error, although again with \hat{N}_u tending to be ahead.

In the above discussion a deliberate attempt has been made not to place too much emphasis on mean square error. This being due to the fact that mean square error is known to reward negative bias. So that when only a small proportion of the population is seen during sampling, that is when \hat{N}_u is negatively biased with 'small' variance and \hat{N} is positively biased with 'large' variance, one would expect mean square error to perhaps unfairly favour the estimator \hat{N}_u . It is true that this can occasionally happen : for an example consider table 1.2a. When $N = 25$, $p = 0.05$, $R = 5$ and $t = 5$ the mean, standard deviation and root mean square error of \hat{N}_u are respectively 18.1, 9.47 and 11.7; whereas the corresponding values for \hat{N} are respectively 26.2, 17.4 and 17.4. In this situation, on the basis of root mean square error alone, one would choose the CUE, however it could be argued that an alternative loss criterion which places more weight on the mean of an estimator might more sensibly favour the MLE. Examples of this type however are few and far between. Generally the mean of \hat{N}_u is as good as or better than that of \hat{N} , and as a result of \hat{N}_u also tending to have a smaller standard deviation, it may be concluded that one should always use the CUE in preference to the MLE.

Table 1.1a

N = 10		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	2.96	1.24	7.15	4.40	1.46	5.79	5.54	1.51	4.71	6.48	1.49	3.82
	P	2.96	1.24	7.15	4.40	1.46	5.79	5.54	1.51	4.71	6.48	1.49	3.82
	CUE	3.93	2.32	6.50	6.63	3.31	4.72	8.46	3.51	3.83	9.47	3.32	3.36
	MLE	4.30	3.02	6.45	7.92	5.03	5.44	9.79	5.44	5.44	10.4	4.95	4.97
	P(inf mle)	0.7957			0.4063			0.1534			0.0465		
5	X1	2.31	1.33	7.80	4.03	1.55	6.17	5.37	1.58	4.89	6.42	1.52	3.89
	P	6.56	5.11	6.16	9.24	5.90	5.95	9.93	5.36	5.36	10.0	4.50	4.50
	CUE	6.60	4.54	5.67	9.27	5.00	5.05	9.87	4.18	4.18	10.0	3.25	3.25
	MLE	9.30	7.64	7.67	11.3	8.12	8.22	10.6	6.08	6.11	10.0	4.10	4.10
	P(inf mle)	0.2207			0.0313			0.0033			0.0003		
10	X1	2.27	1.32	7.84	4.01	1.55	6.18	5.37	1.58	4.89	6.42	1.52	3.89
	P	8.63	6.74	6.88	9.95	5.96	5.96	10.0	4.53	4.53	10.0	3.56	3.56
	CUE	8.52	6.14	6.31	9.87	5.26	5.26	9.96	3.84	3.84	9.99	2.93	2.93
	MLE	11.6	10.5	10.7	11.0	7.57	7.64	10.1	4.56	4.56	9.77	3.16	3.17
	P(inf mle)	0.0612			0.0024			0.0001			0.0000		

Table 1.1b

N = 10		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	4.49	1.47	5.71	6.57	1.48	3.73	7.95	1.28	2.42	8.78	1.03	1.60
	P	4.49	1.47	5.71	6.57	1.48	3.73	7.95	1.28	2.42	8.78	1.03	1.60
	CUE	6.62	3.10	4.59	9.38	3.21	3.27	9.97	2.31	2.31	10.0	1.57	1.57
	MLE	7.60	4.43	5.04	10.2	4.93	4.93	9.94	3.09	3.09	9.61	1.72	1.76
	P(inf mle)	0.4275			0.0467			0.0025			0.0001		
5	X1	4.11	1.55	6.09	6.51	1.51	3.80	7.94	1.28	2.42	8.78	1.03	1.60
	P	9.27	5.82	5.87	10.0	4.40	4.40	10.1	3.02	3.02	10.1	2.16	2.16
	CUE	9.29	4.82	4.87	9.99	3.20	3.20	9.99	2.02	2.02	10.0	1.34	1.34
	MLE	11.3	7.99	8.09	10.1	3.98	3.98	9.64	2.12	2.15	9.55	1.43	1.50
	P(inf mle)	0.0307			0.0002			0.0000			0.0000		
10	X1	4.10	1.55	6.11	6.51	1.51	3.80	7.94	1.28	2.42	8.78	1.03	1.60
	P	9.96	5.85	5.85	10.1	3.47	3.47	10.1	2.30	2.30	10.1	1.63	1.63
	CUE	9.90	5.13	5.13	10.0	2.88	2.88	10.0	1.88	1.88	10.0	1.29	1.29
	MLE	10.9	7.43	7.47	9.81	3.10	3.11	9.59	1.92	1.96	9.51	1.36	1.44
	P(inf mle)	0.0022			0.0000			0.0000			0.0000		

Table 1.1c

N = 10		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	6.78	1.46	3.53	8.93	0.98	1.45	9.65	0.58	0.68	9.88	0.34	0.36
	P	6.78	1.46	3.53	8.93	0.98	1.45	9.65	0.58	0.68	9.88	0.34	0.36
	CUE	9.58	3.08	3.11	9.98	1.44	1.44	9.86	0.75	0.76	9.90	0.35	0.37
	MLE	10.2	4.50	4.51	9.59	1.56	1.61	9.69	0.63	0.70	9.89	0.34	0.36
	P(inf mle)	0.0475			0.0001			0.0000			0.0000		
5	X1	6.72	1.48	3.60	8.93	0.98	1.45	9.65	0.58	0.68	9.88	0.34	0.36
	P	10.1	4.20	4.20	10.1	2.00	2.00	10.0	1.07	1.07	10.0	0.60	0.60
	CUE	10.1	3.06	3.06	10.0	1.29	1.29	9.85	0.73	0.75	9.89	0.34	0.36
	MLE	10.0	3.76	3.76	9.52	1.33	1.42	9.66	0.60	0.69	9.88	0.34	0.36
	P(inf mle)	0.0002			0.0000			0.0000			0.0000		
10	X1	6.72	1.48	3.60	8.93	0.98	1.45	9.65	0.58	0.68	9.88	0.34	0.36
	P	10.1	3.30	3.30	10.0	1.52	1.52	10.0	0.84	0.84	10.0	0.48	0.48
	CUE	10.0	2.74	2.74	10.0	1.21	1.21	9.80	0.70	0.73	9.89	0.34	0.36
	MLE	9.75	2.95	2.96	9.50	1.27	1.37	9.66	0.59	0.68	9.88	0.34	0.36
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.2a

N = 25		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	6.24	2.03	18.9	10.2	2.41	15.0	13.4	2.48	11.8	16.0	2.40	9.28
	P	6.24	2.03	18.9	10.2	2.41	15.0	13.4	2.48	11.8	16.0	2.40	9.28
	CUE	11.5	5.91	14.7	21.2	8.89	9.65	24.5	8.35	8.36	25.0	6.39	6.39
	MLE	15.1	9.53	13.8	27.1	15.5	15.6	27.5	13.4	13.6	25.9	8.44	8.49
	P(inf mle)	0.5648			0.1052			0.0092			0.0005		
5	X1	5.74	2.09	19.4	10.0	2.45	15.2	13.4	2.49	11.8	16.0	2.40	9.28
	P	17.5	11.0	13.3	23.7	13.5	13.6	24.8	12.2	12.2	25.0	10.1	10.1
	CUE	18.1	9.47	11.7	24.4	10.4	10.4	25.0	7.68	7.68	25.0	5.52	5.52
	MLE	26.2	17.4	17.4	28.7	17.2	17.6	26.2	9.99	10.1	25.2	6.14	6.15
	P(inf mle)	0.1567			0.0081			0.0002			0.0000		
10	X1	5.68	2.09	19.4	10.0	2.45	15.2	13.4	2.49	11.8	16.0	2.40	9.28
	P	22.1	14.3	14.6	24.9	12.9	12.9	25.0	9.62	9.62	25.0	7.44	7.44
	CUE	22.1	12.1	12.5	24.9	9.96	9.96	25.0	6.85	6.85	25.0	5.09	5.09
	MLE	30.1	22.3	22.9	27.4	14.1	14.3	25.6	7.80	7.82	25.0	5.36	5.36
	P(inf mle)	0.0435			0.0006			0.0000			0.0000		
25	X1	5.66	2.09	19.5	10.0	2.45	15.2	13.4	2.49	11.8	16.0	2.40	9.28
	P	24.9	14.5	14.5	25.0	9.13	9.13	25.0	6.77	6.77	25.0	5.39	5.39
	CUE	24.9	13.4	13.4	25.0	8.19	8.19	25.0	5.85	5.85	25.0	4.51	4.51
	MLE	28.5	19.6	19.9	25.6	8.98	9.00	25.0	6.08	6.08	24.7	4.63	4.64
	P(inf mle)	0.0009			0.0000			0.0000			0.0000		

Table 1.2b

N = 25		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	10.4	2.41	14.8	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
	P	10.4	2.41	14.8	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
	CUE	20.9	8.44	9.39	25.0	6.25	6.25	25.0	3.56	3.56	25.0	2.29	2.29
	MLE	26.5	14.7	14.8	25.9	8.29	8.34	24.8	3.78	3.79	24.6	2.35	2.38
	P(inf mle)	0.1195			0.0005			0.0000			0.0000		
5	X1	10.2	2.46	15.0	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
	P	23.7	13.4	13.5	25.0	9.92	9.92	25.0	6.70	6.70	25.0	4.78	4.78
	CUE	24.3	10.3	10.3	25.0	5.41	5.41	25.0	3.26	3.26	25.0	2.21	2.21
	MLE	28.6	17.2	17.6	25.2	6.05	6.05	24.7	3.36	3.37	24.6	2.23	2.27
	P(inf mle)	0.0086			0.0000			0.0000			0.0000		
10	X1	10.2	2.46	15.0	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
	P	24.9	12.7	12.7	25.0	7.26	7.26	25.0	4.90	4.90	24.9	3.54	3.54
	CUE	24.9	9.94	9.94	25.0	4.96	4.96	25.0	3.14	3.14	25.0	2.14	2.14
	MLE	27.2	14.1	14.3	25.0	5.24	5.24	24.6	3.18	3.20	24.5	2.16	2.20
	P(inf mle)	0.0006			0.0000			0.0000			0.0000		
25	X1	10.2	2.46	15.0	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
	P	25.0	8.95	8.95	25.0	5.27	5.27	25.0	3.63	3.63	25.1	2.57	2.57
	CUE	25.0	8.06	8.06	25.0	4.41	4.41	25.0	2.91	2.91	25.0	2.05	2.05
	MLE	25.6	8.83	8.85	24.8	4.53	4.53	24.6	2.93	2.96	24.5	2.06	2.11
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.2c

N = 25		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
	P	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
	CUE	24.9	6.12	6.12	25.0	2.14	2.14	25.1	1.07	1.07	24.8	0.61	0.64
	MLE	25.7	8.15	8.18	24.6	2.18	2.22	24.4	1.08	1.21	24.7	0.54	0.61
	P(inf mle)	0.0005			0.0000			0.0000			0.0000		
5	X1	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
	P	25.1	9.45	9.45	25.0	4.44	4.44	25.0	2.40	2.40	25.0	1.34	1.34
	CUE	25.1	5.26	5.26	25.0	2.05	2.05	25.1	1.03	1.04	24.8	0.60	0.64
	MLE	25.2	5.80	5.81	24.5	2.07	2.12	24.4	1.08	1.21	24.7	0.53	0.61
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
10	X1	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
	P	25.0	6.88	6.88	24.9	3.29	3.29	25.0	1.86	1.86	25.0	1.11	1.11
	CUE	25.0	4.81	4.81	25.0	1.99	1.99	25.1	1.02	1.02	24.8	0.58	0.63
	MLE	24.9	5.04	5.04	24.5	2.01	2.07	24.4	1.07	1.22	24.7	0.53	0.61
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
25	X1	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
	P	25.0	5.02	5.02	25.1	2.38	2.38	25.0	1.32	1.32	25.0	0.76	0.76
	CUE	25.0	4.26	4.26	25.0	1.91	1.91	25.1	0.98	0.98	24.7	0.56	0.62
	MLE	24.7	4.33	4.34	24.5	1.92	1.98	24.4	1.06	1.22	24.7	0.53	0.61
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.3a

N = 50		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	11.7	2.90	38.4	20.1	3.45	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	P	11.7	2.90	38.4	20.1	3.45	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	CUE	30.1	13.4	24.0	48.6	18.3	18.3	49.9	12.7	12.7	50.0	8.64	8.64
	MLE	43.3	24.4	25.4	57.7	31.7	32.7	52.3	16.2	16.3	50.6	9.37	9.39
	P(inf mle)	0.3190			0.0111			0.0001			0.0000		
5	X1	11.4	2.95	38.7	20.1	3.47	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	P	37.0	20.8	24.5	47.7	25.9	26.0	49.6	23.4	23.4	50.0	19.5	19.5
	CUE	40.8	18.7	20.9	49.8	17.7	17.7	50.0	11.3	11.3	50.0	8.01	8.01
	MLE	57.9	36.3	37.1	55.2	26.3	26.8	51.4	12.8	12.9	50.3	8.43	8.44
	P(inf mle)	0.0885			0.0009			0.0000			0.0000		
10	X1	11.3	2.96	38.8	20.1	3.47	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	P	45.4	27.5	27.9	49.9	24.3	24.3	50.0	17.9	17.9	50.0	13.9	13.9
	CUE	46.3	22.3	22.6	50.0	16.0	16.0	50.0	10.4	10.4	50.0	7.58	7.58
	MLE	61.5	41.8	43.4	53.4	20.6	20.9	50.9	11.3	11.3	50.2	7.88	7.88
	P(inf mle)	0.0245			0.0001			0.0000			0.0000		
25	X1	11.3	2.96	38.8	20.1	3.47	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	P	49.9	25.7	25.7	50.0	15.9	15.9	50.0	11.8	11.8	50.0	9.36	9.36
	CUE	49.8	22.5	22.5	50.0	13.0	13.0	50.0	9.07	9.07	50.0	6.87	6.87
	MLE	56.8	33.1	33.7	51.4	14.2	14.2	50.3	9.41	9.41	49.9	7.00	7.00
	P(inf mle)	0.0005			0.0000			0.0000			0.0000		
50	X1	11.3	2.96	38.8	20.1	3.47	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	P	50.0	19.7	19.7	50.0	12.5	12.5	50.0	9.43	9.43	50.0	7.55	7.55
	CUE	50.0	18.6	18.6	50.0	11.3	11.3	50.0	8.18	8.18	50.0	6.33	6.33
	MLE	52.7	21.2	21.3	50.5	11.7	11.7	50.0	8.32	8.32	49.8	6.40	6.40
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.3b

N = 50		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	20.5	3.46	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	P	20.5	3.46	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	CUE	48.2	18.1	18.2	50.0	8.54	8.54	50.0	4.86	4.86	50.0	3.21	3.21
	MLE	57.7	31.8	32.7	50.6	9.29	9.31	49.8	4.97	4.97	49.6	3.24	3.27
	P(inf mle)	0.0143			0.0000			0.0000			0.0000		
5	X1	20.5	3.48	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	P	47.9	25.8	25.9	50.0	19.1	19.1	50.0	12.8	12.8	50.0	9.09	9.09
	CUE	49.8	17.7	17.7	50.0	7.89	7.89	50.0	4.71	4.71	50.0	3.14	3.14
	MLE	55.4	26.8	27.3	50.2	8.30	8.31	49.7	4.76	4.77	49.6	3.15	3.18
	P(inf mle)	0.0010			0.0000			0.0000			0.0000		
10	X1	20.5	3.48	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	P	49.9	23.9	23.9	50.0	13.5	13.5	50.0	9.07	9.07	50.0	6.58	6.58
	CUE	50.0	16.0	16.0	50.0	7.48	7.48	50.0	4.55	4.55	50.0	3.08	3.08
	MLE	53.4	20.9	21.1	50.1	7.78	7.78	49.7	4.61	4.62	49.6	3.10	3.14
	P(inf mle)	0.0001			0.0000			0.0000			0.0000		
25	X1	20.5	3.48	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	P	50.0	15.6	15.6	50.0	9.15	9.15	50.0	6.30	6.30	50.1	4.54	4.54
	CUE	50.0	12.9	12.9	50.0	6.75	6.75	50.0	4.31	4.31	50.0	2.97	2.97
	MLE	51.3	14.1	14.2	49.9	6.86	6.86	49.6	4.33	4.35	49.5	2.97	3.01
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
50	X1	20.5	3.48	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	P	50.0	12.3	12.3	50.0	7.39	7.39	50.0	5.13	5.13	50.1	3.68	3.68
	CUE	50.0	11.2	11.2	50.0	6.20	6.20	50.0	4.09	4.09	50.0	2.88	2.88
	MLE	50.5	11.6	11.6	49.7	6.25	6.26	49.6	4.10	4.12	49.5	2.89	2.92
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.3c

N = 50		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	P	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	CUE	49.9	8.30	8.30	50.0	2.98	2.98	50.0	1.50	1.50	50.0	0.92	0.92
	MLE	50.5	9.05	9.07	49.6	3.00	3.03	49.5	1.48	1.55	49.5	0.78	0.95
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
5	X1	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	P	50.0	18.2	18.2	50.0	8.43	8.43	50.0	4.56	4.56	50.0	2.56	2.56
	CUE	50.0	7.66	7.66	50.0	2.91	2.91	50.0	1.49	1.49	50.0	0.92	0.92
	MLE	50.2	8.07	8.07	49.6	2.92	2.95	49.5	1.46	1.54	49.5	0.78	0.95
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
10	X1	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	P	50.0	12.8	12.8	50.0	6.12	6.12	50.0	3.35	3.35	50.0	1.88	1.88
	CUE	49.9	7.20	7.20	50.0	2.86	2.86	50.0	1.49	1.49	50.1	0.91	0.91
	MLE	50.0	7.50	7.50	49.5	2.88	2.91	49.5	1.46	1.54	49.4	0.77	0.95
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
25	X1	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	P	50.0	8.71	8.71	50.1	4.21	4.21	50.0	2.31	2.31	50.0	1.32	1.32
	CUE	50.0	6.46	6.46	50.0	2.76	2.76	50.0	1.47	1.47	50.1	0.91	0.92
	MLE	49.8	6.57	6.57	49.5	2.78	2.82	49.5	1.43	1.52	49.4	0.77	0.95
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
50	X1	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	P	50.0	7.05	7.05	50.1	3.40	3.40	50.0	1.86	1.86	50.0	1.07	1.07
	CUE	50.0	5.96	5.96	50.0	2.68	2.68	50.0	1.44	1.44	50.1	0.91	0.92
	MLE	49.7	6.02	6.03	49.5	2.68	2.72	49.5	1.42	1.51	49.4	0.76	0.95
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.4a

N = 100		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	22.8	4.15	77.3	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	P	22.8	4.15	77.3	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	CUE	80.6	33.2	38.5	99.9	29.1	29.1	100	17.2	17.2	100	11.9	11.9
	MLE	115	66.4	68.0	108	39.2	39.9	102	18.5	18.6	101	12.3	12.3
	P(inf mle)	0.1018			0.0001			0.0000			0.0000		
5	X1	22.7	4.18	77.4	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	P	77.2	40.1	46.1	95.4	50.2	50.5	99.1	45.7	45.7	99.9	38.1	38.1
	CUE	92.0	39.8	40.6	100	26.3	26.3	100	16.2	16.2	100	11.5	11.5
	MLE	123	77.9	81.2	105	31.9	32.4	101	17.2	17.2	100	11.8	11.8
	P(inf mle)	0.0282			0.0000			0.0000			0.0000		
10	X1	22.6	4.18	77.5	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	P	93.1	54.5	54.9	99.7	46.8	46.8	100	34.5	34.5	100	26.6	26.6
	CUE	97.1	42.6	42.7	100	24.3	24.3	100	15.5	15.5	100	11.1	11.1
	MLE	122	77.6	80.6	104	27.7	28.0	101	16.2	16.2	100	11.4	11.4
	P(inf mle)	0.0078			0.0000			0.0000			0.0000		
25	X1	22.6	4.18	77.5	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	P	99.8	48.0	48.0	100	29.2	29.2	100	21.5	21.5	100	17.1	17.1
	CUE	99.9	38.5	38.5	100	20.7	20.7	100	14.0	14.0	100	10.4	10.4
	MLE	111	54.2	55.4	102	22.2	22.3	101	14.4	14.4	100	10.5	10.5
	P(inf mle)	0.0002			0.0000			0.0000			0.0000		
50	X1	22.6	4.18	77.5	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	P	100	34.4	34.4	100	21.7	21.7	100	16.3	16.3	100	13.1	13.1
	CUE	100	30.9	30.9	100	18.0	18.0	100	12.7	12.7	100	9.65	9.65
	MLE	105	35.1	35.5	101	18.7	18.7	100	12.9	12.9	99.9	9.73	9.73
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
100	X1	22.6	4.18	77.5	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	P	100	26.9	26.9	100	17.5	17.5	100	13.2	13.2	100	10.6	10.6
	CUE												
	MLE	102	26.9	27.0	100	16.1	16.1	100	11.6	11.6	99.7	8.98	8.99
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.4b

N = 100		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	P	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	CUE	99.9	29.8	29.8	100	11.7	11.7	100	6.83	6.83	100	4.50	4.50
	MLE	108	41.4	42.2	100	12.2	12.2	99.8	6.89	6.89	99.6	4.51	4.53
	P(inf mle)	0.0002			0.0000			0.0000			0.0000		
5	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	P	95.8	50.2	50.4	99.9	37.3	37.3	100	25.1	25.1	100	17.7	17.7
	CUE	100	26.7	26.7	100	11.3	11.3	100	6.69	6.69	100	4.46	4.46
	MLE	106	33.0	33.5	100	11.6	11.6	99.7	6.75	6.75	99.6	4.47	4.49
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
10	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	P	99.8	46.0	46.0	100	26.0	26.0	100	17.4	17.4	100	12.5	12.5
	CUE	100	24.5	24.5	100	11.0	11.0	100	6.58	6.58	100	4.41	4.41
	MLE	104	28.3	28.6	100	11.2	11.2	99.7	6.63	6.64	99.6	4.43	4.45
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
25	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	P	100	28.6	28.6	100	16.7	16.7	100	11.5	11.5	100	8.39	8.39
	CUE	100	20.7	20.7	100	10.2	10.2	100	6.33	6.33	100	4.31	4.31
	MLE	102	22.2	22.3	100	10.4	10.4	99.7	6.36	6.37	99.6	4.32	4.34
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
50	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	P	100	21.3	21.3	100	12.8	12.8	100	8.86	8.86	100	6.47	6.47
	CUE	100	17.9	17.9	100	9.48	9.48	100	6.07	6.07	100	4.19	4.19
	MLE	101	18.6	18.6	99.9	9.56	9.56	99.6	6.08	6.09	99.5	4.19	4.22
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
100	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	P	100	17.2	17.2	100	10.4	10.4	100	7.22	7.22	100	5.28	5.28
	CUE												
	MLE	100	15.9	15.9	99.7	8.81	8.81	99.6	5.78	5.79	99.5	4.06	4.09
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

Table 1.4c

N = 100		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
0	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	P	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	CUE	100	11.4	11.4	100	4.17	4.17	100	2.08	2.08	100	1.12	1.12
	MLE	100	11.8	11.8	99.6	4.19	4.21	99.5	2.08	2.14	99.5	1.19	1.27
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
5	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	P	100	35.6	35.6	100	16.5	16.5	100	8.85	8.85	100	4.99	4.99
	CUE	100	11.0	11.0	100	4.14	4.14	100	2.07	2.07	99.9	1.12	1.12
	MLE	100	11.3	11.3	99.6	4.14	4.16	99.5	2.07	2.13	99.5	1.18	1.27
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
10	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	P	100	24.6	24.6	100	11.7	11.7	100	6.43	6.43	100	3.61	3.61
	CUE	100	10.6	10.6	100	4.09	4.09	100	2.06	2.06	99.9	1.11	1.12
	MLE	100	10.9	10.9	99.6	4.10	4.12	99.5	2.06	2.12	99.6	1.18	1.26
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
25	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	P	100	15.9	15.9	100	7.80	7.80	100	4.27	4.27	100	2.40	2.40
	CUE	99.9	9.84	9.84	100	4.00	4.00	100	2.05	2.05	99.9	1.11	1.11
	MLE	99.9	9.98	9.98	99.5	4.02	4.04	99.5	2.05	2.11	99.6	1.18	1.25
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
50	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	P	100	12.2	12.2	100	6.01	6.01	100	3.25	3.25	100	1.85	1.85
	CUE	100	9.11	9.11	100	3.89	3.89	100	2.02	2.02	99.9	1.10	1.11
	MLE	99.8	9.18	9.19	99.5	3.91	3.93	99.5	2.03	2.09	99.6	1.17	1.24
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		
100	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	P	100	9.92	9.92	100	4.90	4.90	100	2.62	2.62	100	1.51	1.51
	CUE												
	MLE	99.7	8.44	8.44	99.5	3.78	3.81	99.5	2.00	2.06			
	P(inf mle)	0.0000			0.0000			0.0000			0.0000		

§ 1.10 : The Unconditional Performance of the CUE and the Peterson-Type Estimators

In the previous section, in order to compare the relative performance of all three estimators considered within this chapter, it was necessary to consider the distribution of each conditional on the event $C=\{Z>X_1\}$. This was necessary because the MLE is known to yield infinite estimates when $Z=X_1$, and this would result in the MLE having an infinite mean, standard deviation and mean square error over the entire joint distribution of X_1 and Z . The discussion of the previous section recommended that, conditional on the event $C=\{Z>X_1\}$, one should favour the CUE \hat{N}_u . In view of this, and due to the fact that the CUE and Peterson-type estimators are both finite with probability one, the distributional properties of \hat{N}_u and \hat{N}_p are now presented unconditionally over the entire joint distribution of X_1 and Z . The results are contained in tables 1.5a,b,c, 1.6a,b,c, 1.7a,b,c and 1.8a,b,c. The values of N , t , p and R which are considered are identical to those of section 1.9. Notation is the same as in previous section.

§ 1.10a : Discussion

The performance, and relative performance, of the estimators over the complete sample space is seen to be very similar to their performance in the previous section. As one would expect, the difference is most noticeable when the probability of X_1 being equal to Z is large.

Tables 1.5 to 1.8 clearly indicate that the overall performance of the CUE is superior to that of \hat{N}_p . In the absence of plants \hat{N}_p reduces to X_1 , so it is not surprising that \hat{N}_u is seen to clearly outperform \hat{N}_p in this situation. When $R>0$, both perform very well in terms of bias, with \hat{N}_u almost always being the less biased of the two. The standard deviation of \hat{N}_u is, in all but four of the situations considered, less than that of \hat{N}_p : the difference, when not in favour of \hat{N}_u , is small. The standard deviation of \hat{N}_p tends to be large when R , greater than zero, is small relative to N , and it is in these situations that the standard deviation of \hat{N}_u is appreciably smaller than that of \hat{N}_p . In each of the four situations where, for $R>0$, the standard deviation of \hat{N}_u is greater than that of \hat{N}_p , the CUE exhibits a smaller bias. In conclusion, one should always use the CUE in preference to the Peterson-type estimator \hat{N}_p .

Table 1.5a

N = 10		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	2.26	1.32	7.85	4.01	1.55	6.18	5.37	1.58	4.89	6.42	1.52	3.89
0	P	2.26	1.32	7.85	4.01	1.55	6.18	5.37	1.58	4.89	6.42	1.52	3.89
	CUE	3.70	3.13	7.04	7.31	4.60	5.33	9.10	4.62	4.71	9.80	4.02	4.03
5	P	7.87	6.29	6.64	9.60	6.35	6.36	9.98	5.46	5.46	10.0	4.51	4.51
	CUE	8.24	6.79	7.02	9.78	6.14	6.15	9.95	4.49	4.49	10.0	3.31	3.31
10	P	9.50	8.16	8.18	10.0	6.16	6.16	10.0	4.54	4.54	10.0	3.56	3.56
	CUE	9.50	8.11	8.12	9.95	5.59	5.59	9.97	3.87	3.87	9.99	2.93	2.93

Table 1.5b

N = 10		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	4.10	1.56	6.11	6.51	1.51	3.80	7.94	1.28	2.42	8.78	1.03	1.60
0	P	4.10	1.56	6.11	6.51	1.51	3.80	7.94	1.28	2.42	8.78	1.03	1.60
	CUE	7.27	4.41	5.18	9.71	3.94	3.95	10.0	2.44	2.44	10.0	1.58	1.58
5	P	9.64	6.31	6.32	10.1	4.42	4.42	10.1	3.02	3.02	10.1	2.16	2.16
	CUE	9.81	6.00	6.01	10.0	3.25	3.25	9.99	2.02	2.02	10.0	1.34	1.34
10	P	10.0	6.05	6.05	10.1	3.47	3.48	10.1	2.30	2.30	10.1	1.63	1.63
	CUE	9.97	5.45	5.45	10.0	2.88	2.88	10.0	1.88	1.88	10.0	1.29	1.29

Table 1.5c

N = 10		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	6.72	1.48	3.60	8.93	0.98	1.45	9.65	0.58	0.68	9.88	0.34	0.36
0	P	6.72	1.48	3.60	8.93	0.98	1.45	9.65	0.58	0.68	9.88	0.34	0.36
	CUE	9.91	3.79	3.79	9.99	1.45	1.45	9.86	0.75	0.76	9.90	0.35	0.37
5	P	10.1	4.21	4.22	10.1	2.00	2.00	10.0	1.07	1.07	10.0	0.60	0.60
	CUE	10.1	3.10	3.10	10.0	1.29	1.29	9.85	0.73	0.75	9.89	0.34	0.36
10	P	10.1	3.30	3.30	10.0	1.52	1.52	10.0	0.84	0.84	10.0	0.48	0.48
	CUE	10.0	2.74	2.74	10.0	1.21	1.21	9.80	0.70	0.73	9.89	0.34	0.36

Table 1.6a

N = 25		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	5.66	2.09	19.5	10.0	2.45	15.2	13.4	2.49	11.8	16.0	2.40	9.28
0	P	5.66	2.09	19.5	10.0	2.45	15.2	13.4	2.49	11.8	16.0	2.40	9.28
	CUE	13.8	8.98	14.3	23.3	12.3	12.4	24.9	9.56	9.56	25.0	6.56	6.56
5	P	19.7	12.3	13.4	23.9	13.8	13.8	24.8	12.2	12.2	25.0	10.1	10.1
	CUE	21.8	14.6	15.0	24.9	12.0	12.0	25.0	7.82	7.82	25.0	5.53	5.53
10	P	23.6	16.4	16.4	25.0	13.1	13.1	25.0	9.62	9.62	25.0	7.44	7.44
	CUE	24.1	16.4	16.4	25.0	10.3	10.3	25.0	6.86	6.86	25.0	5.09	5.09
25	P	25.0	14.9	14.9	25.0	9.13	9.13	25.0	6.77	6.77	25.0	5.39	5.39
	CUE	25.0	14.1	14.1	25.0	8.19	8.19	25.0	5.85	5.85	25.0	4.51	4.51

Table 1.6b

N = 25		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	10.2	2.46	15.0	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
0	P	10.2	2.46	15.0	16.3	2.38	9.04	19.9	2.02	5.53	22.0	1.63	3.45
	CUE	23.1	12.0	12.1	25.1	6.43	6.43	25.0	3.56	3.56	25.0	2.29	2.29
5	P	24.0	13.7	13.8	25.0	9.92	9.92	25.0	6.70	6.70	25.0	4.78	4.78
	CUE	24.9	12.0	12.0	25.0	5.41	5.41	25.0	3.26	3.26	25.0	2.21	2.21
10	P	25.0	12.8	12.8	25.0	7.26	7.26	25.0	4.90	4.90	24.9	3.54	3.54
	CUE	25.0	10.3	10.3	25.0	4.96	4.96	25.0	3.14	3.14	25.0	2.14	2.14
25	P	25.0	8.95	8.95	25.0	5.27	5.27	25.0	3.63	3.63	25.1	2.57	2.57
	CUE	25.0	8.06	8.06	25.0	4.41	4.41	25.0	2.91	2.91	25.0	2.05	2.05

Table 1.6c

N = 25		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
0	P	16.8	2.35	8.52	22.3	1.55	3.10	24.1	0.92	1.27	24.7	0.53	0.61
	CUE	25.0	6.30	6.30	25.0	2.14	2.14	25.1	1.07	1.07	24.8	0.61	0.64
5	P	25.1	9.45	9.45	25.0	4.44	4.44	25.0	2.40	2.40	25.0	1.34	1.34
	CUE	25.1	5.26	5.26	25.0	2.05	2.05	25.1	1.03	1.04	24.8	0.60	0.64
10	P	25.0	6.88	6.88	24.9	3.29	3.29	25.0	1.86	1.86	25.0	1.11	1.11
	CUE	25.0	4.81	4.81	25.0	1.99	1.99	25.1	1.02	1.02	24.8	0.58	0.63
25	P	25.0	5.02	5.02	25.1	2.38	2.38	25.0	1.32	1.32	25.0	0.76	0.76
	CUE	25.0	4.26	4.26	25.0	1.91	1.91	25.1	0.98	0.98	24.7	0.56	0.62

Table 1.7a

N = 50		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	11.3	2.96	38.8	20.1	3.47	30.1	26.8	3.53	23.4	32.1	3.39	18.2
0	P	11.3	2.96	38.8	20.1	3.47	30.1	26.8	3.53	23.4	32.1	3.39	18.2
	CUE	37.4	21.5	24.9	50.0	21.8	21.8	49.9	12.8	12.8	50.0	8.65	8.65
5	P	39.3	21.8	24.3	47.8	25.9	26.0	49.6	23.4	23.4	50.0	19.5	19.5
	CUE	46.5	28.3	28.5	50.0	18.6	18.6	50.0	11.3	11.3	50.0	8.01	8.01
10	P	47.1	29.6	29.7	49.9	24.3	24.3	50.0	17.9	17.9	50.0	13.9	13.9
	CUE	49.0	29.2	29.2	50.0	16.1	16.1	50.0	10.4	10.4	50.0	7.58	7.58
25	P	50.0	26.2	26.2	50.0	15.9	15.9	50.0	11.8	11.8	50.0	9.36	9.36
	CUE	49.9	23.4	23.4	50.0	13.0	13.0	50.0	9.07	9.07	50.0	6.87	6.87
50	P	50.0	19.7	19.7	50.0	12.5	12.5	50.0	9.43	9.43	50.0	7.55	7.55
	CUE	50.0	18.6	18.6	50.0	11.3	11.3	50.0	8.18	8.18	50.0	6.33	6.33

Table 1.7b

N = 50		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	20.5	3.48	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
0	P	20.5	3.48	29.7	32.6	3.37	17.8	39.7	2.86	10.7	43.9	2.31	6.50
	CUE	49.6	22.0	22.0	50.0	8.54	8.54	50.0	4.86	4.86	50.0	3.21	3.21
5	P	47.9	25.9	26.0	50.0	19.1	19.1	50.0	12.8	12.8	50.0	9.09	9.09
	CUE	50.0	18.8	18.8	50.0	7.89	7.89	50.0	4.71	4.71	50.0	3.14	3.14
10	P	49.9	23.9	23.9	50.0	13.5	13.5	50.0	9.07	9.07	50.0	6.58	6.58
	CUE	50.1	16.2	16.2	50.0	7.48	7.48	50.0	4.55	4.55	50.0	3.08	3.08
25	P	50.0	15.6	15.6	50.0	9.15	9.15	50.0	6.30	6.30	50.1	4.54	4.54
	CUE	50.0	12.9	12.9	50.0	6.75	6.75	50.0	4.31	4.31	50.0	2.97	2.97
50	P	50.0	12.3	12.3	50.0	7.39	7.39	50.0	5.13	5.13	50.1	3.68	3.68
	CUE	50.0	11.2	11.2	50.0	6.20	6.20	50.0	4.09	4.09	50.0	2.88	2.88

Table 1.7c

N = 50		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
0	P	33.6	3.32	16.7	44.6	2.19	5.80	48.2	1.30	2.19	49.4	0.75	0.95
	CUE	49.9	8.30	8.30	50.0	2.98	2.98	50.0	1.50	1.50	50.0	0.92	0.92
5	P	50.0	18.2	18.2	50.0	8.43	8.43	50.0	4.56	4.56	50.0	2.56	2.56
	CUE	50.0	7.66	7.66	50.0	2.91	2.91	50.0	1.49	1.49	50.0	0.92	0.92
10	P	50.0	12.8	12.8	50.0	6.12	6.12	50.0	3.35	3.35	50.0	1.88	1.88
	CUE	49.9	7.20	7.20	50.0	2.86	2.86	50.0	1.49	1.49	50.1	0.91	0.91
25	P	50.0	8.71	8.71	50.1	4.21	4.21	50.0	2.31	2.31	50.0	1.32	1.32
	CUE	50.0	6.46	6.46	50.0	2.76	2.76	50.0	1.47	1.47	50.1	0.91	0.92
50	P	50.0	7.05	7.05	50.1	3.40	3.40	50.0	1.86	1.86	50.0	1.07	1.07
	CUE	50.0	5.96	5.96	50.0	2.68	2.68	50.0	1.44	1.44	50.1	0.91	0.92

Table 1.8a

N = 100		t											
p = 0.05		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	22.6	4.18	77.5	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
0	P	22.6	4.18	77.5	40.1	4.90	60.1	53.7	4.99	46.6	64.2	4.80	36.2
	CUE	92.0	51.6	52.2	100	29.6	29.6	100	17.2	17.2	100	11.9	11.9
5	P	78.6	40.5	45.8	95.5	50.2	50.5	99.1	45.7	45.7	99.9	38.1	38.1
	CUE	97.8	54.1	54.2	100	26.4	26.4	100	16.2	16.2	100	11.5	11.5
10	P	94.1	55.7	56.0	99.7	46.8	46.8	100	34.5	34.5	100	26.6	26.6
	CUE	99.4	51.1	51.1	100	24.3	24.3	100	15.5	15.5	100	11.1	11.1
25	P	99.9	48.4	48.4	100	29.2	29.2	100	21.5	21.5	100	17.1	17.1
	CUE	100	39.4	39.4	100	20.7	20.7	100	14.0	14.0	100	10.4	10.4
50	P	100	34.4	34.4	100	21.7	21.7	100	16.3	16.3	100	13.1	13.1
	CUE	100	30.9	30.9	100	18.0	18.0	100	12.7	12.7	100	9.65	9.65
100	P	100	26.9	26.9	100	17.5	17.5	100	13.2	13.2	100	10.6	10.6
	CUE												

Table 1.8b

N = 100		t											
p = 0.10		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
0	P	41.0	4.92	59.3	65.1	4.77	35.2	79.4	4.04	21.0	87.8	3.27	12.6
	CUE	100	30.5	30.5	100	11.7	11.7	100	6.83	6.83	100	4.50	4.50
5	P	95.8	50.2	50.4	99.9	37.3	37.3	100	25.1	25.1	100	17.7	17.7
	CUE	100	26.9	26.9	100	11.3	11.3	100	6.69	6.69	100	4.46	4.46
10	P	99.8	46.0	46.0	100	26.0	26.0	100	17.4	17.4	100	12.5	12.5
	CUE	100	24.5	24.5	100	11.0	11.0	100	6.58	6.58	100	4.41	4.41
25	P	100	28.6	28.6	100	16.7	16.7	100	11.5	11.5	100	8.39	8.39
	CUE	100	20.7	20.7	100	10.2	10.2	100	6.33	6.33	100	4.31	4.31
50	P	100	21.3	21.3	100	12.8	12.8	100	8.86	8.86	100	6.47	6.47
	CUE	100	17.9	17.9	100	9.48	9.48	100	6.07	6.07	100	4.19	4.19
100	P	100	17.2	17.2	100	10.4	10.4	100	7.22	7.22	100	5.28	5.28
	CUE												

Table 1.8c

N = 100		t											
p = 0.20		5			10			15			20		
R	Estimator	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse	exp.	s.d.	rmse
>=0	X1	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
0	P	67.2	4.69	33.1	89.3	3.10	11.2	96.5	1.84	3.97	98.8	1.07	1.57
	CUE	100	11.4	11.4	100	4.17	4.17	100	2.08	2.08	100	1.12	1.12
5	P	100	35.6	35.6	100	16.5	16.5	100	8.85	8.85	100	4.99	4.99
	CUE	100	11.0	11.0	100	4.14	4.14	100	2.07	2.07	99.9	1.12	1.12
10	P	100	24.6	24.6	100	11.7	11.7	100	6.43	6.43	100	3.61	3.61
	CUE	100	10.6	10.6	100	4.09	4.09	100	2.06	2.06	99.9	1.11	1.12
25	P	100	15.9	15.9	100	7.80	7.80	100	4.27	4.27	100	2.40	2.40
	CUE	99.9	9.84	9.84	100	4.00	4.00	100	2.05	2.05	99.9	1.11	1.11
50	P	100	12.2	12.2	100	6.01	6.01	100	3.25	3.25	100	1.85	1.85
	CUE	100	9.11	9.11	100	3.89	3.89	100	2.02	2.02	99.9	1.10	1.11
100	P	100	9.92	9.92	100	4.90	4.90	100	2.62	2.62	100	1.51	1.51
	CUE												

§ 1.11 : The Performance of Plant-Capture

When Applied to the Model M_0 : Under Discrete Time Sampling

In previous sections it has been argued that the overall performance of the CUE \hat{N}_u should always be considered superior to that of the MLE and Peterson-type estimators. Rather than only discussing the way in which the information gained through plants may improve the performance of the CUE, this section considers how the method of plant-capture can affect the performance of all the estimators described within this chapter. This approach is taken since, in spite of the evidence of the previous sections, it is believed that more traditionally minded practitioners may still prefer to use the MLE. The following discussion is based on an inspection of all the 24 tables of this chapter.

It has previously been mentioned that mean square error is known to reward negative bias, and that this characteristic can lead to incorrect conclusions being drawn, that is if one places too much emphasis on mean square error alone. When comparing the performance of estimators, one should always, where possible, consider firstly their mean and standard deviation, and only then should one consider mean square error, or alternative loss functions such as mean absolute deviation. This approach is taken in the following discussion ; consideration of mean square error alone can lead to counter intuitive conclusions. For example, consider the performance of the CUE in table 1.6a, where $N = 25$, $p = 0.05$ and $t = 5$. As R is increased from 0 to 10 the mean square error of \hat{N}_u increases from 14.3 to 16.4 ! However, only when one considers the way in which the bias of \hat{N}_u is being significantly reduced can one see that the extra information gained from the plants is in fact improving the performance of the CUE.

This last example is quite typical of the way in which the information gained from plants enhances the performance of the estimators in situations where very little information is gained from the target population, however in many of these situations the improvement in bias is accompanied by a reduction in mean square error.

Except for situations where only a small amount of information is available, the CUE \hat{N}_u is usually unbiased, and where not its bias is negligible. In those situations where only a small amount of information is available, \hat{N}_u tends to be negatively biased, with this bias reducing significantly and uniformly as the number of plants is increased. This behaviour is intuitively very reasonable, since the CUE is unbiased conditional on the event $Z = Z_1 + Z_2 \geq N$. That is because $Z_2 \sim \text{Bin}(Rt, p)$, the event $Z = Z_1 + Z_2 \geq N$ is more and more likely to occur as R is increased. The standard deviation of \hat{N}_u is generally seen to reduce uniformly as more and more plants are used. Where the standard deviation of the CUE is not reduced by an increase in R , this is always due to its bias being significantly improved.

In the absence of plants, the Peterson-type estimator reduces to X_1 , commonly referred to as the 'enumeration estimator'. For this reason, \hat{N}_p is only considered here when plants are used. In terms of bias, the Peterson-type estimator behaves in a very similar way to the CUE, although \hat{N}_u is on almost all occasions less biased. When the number of plants is small relative to the size of the target population, \hat{N}_p tends to have a relatively large variance, however this is on almost all occasions seen to reduce uniformly as R is increased.

The CUE is seen to utilise the information gained through the plants in a very 'smooth' way. That is, as R is increased, usually either the mean of \hat{N}_u is significantly improved at the expense of a slight increase in standard deviation or both the bias and the standard deviation are reduced. The behaviour of the MLE when in situations where little information is available is less predictable as more plants are introduced. Consider for example table 1.3a, in which $N = 50$ and $p = 0.05$. When the number of sampling occasions is equal to 5, for $R = 0, 5, 10, 25$ and 50 the mean and standard deviation of \hat{N} are respectively 43.3, 24.4; 57.9, 36.3; 61.5, 41.8; 56.8, 33.1 and 52.7, 21.2 : the corresponding values taken by \hat{N}_u are respectively 30.1, 13.4; 40.8, 18.7; 46.3, 22.3; 49.8, 22.5 and 50.0, 18.6. The CUE in this example behaves in the manner described above, i.e. as R is increased its performance improves 'smoothly'. However, as the number of plants is increased from 0 to 5 to 10, both the mean and standard deviation of the MLE are seen to become worse ! This result appears counter intuitive, that is until one considers the way in which the value of R affects the probability of obtaining a finite MLE. In the above situation, where $t = 5$, for $R = 0, 5, 10, 25$ and 50, the probabilities of obtaining an infinite MLE are respectively 0.3190, 0.0885, 0.0245, 0.0005 and 0.0000. In other words the introduction of plants is seen to dramatically improve the probability of obtaining a useful MLE. When this advantage is considered along with the performance of the MLE, it can be argued that even in situations where, as in the above example, very little information is obtained from the target population, the presence of plants is beneficial to the overall performance of the MLE. Other than those extreme situations in which very little information is present, an increase in the number of plants is generally seen to improve the performance of the MLE via a reduction in both bias and standard deviation. And where both statistics are not improved, one of the two is.

In conclusion, the introduction of plants can be seen to enhance the performance of all three of the estimators which have been considered within this chapter, this being under the assumption that the planted individuals do indeed behave in an identical manner to members of the target population. In particular the plants are seen to be of most use when only little information has been gained from the target population. Furthermore, on the basis of the above discussions, whether sampling with or without

plants, it is recommended that the CUE be considered superior to both the MLE and Peterson-type estimator.

Chapter 2 : A Plant-Capture Approach for Sequential Tagging

§ 2.1 : Introduction

This chapter considers how the method of plant-capture may be used to aid the problem of estimating population size when the population in question behaves according to a continuous time analogue of the standard capture-recapture model known as M_0 . The model M_0 is one of the sequence of models described by Otis et al. (1978) for capture-recapture data in closed populations. The sampling procedure considered within this chapter assumes that the population in question is under constant observation for some period of time, and that individuals are seen one at a time. Existing methods for estimating the value of N are based on either truncated sampling, in which sampling continues for a fixed predetermined amount of time, or censored sampling, in which sampling continues until a predetermined number of tagged individuals have been seen. Within this chapter consideration is given to the problem of estimation under the more commonly used method of truncated sampling. In the absence of plants this version of the problem has previously been studied by Nayak(1988), who derived a maximum likelihood estimator.

§ 2.2 : Sampling Procedure and Assumptions

Prior to the commencement of the experiment it is assumed that the target population, whose size N we wish to estimate, is augmented by the insertion of a known number R of planted individuals. Each planted individual is assumed to have received a unique tag prior to its release. Sightings of any particular member of the target population form a homogeneous Poisson process of rate λ . The augmented population, of size $N+R$, is randomly mixed. It is assumed that the planted individuals behave exactly as members of the target population, so that the augmented population constitutes $N+R$ independent homogeneous Poisson processes each of rate λ . One member of the augmented population is randomly selected at a time: individuals that are seen for the first time receive a unique tag, so that they may be recognised on subsequent occasions. Individuals having been seen are then immediately released into the population. The augmented population is assumed to be closed and under continuous observation during the predetermined time period $[0, \tau]$.

§ 2.3 : A Note on Software Reliability

In the above, N has been referred to as being the size of ' a population '. More specifically, N may represent the size of a wildlife population, in which case the above sampling procedure constitutes a sequential Schnabel census with samples of size one, see Schnabel(1938) or, for a more comprehensive review, Seber(1982). The theory discussed here however is equally applicable to the problem of estimating the original number of faults, N , in a reliability system. Only the interpretation of the theory in both cases is a little different : probabilistically both situations are identical. In the situation where N represents the unknown number of errors in a piece of computer software the above model was originally proposed by Jelinski & Moranda (1972). Other models which aim to describe the stochastic failures of a piece of software have been proposed however the Jelinski & Moranda model is commonly regarded as being central to the topic of software reliability, see Langberg & Singpurwalla (1985). Originally attempts to estimate the value of N based upon the Jelinski & Moranda model assumed that only the times at which errors were first detected would be recorded. Nayak(1988) introduced a design called recapture debugging in which he developed a sampling procedure, which is analogous to the sequential Schnabel census, in an attempt to get extra information from the population prior to estimating the value of N . In Nayak's model the software is assumed to originally contain N errors. Whenever an error is detected it is corrected, without further errors being inserted, but a counter is added to record how often that area of the software is accessed during the remainder of sampling time. This ensures that recapture debugging uses the available sampling time more efficiently.

§ 2.4 : The Sufficient Statistics

Nayak(1988) determined the sufficient statistics for the situation in which no plants are present. This section utilizes the theory of Nayak(1988) in order to determine the sufficient statistics for situations in which the number of plants is greater than or equal to zero. The following notation is used :

X_1 = the number of distinct unplanted individuals seen in time $[0, \tau]$.

X_2 = the number of distinct planted individuals seen in time $[0, \tau]$.

Z_1 = the total number of sightings made from the target population
in time $[0, \tau]$.

Z_2 = the total number of sightings made from the planted population
in time $[0, \tau]$.

Z = the total number of sightings made from the augmented population
in time $[0, \tau]$.

Within this section it is more convenient to let

$N^{(1)} = N$ = the size of the target population

and $N^{(2)} = R$ = the size of the planted population.

Suppose that the individuals within each population are labelled as 1, 2, 3, according to the order in which they were seen.

Consider firstly the target population.

Let

$T_{(i)}^{(1)}$ = the time at which individual i is first seen, for $i = 1, 2, \dots, X_1$.

It follows that $0 \leq T_{(1)}^{(1)} \leq T_{(2)}^{(1)} \leq \dots \leq T_{(X_1)}^{(1)} \leq \tau$.

and let

$M_i^{(1)}$ = the number of times individual i is recaptured after its initial capture, $i = 1, 2, \dots, X_1$. In other words $M_i^{(1)} = Y_i^{(1)} - 1$, where $Y_i^{(1)}$ is the total number of times that individual i is seen during the time interval $[0, \tau]$.

Similarly for the planted population let

$T_{(i)}^{(2)}$ = the time at which individual i is first seen, for $i = 1, 2, \dots, X_2$.

It follows that $0 \leq T_{(1)}^{(2)} \leq T_{(2)}^{(2)} \leq \dots \leq T_{(X_2)}^{(2)} \leq \tau$.

and let

$M_i^{(2)}$ = the number of times individual i is recaptured after its initial capture, $i = 1, 2, \dots, X_2$. In other words $M_i^{(2)} = Y_i^{(2)} - 1$, where $Y_i^{(2)}$ is the total number of times that individual i is seen during the time interval $[0, \tau]$.

Now define the vectors

$$U_{(1)} = (X_1, T_{(1)}^{(1)}, T_{(2)}^{(1)}, \dots, T_{(X_1)}^{(1)})$$

$$V_{(1)} = (M_1^{(1)}, M_2^{(1)}, \dots, M_{X_1}^{(1)})$$

and

$$U_{(2)} = (X_2, T_{(1)}^{(2)}, T_{(2)}^{(2)}, \dots, T_{(X_2)}^{(2)})$$

$$V_{(2)} = (M_1^{(2)}, M_2^{(2)}, \dots, M_{X_2}^{(2)}).$$

By the independence of the target and planted populations, it follows that

$$\begin{aligned} \text{Prob}(U_{(1)}, V_{(1)}, U_{(2)}, V_{(2)}) &= \prod_{i=1}^2 \text{Prob}(U_{(i)} = u_{(i)}, V_{(i)} = v_{(i)}) \\ &= \prod_{i=1}^2 \text{Prob}(V_{(i)} = v_{(i)} | U_{(i)} = u_{(i)}) \text{Prob}(U_{(i)} = u_{(i)}). \end{aligned} \quad (2.1)$$

Under the model, the following distributional results hold :

(i) Given $U_{(i)}$, $M_j^{(i)} \sim P(\lambda(\tau - t_{(j)}^{(i)}))$, for $j = 1, 2, \dots, X_i$; $i = 1, 2$,

where $P(\lambda)$ denotes a Poisson distribution with mean λ .

This implies that

$$\begin{aligned} \text{Prob}(V_{(i)}|U_{(i)}) &= \prod_{j=1}^{X_i} \text{Prob}(M_j^{(i)} = m_j^{(i)}) && \text{by independence} \\ &= \prod_{j=1}^{X_i} \frac{[\lambda(\tau - t_{(j)}^{(i)})]^{m_j^{(i)}} \exp[-\lambda(\tau - t_{(j)}^{(i)})]}{m_j^{(i)}!} \\ &= \frac{\lambda^{m^{(i)}} \exp\left[-\lambda X_i \tau + \lambda \sum_{j=1}^{X_i} t_{(j)}^{(i)}\right] \prod_{j=1}^{X_i} (\tau - t_{(j)}^{(i)})^{m_j^{(i)}}}{\prod_{j=1}^{X_i} m_j^{(i)}!}, \end{aligned}$$

$$\text{where } m^{(i)} = \sum_{j=1}^{X_i} m_j^{(i)}.$$

$$\begin{aligned} \text{(ii) } \text{Prob}(U_{(i)}) &= \text{Prob}(X_{(i)} = x_{(i)}) \text{Prob}(T_{(1)}^{(i)}, T_{(2)}^{(i)}, \dots, T_{(X_i)}^{(i)} | X_{(i)} = x_{(i)}), \\ &\text{where } X_{(i)} \sim \text{Bin}(N^{(i)}, 1 - \exp[-\lambda\tau]). \end{aligned}$$

Given that an individual is seen by time τ , its conditional time to detection has probability density function

$$\frac{\lambda \exp[-\lambda t]}{1 - \exp[-\lambda\tau]}, \quad 0 \leq t \leq \tau.$$

It then follows that the joint conditional probability distribution function of the order statistics of the X_i seen by time τ is

$$X_i! \prod_{j=1}^{X_i} \frac{\lambda \exp[-\lambda t_{(j)}^{(i)}}{1 - \exp[-\lambda\tau]} = \frac{X_i! \lambda^{X_i} \exp\left[-\lambda \sum_{j=1}^{X_i} t_{(j)}^{(i)}\right]}{(1 - \exp[-\lambda\tau])^{X_i}}, \quad \text{for } i = 1, 2.$$

Hence

$$\begin{aligned} \text{Prob}(U_{(i)} = u_{(i)}) &= \binom{N^{(i)}}{X_i} (1 - \exp[-\lambda\tau])^{X_i} (\exp[-\lambda\tau])^{N^{(i)} - X_i} \frac{X_i! \lambda^{X_i} \exp\left[-\lambda \sum_{j=1}^{X_i} t_{(j)}^{(i)}\right]}{(1 - \exp[-\lambda\tau])^{X_i}} \\ &= \binom{N^{(i)}}{X_i} (\exp[-\lambda\tau])^{N^{(i)} - X_i} X_i! \lambda^{X_i} \exp\left[-\lambda \sum_{j=1}^{X_i} t_{(j)}^{(i)}\right]. \end{aligned}$$

Substituting these results into equation (2.1) yields the following :

$$\begin{aligned} \text{Prob}(U_{(1)}, V_{(1)}, U_{(2)}, V_{(2)}) &= \prod_{i=1}^2 \text{Prob}(V_{(i)} = v_{(i)} | U_{(i)} = u_{(i)}) \text{Prob}(U_{(i)} = u_{(i)}) \\ &= \prod_{i=1}^2 \frac{\lambda^{m^{(i)}} \exp\left[-\lambda X_i \tau + \lambda \sum_{j=1}^{X_i} t_{(j)}^{(i)}\right] \prod_{j=1}^{X_i} (\tau - t_{(j)}^{(i)})^{m_j^{(i)}}}{\prod_{j=1}^{X_i} m_j^{(i)}!} \binom{N^{(i)}}{X_i} (\exp[-\lambda\tau])^{N^{(i)} - X_i} X_i! \lambda^{X_i} \exp\left[-\lambda \sum_{j=1}^{X_i} t_{(j)}^{(i)}\right] \end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^2 \frac{N^{(i)}! \lambda^{X_i + m^{(i)}} \exp[-N^{(i)} \lambda \tau] \prod_{j=1}^{x_i} (\tau - t_{(i)}^{(j)})^{m_j^{(i)}}}{(N^{(i)} - X_i)! \prod_{j=1}^{x_i} m_j^{(i)}!} \\
&= \prod_{i=1}^2 \frac{N^{(i)}! \lambda^{Z_i} \exp[-N^{(i)} \lambda \tau] \prod_{j=1}^{x_i} (\tau - t_{(i)}^{(j)})^{m_j^{(i)}}}{(N^{(i)} - X_i)! \prod_{j=1}^{x_i} m_j^{(i)}!}.
\end{aligned}$$

It follows that the likelihood function for λ and $N = N^{(1)}$ may be written as

$$L(\lambda, N^{(1)}) \propto \binom{N^{(1)}}{X_1} \lambda^{Z_1 + Z_2} \exp[-(N^{(1)} + N^{(2)}) \lambda \tau].$$

Hence, by the Neyman-Pearson factorisation theorem, the sufficient statistics for λ and $N = N^{(1)}$ are X_1 and $Z = Z_1 + Z_2$.

§ 2.5 : The Distribution Function of the Sufficient Statistics

The most direct way of deriving the joint probability function of X_1 and Z is to consider the decomposition

$$\text{Prob}(X_1 = x_1, Z = z) = \text{Prob}(X_1 = x_1 | Z = z) \text{Prob}(Z = z). \quad (2.2)$$

Firstly, from the above assumptions it follows that Z has a Poisson distribution with parameter $(N + R)\lambda\tau$.

$$\text{Explicitly} \quad \text{Prob}(Z = z) = \frac{[(N + R)\lambda\tau]^z \exp[-(N + R)\lambda\tau]}{z!}, \quad z = 0, 1, 2, \dots \quad (2.3)$$

The conditional distribution of X_1 given Z has previously appeared in an urn model context, see Johnson and Kotz(1977) p.122. Suppose one thinks of the $N+R$ members of the augmented population as $N+R$ urns and that each time an individual is seen a ball is placed into the urn representing it. Initially let N of these urns be empty and R contain one ball. So that at any subsequent time the number of tagged individuals in the population is represented by the number of urns containing at least one ball. Now the probability of X_1 given Z is the probability that X_1 of the initially empty N urns contain at least one ball given that Z balls have been randomly allocated to the $N+R$ urns, the balls being allocated to the urns in such a way that the probability of a ball being allocated to any one urn is $\frac{1}{N + R}$. The distribution of X_1 given Z is then exactly the variation of the classical occupancy situation discussed in Johnson and Kotz (1977) p.122, where a derivation of the probability function of X_1 given Z may be found.

Alternatively one may obtain the probability function of X_1 given Z more directly as follows.

$$\text{Prob}(X_1 = x_1 | Z = z) = \sum_{z_1=0}^z \text{Prob}(X_1 = x_1 | Z_1 = z_1, Z = z) \text{Prob}(Z_1 = z_1 | Z = z),$$

using the theorem of total conditional probability.

$$= \sum_{z_1=0}^z \text{Prob}(X_1 = x_1 | Z_1 = z_1) \text{Prob}(Z_1 = z_1 | Z = z). \quad (2.3a)$$

From assumptions, it is known that Z_1 has a Poisson distribution with mean $N\lambda\tau$. Since Z has a Poisson distribution with mean $(N+R)\lambda\tau$, it is easy to show that the distribution of $Z_1|Z$ is Binomial : explicitly $Z_1|Z \sim \text{Bin}\left(Z, \frac{N}{N+R}\right)$.

The conditional distribution of X_1 given Z_1 constitutes what is known in the literature as a Classical Occupancy distribution. For completeness, the Classical Occupancy distribution is described in Appendix 1, wherein its probability function is derived.

Explicitly

$$\text{Prob}(X_1 = x_1 | Z_1 = z_1) = N^{-z_1} \binom{N}{x_1} x_1! S(x_1, z_1),$$

$$x_1 = 0, 1, 2, \dots, \min(N, z_1),$$

where $S(x_1, z_1) = \frac{1}{x_1!} \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (x_1 - k)^{z_1}$ is a Stirling Number of the second kind.

The conditional distribution of X_1 given Z may now be obtained as follows :

$$\text{Prob}(X_1 = x_1 | Z = z) = \sum_{z_1=0}^z \text{Prob}(X_1 = x_1 | Z_1 = z_1) \text{Prob}(Z_1 = z_1 | Z = z) \quad \text{from (2.3a)}$$

$$= \sum_{z_1=0}^z N^{-z_1} \binom{N}{x_1} x_1! S(x_1, z_1) \binom{z}{z_1} \left(\frac{N}{N+R}\right)^{z_1} \left(\frac{R}{N+R}\right)^{z-z_1}$$

$$= \binom{N}{x_1} \frac{x_1!}{(N+R)^z} \sum_{z_1=0}^z S(x_1, z_1) \binom{z}{z_1} R^{z-z_1}$$

$$= \binom{N}{x_1} \frac{x_1!}{(N+R)^z} \sum_{z_1=0}^z \left(\frac{1}{x_1!} \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (x_1 - k)^{z_1} \right) \binom{z}{z_1} R^{z-z_1}$$

$$= \binom{N}{x_1} \frac{1}{(N+R)^z} \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k \sum_{z_1=0}^z \binom{z}{z_1} (x_1 - k)^{z_1} R^{z-z_1}$$

$$= \binom{N}{x_1} \frac{1}{(N+R)^z} \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R + x_1 - k)^z, \quad (2.4)$$

$$x_1 = 0, 1, 2, \dots, \min(N, z).$$

Substitution of (2.3) and (2.4) into equation (2.2) then yields the joint probability function for X_1 and Z :

$$\begin{aligned} \text{Prob}(X_1 = x_1, Z = z) &= \text{Prob}(X_1 = x_1 | Z = z) \text{Prob}(Z = z) \\ &= \binom{N}{x_1} \frac{1}{(N+R)^z} \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R+x_1-k)^z \frac{[(N+R)\lambda\tau]^z \exp[-(N+R)\lambda\tau]}{z!} \\ &= \frac{(\lambda\tau)^z \exp[-(N+R)\lambda\tau]}{z!} \binom{N}{x_1} \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R+x_1-k)^z \end{aligned} \quad (2.5)$$

$x_1 = 0, 1, 2, \dots, N,$
 $z = x_1, x_1 + 1, x_1 + 2, \dots$

§ 2.6 : The Q - Numbers

The Q - numbers are defined as follows :

$$Q(x_1, z; R) = \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R+x_1-k)^z,$$

$$x_1 = 0, 1, 2, \dots$$

$$z = x_1, x_1+1, x_1+2, \dots$$

These Q - numbers are of importance since they appear within the joint probability distribution function of the sufficient statistics, as given in the previous section.

The Q - numbers are a generalisation of the Stirling numbers of the second kind. Explicitly the relationship is given by the equation

$$Q(x_1, z; R = 0) = x_1! S(x_1, z). \quad (2.6)$$

In order to investigate the distributional properties of the estimators which are considered further on in this chapter it is necessary to evaluate the Q - numbers over some particular range of parameter values. This can lead to computational problems since the form of the Q - numbers is clearly not desirable from a computational point of view. That is the alternating sign within the summation means that very large numbers are repeatedly being added to and in particular subtracted from one another, and this is a major source of rounding error. To help avoid this, and other significant computational problems, one may consider the following 'triangular' recurrence relation of the Q - numbers :

$$Q(x_1, z; R) = x_1 Q(x_1 - 1, z - 1; R) + (R + x_1) Q(x_1, z - 1; R). \quad (2.7)$$

A direct proof of (2.7) is as follows :

$$\begin{aligned}
 & x_1 Q(x_1 - 1, z - 1; R) + (R + x_1) Q(x_1, z - 1; R) \\
 &= x_1 \sum_{k=0}^{x_1-1} \binom{x_1-1}{k} (-1)^k (R + x_1 - 1 - k)^{z-1} + (R + x_1) \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R + x_1 - k)^{z-1} \\
 &= x_1 \sum_{j=1}^{x_1} \binom{x_1-1}{j-1} (-1)^{j-1} (R + x_1 - j)^{z-1} + (R + x_1) \sum_{k=1}^{x_1} \binom{x_1}{k} (-1)^k (R + x_1 - k)^{z-1} \\
 &\quad + (R + x_1)^z \\
 &= (R + x_1)^z + \sum_{k=1}^{x_1} \left\{ -x_1 \binom{x_1-1}{k-1} + (R + x_1) \binom{x_1}{k} \right\} (-1)^k (R + x_1 - k)^{z-1} \\
 &= (R + x_1)^z + \sum_{k=1}^{x_1} \left\{ -x_1 \frac{(x_1-1)!}{(k-1)!(x_1-k)!} + (R + x_1) \binom{x_1}{k} \right\} (-1)^k (R + x_1 - k)^{z-1} \\
 &= (R + x_1)^z + \sum_{k=1}^{x_1} \left\{ -k \frac{x_1!}{k!(x_1-k)!} + (R + x_1) \binom{x_1}{k} \right\} (-1)^k (R + x_1 - k)^{z-1} \\
 &= (R + x_1)^z + \sum_{k=1}^{x_1} \left\{ (R + x_1 - k) \binom{x_1}{k} \right\} (-1)^k (R + x_1 - k)^{z-1} \\
 &= (R + x_1)^z + \sum_{k=1}^{x_1} \binom{x_1}{k} (-1)^k (R + x_1 - k)^z \\
 &= \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R + x_1 - k)^z \\
 &= Q(x_1, z; R).
 \end{aligned}$$

Upon using the identity (2.6), the recurrence relation (2.7), when $R = 0$, can be shown to reduce to the well known relationship between Stirling numbers of the second kind, namely

$$S(x_1, z) = S(x_1 - 1, z - 1) + x_1 S(x_1, z - 1).$$

The triangular recurrence relation (2.7) along with the initial conditions

$$Q(0, z; R) = R^z \text{ and } Q(x_1, x_1; R) = x_1! \tag{2.7a}$$

enables one to evaluate the required Q - numbers without having to perform any subtraction operations whatsoever, and hence one can more easily avoid computational rounding error.

N.B. The first initial condition is easy to show directly. The second can be shown to hold as follows. Firstly substituting $z = x_1$ into (2.7) implies that $Q(x_1, x_1; R) = x_1 Q(x_1 - 1, x_1 - 1; R)$, then after observing that $Q(0, 0; R) = 1$ it is easy to see that $Q(x_1, x_1; R) = x_1!$ for all $x_1 \geq 0$.

Comments

Using (2.7), one can show that a similar 'triangular' recurrence relation exists between the probabilities of the joint distribution of X_1 and Z , given by equation (2.5). It can be shown that

$$P_{x_1, z} = \frac{\lambda \tau}{z} \left[(R + x_1) P_{x_1, z-1} + (N - x_1 + 1) P_{x_1-1, z-1} \right] \quad (2.8)$$

where $P_{x_1, z} = \text{Prob}(X_1 = x_1, Z = z)$.

It is also straightforward to show that (2.8) is subject to the initial conditions

$$P_{0, z} = \frac{(\lambda \tau)^z \exp[-(N + R)\lambda \tau]}{z!} R^z \quad z = 0, 1, 2, \dots \quad (2.8a)$$

$$\text{and } P_{x_1, x_1} = (\lambda \tau)^{x_1} \exp[-(N + R)\lambda \tau] \binom{N}{x_1} \quad x_1 = 0, 1, 2, \dots, N. \quad (2.8b)$$

Again in an attempt to avoid numerical computational problems, one may use the following recurrence relations to determine the initial conditions (2.8a) and (2.8b) :

$$(i) \quad P_{0, z} = \left[\frac{(\lambda \tau)R}{z} \right] P_{0, z-1} \quad z = 1, 2, 3, \dots$$

$$(ii) \quad P_{x_1, x_1} = \left[\frac{(\lambda \tau)(N - x_1 + 1)}{x_1} \right] P_{x_1-1, x_1-1} \quad x_1 = 1, 2, 3, \dots, N.$$

Where the appropriate initial condition for both (i) and (ii) is $P_{0,0} = \exp[-(N + R)\lambda \tau]$.

§ 2.7 : The Maximum Likelihood Estimator

It follows from equation (2.5) that the joint likelihood function for N and λ may be written as

$$L(N, \lambda) \propto \lambda^z \exp[-(N + R)\lambda \tau] \binom{N}{x_1}. \quad (2.9)$$

This is now maximised over λ :

$$\frac{\partial L}{\partial \lambda} \propto \lambda^z (-(N + R)\tau) \exp[-(N + R)\lambda \tau] + z \lambda^{z-1} \exp[-(N + R)\lambda \tau],$$

equated to zero to obtain $\hat{\lambda}$:

$$\hat{\lambda}^z (N + R)\tau \exp[-(N + R)\hat{\lambda} \tau] = z \hat{\lambda}^{z-1} \exp[-(N + R)\hat{\lambda} \tau]$$
$$\Rightarrow \quad \hat{\lambda} = \frac{z}{(N + R)\tau}.$$

$\hat{\lambda}$ may now be substituted into (2.9) to obtain the profile likelihood for N :

$$L(N) \propto \binom{N}{x_1} (N+R)^{-z}, \quad N = x_1, x_1 + 1, x_1 + 2, \dots$$

There is no closed form expression for the value of N which maximises this profile likelihood function. However, for given values of x_1 and z , one may determine the estimate produced by the maximum likelihood estimator \hat{N} , using the following method.

Firstly, if $z = x_1$ the profile likelihood for N is clearly increasing and hence $\hat{N} = \infty$.

Now if $z > x_1$ one may observe that the profile likelihood function is uni-modal. Hence $\hat{N} = k$, where k is the smallest integer in the set $\{x_1, x_1 + 1, x_1 + 2, \dots\}$ able to

satisfy the condition $L(k) > L(k+1) \Leftrightarrow \frac{x_1}{k+1} + \left(\frac{k+R}{k+R+1}\right)^z < 1$.

Once \hat{N} has been determined, this value may then be used in the calculation of the maximum likelihood estimate of λ :

$$\hat{\lambda} = \frac{z}{(\hat{N} + R)\tau}$$

§ 2.8 : The Harmonic Mean Estimator

The Harmonic mean estimator was first considered by Joe and Reid (1985). Explicitly a point estimate of N is given by

$$\hat{N}_h = \left[0.5 + \left\{ \frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} \right\} \right] \quad \text{where } [.] \text{ denotes the integer part.}$$

The values of n_1 and n_2 are defined as follows :

$$n_1 = \inf\{ N \geq x : L(N|x) \geq cL(\hat{N}|x) \}$$

and $n_2 = \sup\{ N \geq x : L(N|x) \geq cL(\hat{N}|x) \},$

where $c \in (0,1]$.

§ 2.9 : A Peterson-Type Estimator

This section introduces an estimator of population size which is only dependent upon the observed numbers of distinct animals seen from the target and planted populations. The estimator is derived from the conditional distribution of X_1 given X .

Within section 2.4 the following distributional results were observed to be true

$$X_1 \sim \text{Bin}(N, 1 - \exp[-\lambda\tau]),$$

$$X_2 \sim \text{Bin}(R, 1 - \exp[-\lambda\tau])$$

and
$$X \sim \text{Bin}(N + R, 1 - \exp[-\lambda\tau]).$$

It is then easy to show that the distribution of $X_1|X$ is in fact hypergeometric with probability function

$$\text{Prob}(X_1 = x_1 | X = x) = \frac{\binom{N}{x_1} \binom{R}{x - x_1}}{\binom{N + R}{x}}, \quad \max(0, x - R) \leq x_1 \leq \min(N, x).$$

The likelihood function for N based on this probability function is maximised by the Peterson-type estimator $\tilde{N}_p = RX_1/X_2$. To avoid an estimator which becomes infinite when $X_2=0$, \tilde{N}_p is now slightly modified : from this point consideration is given to the estimator $\hat{N}_p = \left[0.5 + \frac{(R+1)X_1}{(X_2+1)} \right]$, where $[.]$ denotes the integer part of.

§2.9a : A New Estimator for Homogeneous Populations

A simple closed form estimator may be found by considering the expected value of $f_1 \sim$ the number of animals seen exactly once during the experiment.

From the assumptions made above it follows that

$$f_1 = \sum_{i=1}^{N+R} I(X_i = 1) \quad \text{where } I(X_i = 1) = \begin{cases} 1 & \text{w.p. } \lambda\tau \cdot \exp(-\lambda\tau) \\ 0 & \text{w.p. } 1 - \lambda\tau \cdot \exp(-\lambda\tau) \end{cases}$$

The expected value of f_1 is then given by

$$\begin{aligned} E[f_1] &= \sum_{i=1}^{N+R} E[I(X_i = 1)] \\ &= \sum_{i=1}^{N+R} \lambda\tau \cdot \exp(-\lambda\tau) \\ &= (N + R)\lambda\tau \cdot \exp(-\lambda\tau). \end{aligned}$$

By equating f_1 to its expected value one can obtain the equation

$$f_1 = (N + R)\lambda\tau \cdot \exp\left(\frac{-(N + R)\lambda\tau}{(N + R)}\right) \tag{2.9a}$$

The total number of sightings Z has a Poisson distribution with mean $(N + R)\lambda\tau$. Hence z may be used as an estimate of $(N + R)\lambda\tau$, and substituting this into equation (2.9a) yields

$$f_1 = z \cdot \exp\left(\frac{-z}{N + R}\right).$$

Solving this for N gives $N = \frac{z}{\ln\left\{\frac{z}{f_1}\right\}} - R$, and so a point estimate of population size N

is given by $\hat{N}_{fz} = \frac{z}{\ln\left\{\frac{z}{f_1}\right\}} - R$.

Care must be taken with this estimator since the estimate it produces is only dependent upon the values of f_1 and z . That is one may obtain a distorted view of population size if one observes the population only through the values of f_1 and z . This problem only occurs when sampling for a very long period of time and so in most practical situations it is expected that the above estimator may be considered for use in connection with the majority of experiments. Consider, for example, the data set contained within Table 4.2 of Seber(1982) p.137 which gives the capture-recapture data from a population of butterflies : from Craig(1953). For this data set $R=0$, $f_1=258$, $f_2=72$, $f_3=11$ and $z=435$. The maximum likelihood estimate and Darroch & Ratcliff(1980) estimate of population size are 853 and 838 respectively. The above estimator's estimate of population size is $\hat{N}_{fz} = \frac{435}{\ln\left\{\frac{435}{258}\right\}} = 833$.

Simulation results have shown that, provided no more than about 80% of the population is seen during the experiment, the estimator \hat{N}_{fz} competes very well with the maximum likelihood estimator and the estimator of Darroch & Ratcliff(1980). The behaviour of \hat{N}_{fz} , in situations where less than about 80% of the population is seen, is in fact very similar to the behaviour of the maximum likelihood estimator - although the performance of the maximum likelihood estimator is on the whole marginally superior. The performance of \hat{N}_{fz} diminishes after sampling has continued for a very long time, and is unsatisfactory when more than about 80% of the population is seen during the experiment. For this reason and the fact that \hat{N}_{fz} is not a function of the sufficient statistics alone, the estimator \hat{N}_{fz} is not considered any further at this stage.

§ 2.10 : A Conditionally Unbiased Estimator

This section introduces the Conditionally Unbiased Estimator (CUE) \tilde{N}_u , which is an estimator of population size N defined by

$$\tilde{N}_u = \frac{Q(x_1, z+1; R)}{Q(x_1, z; R)} - R. \quad (2.10)$$

$$\text{where } Q(x_1, z; R) = \sum_{k=0}^{x_1} \binom{x_1}{k} (-1)^k (R + x_1 - k)^z,$$

as defined in section 2.6.

The CUE \tilde{N}_u was derived from the conditional distribution of X_1 given Z . When $R=0$, \tilde{N}_u reduces to an estimator which has previously been considered in an urn model context, Harris(1968). In the absence of plants Harris(1968) showed that, provided $Z \geq N$, \tilde{N}_u is a minimum variance unbiased estimator of N with respect to the conditional distribution of X_1 given Z . As previously mentioned in section 1.8 of chapter 1, when $R=0$, \tilde{N}_u is equivalent to a special case of the estimator proposed by Pathak(1964). Berg(1975), using the notation of Pathak(1964), derived a recurrence relation for \tilde{N}_u in the $R = 0$ situation. In section 2.10a a recurrence relation for \tilde{N}_u is derived which allows for values of R greater than or equal to zero.

For values of R greater than or equal to zero, the CUE \tilde{N}_u can be shown to be unbiased with respect to the conditional distribution of X_1 given Z , provided that $Z \geq N$, as follows.

$$\begin{aligned} E(\tilde{N}_u) &= \sum_{x_1} \tilde{N}_u \text{Prob}(X_1 = x_1 | Z = z) \\ &= \sum_{x_1} \left[\frac{Q(x_1, z+1; R)}{Q(x_1, z; R)} - R \right] \binom{N}{x_1} \frac{1}{(N+R)^z} Q(x_1, z; R), \\ &\quad \text{this follows from equation (2.4).} \\ &= \sum_{x_1} \frac{Q(x_1, z+1; R)}{Q(x_1, z; R)} \binom{N}{x_1} \frac{1}{(N+R)^z} Q(x_1, z; R) - R \\ &= (N+R) \sum_{x_1} \binom{N}{x_1} \frac{1}{(N+R)^{z+1}} Q(x_1, z+1; R) - R \\ &= (N+R).1 - R \quad \text{if } Z \geq N. \\ &= N. \end{aligned}$$

Hence \tilde{N}_u is unbiased over the conditional distribution of X_1 given Z , provided that the condition $Z \geq N$ holds. Furthermore, again provided that $Z \geq N$, since X_1 is sufficient for N with respect to the conditional distribution of X_1 given Z , it follows

that \tilde{N}_u is the minimum variance unbiased estimator of N with respect to the conditional distribution of X_1 given Z, Lehmann and Scheffe(1950).

In view of the fact that population size N is integer valued, in later sections consideration is given to the following slightly modified version of \tilde{N}_u :

$$\hat{N}_u = \left[0.5 + \frac{Q(x_1, z+1; R)}{Q(x_1, z; R)} - R \right],$$

where the square brackets have been used to denote the integer part.

§ 2.10a : A Note on the Evaluation of the CUE

Direct use of equation (2.10) to evaluate the estimates produced by the estimator \tilde{N}_u can often be difficult, and involve very cumbersome computation. This is due to the fact that the Q - numbers, present within (2.10), grow rapidly with increasing arguments. To overcome this computational problem, a recurrence relation linking the \tilde{N}_u is now stated and proved.

To make the following proof more easily read some shorthand notation is necessary.

Let $Q_{x_1, z} = Q(x_1, z; R)$

and let $N_{x_1, z} = \tilde{N}_u = \frac{Q(x_1, z+1; R)}{Q(x_1, z; R)} - R = \frac{Q_{x_1, z+1}}{Q_{x_1, z}} - R.$

The $N_{x_1, z}$ are then subject to the following recurrence relation

$$N_{x_1, z} = x_1 + \left(\frac{N_{x_1-1, z-1} + R}{N_{x_1, z-1} + R} \right) (N_{x_1, z-1} - x_1), \quad (2.11)$$

with initial conditions $N_{0, z} = 0$ for $z = 0, 1, 2, \dots$ (2.12)

and $N_{x_1, x_1} = \frac{x_1}{2} [x_1 + 2R + 1]$ for $x_1 \geq 0.$ (2.13)

N.B. Substituting $R = 0$ into (2.11), (2.12) and (2.13) yields the 'Property 5' of Berg(1975) p.92.

Proof of (2.11) :

$$\begin{aligned}
 & x_1 + \left(\frac{N_{x_1-1,z-1} + R}{N_{x_1,z-1} + R} \right) (N_{x_1,z-1} - x_1) \\
 &= x_1 + \left(\frac{\frac{Q_{x_1-1,z} - R + R}{Q_{x_1-1,z-1}}}{\frac{Q_{x_1,z} - R + R}{Q_{x_1,z-1}}} \right) \left(\frac{Q_{x_1,z} - R - x_1}{Q_{x_1,z-1}} \right) \\
 &= x_1 + \frac{Q_{x_1-1,z}}{Q_{x_1-1,z-1}} \frac{Q_{x_1,z-1}}{Q_{x_1,z}} \frac{(Q_{x_1,z} - (R + x_1)Q_{x_1,z-1})}{Q_{x_1,z-1}} \\
 &= x_1 + \frac{Q_{x_1-1,z}}{Q_{x_1-1,z-1}} \frac{Q_{x_1,z-1}}{Q_{x_1,z}} \frac{(x_1 Q_{x_1-1,z-1})}{Q_{x_1,z-1}} && \text{using (2.7)} \\
 &= x_1 + \frac{x_1 Q_{x_1-1,z}}{Q_{x_1,z}} \\
 &= x_1 + \frac{(Q_{x_1,z+1} - (R + x_1)Q_{x_1,z})}{Q_{x_1,z}} && \text{using (2.7) with } z \text{ replaced by } z+1 \\
 &= x_1 + \frac{Q_{x_1,z+1}}{Q_{x_1,z}} - (R + x_1) \\
 &= \frac{Q_{x_1,z+1}}{Q_{x_1,z}} - R \\
 &= N_{x_1,z}.
 \end{aligned}$$

Proof of (2.12) :

$$\begin{aligned}
 N_{0,z} &= \frac{Q_{0,z+1}}{Q_{0,z}} - R \\
 &= \frac{R^{z+1}}{R^z} - R && \text{using (2.7a)} \\
 &= 0.
 \end{aligned}$$

Proof of (2.13) :

As a first step in this proof it is necessary to prove the following identity

$$Q_{x_1, x_1+1} = \frac{x_1!}{2} (2R + x_1)(x_1 + 1), \quad x_1 \geq 0. \quad (2.14)$$

The proof of (2.14) is by induction :

Anchor : (2.14) is clearly true for $x_1 = 0$, since $Q_{0,1} = R$.

Assume true for $x_1 = k$, i.e. assume $Q_{k,k+1} = \frac{k!}{2} (2R + k)(k + 1)$.

Then

$$\begin{aligned} Q_{k+1, k+2} &= (k+1)Q_{k, k+1} + (R+k+1)Q_{k+1, k+1} && \text{using (2.7)} \\ &= (k+1)\frac{k!}{2}(2R+k)(k+1) + (R+k+1)(k+1)! && \text{using assumption} \\ & && \text{and (2.7a)} \\ &= \frac{(k+1)!}{2} [(2R+k)(k+1) + 2(R+k+1)] \\ &= \frac{(k+1)!}{2} [2Rk + 4R + k^2 + 3k + 2] \\ &= \frac{(k+1)!}{2} [(2R+k)(k+2) + k + 2] \\ &= \frac{(k+1)!}{2} [(2R+(k+1))((k+1)+1)]. \end{aligned}$$

This shows that, if (2.14) is true for $x_1 = k$, then it must also be true for $x_1 = k + 1$. Since it has been shown that (2.14) is true for $x_1 = 0$, it follows by induction that (2.14) holds for all $x_1 \geq 0$.

The proof of (2.13) may now be completed :

$$\begin{aligned} N_{x_1, x_1} &= \frac{Q_{x_1, x_1+1}}{Q_{x_1, x_1}} - R \\ &= \frac{\frac{x_1!}{2} (2R + x_1)(x_1 + 1)}{x_1!} - R && \text{using (2.14) and (2.7a)} \\ &= \frac{1}{2} (2R + x_1)(x_1 + 1) - R \\ &= \frac{1}{2} [2Rx_1 + 2R + x_1^2 + x_1 - 2R] \\ &= \frac{x_1}{2} [x_1 + 2R + 1]. \end{aligned}$$

§ 2.11 : A Comparison of All Four Estimators

In order to compare the performance of the four estimators which have so far been discussed, consideration is given to their mean, standard deviation and root mean square error conditional on the event $C = \{Z > X_1\}$. This conditioning is necessary since the maximum likelihood estimator \hat{N} yields infinite estimates when $Z = X_1$. It is important to note however that the Peterson-type estimator \hat{N}_p , harmonic mean estimator \hat{N}_h and conditionally unbiased estimator \hat{N}_u produce finite estimates with probability one. The unconditional performance of \hat{N}_p , \hat{N}_h and \hat{N}_u is considered later on.

Conditional on the event $C = \{Z > X_1\}$, the mean, standard deviation and root mean square error of each estimator are presented in tables 2.1a,b,c, 2.2a,b,c and 2.3a,b,c. These tables summarise the performance of the estimators for each combination from the following factorial design :

		0.1	
		0.2	
		0.3	0
10		0.4	5
$N = 25 \times \lambda = 1 \times \tau = \log\left(\frac{1}{1-\rho}\right) : \rho = 0.5 \times R = 10$		0.6	25
50		0.7	50
		0.8	
		0.9	

where $\rho = E\left(\frac{X_1}{N}\right)$.

Please note however that, for each value of population size N, only values of R up to and including N are considered; this is done for obvious practical reasons.

The notation used within each table is as follows :

Statistics

- exp. ≡ mean or expectation.
- s.d. ≡ standard deviation.
- rmse ≡ root mean square error.
- P(inf mle) ≡ $1 - \text{Prob}(C) = \text{Prob}(\bar{C}) = \text{Prob}(Z = X_1)$, which is the probability of the maximum likelihood estimator producing an infinite estimate.

Estimators

X1	≡	X ₁ ,	the number of distinct individuals seen from the target population.
P	≡	\hat{N}_p ,	the Peterson-type estimator of section 2.9.
CUE	≡	\hat{N}_u ,	the conditionally unbiased estimator of section 2.10.
MLE	≡	\hat{N} ,	the maximum likelihood estimator of section 2.7.
HME	≡	\hat{N}_h ,	the harmonic mean estimator, Joe & Reed(1985)

It is straightforward to obtain the distributions of both \hat{N} and \hat{N}_u given the event C. In order to obtain the conditional distribution of the Peterson-type estimator \hat{N}_p given the event C, one must derive the conditional distribution of X₁ and X₂ given C. The following proof was provided by I. B. J. Goudie (pers. com.).

Firstly recall that C is defined as being the event $\{Z>X_1\}$ and that \bar{C} is used to denote the complementary event $\{Z=X_1\}$.

\bar{C} occurs \Leftrightarrow X₂=0 and each individual in target population is seen at most once.

$$\text{Now } \text{Prob}(X_2 = 0) = [\exp(-\lambda\tau)]^R \tag{2.15}$$

(This is because we know that X₂ ~ Bin(R, 1 - exp(-λτ)))

Let Y_i = the number of sightings of individual i. It follows that Y_i ~ P(λτ).

It may then be observed that

Prob(each individual in target population is seen at most once)

$$\begin{aligned} &= \prod_{i=1}^N \text{Prob}(Y_i \leq 1) \\ &= [(\lambda\tau + 1)\exp(-\lambda\tau)]^N. \end{aligned} \tag{2.16}$$

Use of (2.15) and (2.16) implies that

$$\begin{aligned} \text{Prob}(C) &= 1 - \text{Prob}(\bar{C}) \\ &= 1 - [\exp(-\lambda\tau)]^R [(\lambda\tau + 1)\exp(-\lambda\tau)]^N. \end{aligned}$$

Now

$$\begin{aligned} \text{Prob}(X_1=x_1, X_2=x_2 | Z>X_1) &= \frac{\text{Prob}(X_1=x_1, X_2=x_2, Z>X_1)}{\text{Prob}(Z>X_1)} \\ &= \frac{\text{Prob}(X_1=x_1, X_2=x_2) \text{Prob}(Z>X_1 | X_1=x_1, X_2=x_2)}{\text{Prob}(Z>X_1)} \\ &= \frac{\text{Prob}(X_1=x_1) \text{Prob}(X_2=x_2) \text{Prob}(Z>X_1 | X_1=x_1, X_2=x_2)}{\text{Prob}(Z>X_1)} \end{aligned} \tag{2.17}$$

It is clear that $\text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2) = 1$ if $X_2 > 0$.

When $X_2 = 0$ it may be observed that $Z | X_1, X_2 \equiv Z_1 | X_1$. The distribution of $Z_1 | X_1$ may be characterised as being the sum of X_1 zero truncated Poisson random variables: this distribution is known in the literature as a Stirling distribution of the second kind; the probability function of this distribution is derived in appendix 2.

Explicitly the probability function of $Z_1 | X_1$ is given by

$$\text{Prob}(Z_1 = z_1 | X_1 = x_1) = \frac{x_1! (\lambda\tau)^{z_1} S(x_1, z_1)}{z_1! (\exp(\lambda\tau) - 1)^{x_1}},$$

$$z_1 = x_1, x_1 + 1, x_1 + 2, \dots$$

It follows that

$$\begin{aligned} \text{Prob}(Z > x_1 | X_1 = x_1, X_2 = x_2) &= \text{Prob}(Z_1 > x_1 | X_1 = x_1) \\ &= 1 - \text{Prob}(Z_1 = x_1 | X_1 = x_1) \\ &= 1 - \frac{(\lambda\tau)^{x_1}}{(\exp(\lambda\tau) - 1)^{x_1}}. \end{aligned}$$

Following on from equation (2.17), and using the notation $\tilde{P}(C) = \text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2)$, allows one to write :

$$\text{Prob}(X_1 = x_1, X_2 = x_2 | Z > X_1) = \frac{\text{Prob}(X_1 = x_1) \text{Prob}(X_2 = x_2) \text{Prob}(Z > X_1 | X_1 = x_1, X_2 = x_2)}{\text{Prob}(Z > X_1)}$$

$$= \frac{\binom{N}{x_1} [1 - \exp(-\lambda\tau)]^{x_1} [\exp(-\lambda\tau)]^{N-x_1} \binom{R}{x_2} [1 - \exp(-\lambda\tau)]^{x_2} [\exp(-\lambda\tau)]^{R-x_2} \tilde{P}(C)}{P(C)},$$

$$x_1 = \begin{cases} 0, 1, 2, \dots, N & \text{for } R > 0 \\ 1, 2, \dots, N & \text{for } R = 0 \end{cases}$$

$$x_2 = \begin{cases} 1, 2, \dots, R & \text{for } R > 0, x_1 = 0 \\ 0, 1, 2, \dots, R & \text{for } R > 0, x_1 > 0, \\ 0 & \text{for } R = 0 \end{cases}$$

$$\text{where } \tilde{P}(C) = \begin{cases} 1 & \text{for } X_2 > 0 \\ 1 - \frac{(\lambda\tau)^{x_1}}{(\exp(\lambda\tau) - 1)^{x_1}} & \text{for } X_2 = 0 \end{cases}$$

$$\text{and } P(C) = \text{Prob}(Z > X_1) = 1 - [\exp(-\lambda\tau)]^R [(\lambda\tau + 1)\exp(-\lambda\tau)]^N.$$

§ 2.11a : The c Constant

The Harmonic Mean Estimator, as defined by Joe & Reid(1985), depends upon a constant $c \in (0,1]$. There is no natural or obvious value that this constant should take. The problem of deciding which single value of c should be used in connection with the HME is therefore highly subjective. It is for this reason that the following tables illustrate the performance of the HME when used in connection with a range of possible c values . In practice however it would be necessary to have chosen a priori an appropriate value of c . Hence we now aim to recommend a single value of c that may be considered suitable for general use. This is not straightforward: in some situations the performance of the HME depends heavily upon the value taken by the c constant, whereas in other situations varying the constant has little effect. Joe & Reid(1985), although still concerned with population size estimation, considered a different problem to the one described here. It is worthwhile to note however that one conclusion they reached, bearing in mind that they only considered populations of size $N = 10$ when $R = 0$, was that changing c did not substantially affect the performance of the HME. The same conclusion may be drawn from tables 2.1a,b,c when R is equal to zero. So this, at least, is broadly consistent with Joe & Reid(1985). However when one considers situations in which $N = 10$, $R > 0$ and in particular those in which $N > 10$, $R \geq 0$ one can see that varying the c constant can in fact produce significant change in the performance of the HME, and that the effect of the constant is most significant for the smaller values of τ . For these small values of τ , which represent situations in which little information is available, the likelihood functions can be very spread out, almost flat and although remaining unimodal, are certainly not peaked, their shape representing the paucity of information. It is in each of these situations that the shape of the likelihood function can allow the harmonic mean estimate to differ greatly from the maximum likelihood estimate. The reason for this being that, due to the shape of the likelihood function, it is in each of these situations that the c constant has most effect upon the resulting harmonic mean estimate. Consequently, for small values of τ , the performance of the HME can be greatly affected by changing the c constant. For the larger values of τ , which represent situations in which a large amount of information has been obtained, the likelihood functions become peaked. In each of these situations the estimate produced by the HME is therefore not likely to differ greatly from the maximum likelihood estimate. This is the reason why, for the larger values of τ , the HME and MLE perform in a very similar way. And furthermore, due to the peakedness of the likelihood functions, the c constant has little affect upon each resulting harmonic mean estimate. Hence for the larger values of τ the c constant does not significantly affect the performance of the HME.

Regarding the way in which the c constant affects the HME when its performance is in turn being compared with that of the MLE, the general trend is that for small values of c the HME, when compared with the MLE, tends to have both a smaller mean and variance, and that as c is increased the mean and variance of the HME increase to those of the MLE. For the larger values of c , as one would expect, the performance of the HME is very similar to that of the MLE : in particular when $c = 1$ the HME and MLE are equivalent.

In view of this behaviour, when seeking to choose an appropriate value for the c constant, one must be careful not to place too much emphasis on mean square error. It would be inappropriate to base a choice of c solely upon the effect that this constant may have upon the mean square error of the HME, since this loss function is known to favour estimators possessing negative bias and small variance. Hence consideration of mean square error alone would lead one to choose an unduly small value for the c constant.

Based on an inspection of tables 2.1a,b,c, 2.2a,b,c and 2.3a,b,c it is believed that for general use the most appropriate value of the c constant should be $c = 0.5$. In reaching this decision consideration was given to mean, standard deviation and mean square error. Joe & Reed(1985) recommended use of the HME with a c value of 0.5. Hence from this point when reference is made to the HME a c value of 0.5 is to be assumed.

§ 2.11b : Discussion of Relative Performance of Estimators

The reasons underlying the differing performance of the MLE and HME have been discussed above. For the largest values of τ considered in the tables the MLE and HME perform in a very similar way. However for the smaller values of τ the HME can in some situations be seen to perform significantly better than the MLE : particularly in terms of mean square error, although it is important to note that this is often at the expense of negative bias. On the whole, within the range of population sizes covered in the tables, it can be said that the performance of the HME is either better than or as good as that of the MLE. Whilst it is important to bear in mind the fact that the HME tends to possess a comparatively small mean square error as a result of its negative bias, it is recommended that the HME be preferred to the MLE.

At this stage it is important to observe that, given the proportion of the population seen during the experiment, the performance and relative performance of the MLE, CUE and Peterson-type estimators can be seen to be essentially identical under both the discrete time sampling procedure of the previous chapter and the continuous

time sampling procedure of this chapter. Hence the conclusions reached in chapter 1 regarding the relative performance of the MLE, CUE and Peterson-type estimators are valid here. So that firstly, on the basis of tables 2.1,a,b,c, 2.2a,b,c and 2.3a,b,c, it can be argued that the CUE is clearly a better alternative to the Peterson-type estimator. And that secondly, for the reasons stated fully in chapter 1, one should always use the CUE in preference to the MLE.

The comparison between the HME and CUE is quite straightforward, this being a result of the fact that both estimators behave in a very similar way. When little information is available both the HME and CUE tend to be negatively biased with small variance, this bias then reduces smoothly as more information is obtained. In view of the fact that both the HME and MLE behave in a very similar way with regard to bias, it is worthwhile to note that in all but one of the one-hundred and eight situations considered in the tables, that the standard deviation of the CUE is smaller than that of the HME. For values of τ greater than 0.36, the absolute bias of the CUE is only greater than that of the HME in six out of the seventy-two situations considered, and in terms of mean square error the CUE is uniformly better. For values of τ less than or equal to 0.36, both the CUE and HME can be negatively biased, and in particular for these small values of τ the CUE can be more negatively biased than the HME. The standard deviation of the CUE however remains smaller than that of the HME, and consequently, for values of τ less than or equal to 0.36, the HME possesses a smaller mean square error than the CUE in only nine out of the thirty-six situations considered, and in none of these nine cases is the reduction in mean square error particularly large. In view of the above discussion, it is believed that one should always use the CUE in preference to the HME. This conclusion has been based entirely on consideration of performance : it is worthwhile to note however that for any given data set the estimate produced by the CUE is usually much easier to calculate than the harmonic mean estimate.

Table 2.1a

N = 10																
τ	R	1-P(C) x10 ⁴		X1	P	CUE	MLE	HME								
								0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.11	0	9452	exp.	1.91	1.91	2.25	2.48	3.78	3.77	3.21	2.80	2.89	2.85	2.86	2.67	2.48
			s.d.	0.91	0.91	1.60	2.07	1.79	1.81	1.54	1.86	2.11	2.00	2.04	2.36	2.06
			rmse	8.14	8.14	7.91	7.80	6.47	6.49	6.97	7.44	7.42	7.43	7.43	7.70	7.80
	5	5453	exp.	1.10	3.50	3.39	5.43	2.36	2.39	2.82	2.89	3.25	3.88	4.27	4.36	4.98
			s.d.	0.99	3.73	3.42	5.76	2.40	2.53	3.10	3.34	3.90	3.99	4.53	4.96	5.12
			rmse	8.95	7.49	7.44	7.35	8.01	8.02	7.83	7.86	7.80	7.31	7.30	7.51	7.17
	10	3146	exp.	1.06	5.40	5.27	8.34	2.60	3.17	3.35	4.12	4.42	5.61	6.03	6.81	7.52
			s.d.	0.97	5.61	5.25	9.24	2.92	3.89	4.31	4.81	5.50	6.24	7.01	7.60	8.33
			rmse	8.99	7.26	7.06	9.39	7.95	7.86	7.93	7.60	7.83	7.63	8.06	8.24	8.69
0.22	0	8094	exp.	2.70	2.70	3.71	4.31	5.25	5.25	4.61	4.39	4.69	4.58	4.61	4.63	4.29
			s.d.	1.18	1.18	2.49	3.40	2.31	2.40	2.49	2.68	3.14	3.04	3.16	3.49	3.37
			rmse	7.39	7.39	6.76	6.63	5.28	5.32	5.93	6.22	6.17	6.21	6.25	6.40	6.63
	5	2694	exp.	2.04	6.06	6.03	8.92	4.54	4.69	5.33	5.58	6.29	6.77	7.36	7.84	8.27
			s.d.	1.27	5.02	4.50	7.67	3.41	3.77	4.34	4.83	5.48	5.62	6.20	6.70	7.08
			rmse	8.06	6.38	6.00	7.75	6.44	6.52	6.37	6.55	6.62	6.49	6.74	7.04	7.29
	10	897	exp.	1.99	8.17	8.07	11.46	4.81	5.72	6.28	6.94	7.47	8.57	9.20	9.94	10.72
			s.d.	1.26	6.78	6.18	10.72	3.91	4.94	5.69	6.25	7.06	7.69	8.37	9.00	9.83
			rmse	8.11	7.02	6.47	10.82	6.50	6.53	6.80	6.96	7.50	7.83	8.41	9.00	9.86
0.36	0	5915	exp.	3.57	3.57	5.45	6.50	6.79	6.83	6.36	6.21	6.70	6.59	6.63	6.76	6.45
			s.d.	1.36	1.36	3.16	4.57	2.69	2.91	3.34	3.49	3.92	4.00	4.21	4.35	4.48
			rmse	6.57	6.57	5.54	5.75	4.19	4.30	4.94	5.15	5.13	5.26	5.39	5.43	5.72
	5	978	exp.	3.06	8.12	8.17	10.96	6.64	6.90	7.55	7.95	8.70	9.07	9.59	10.16	10.41
			s.d.	1.45	5.80	4.98	8.59	3.88	4.41	5.01	5.58	6.15	6.48	7.02	7.49	8.12
			rmse	7.09	6.10	5.30	8.64	5.13	5.39	5.58	5.95	6.28	6.55	7.03	7.49	8.13
	10	162	exp.	3.03	9.58	9.50	11.70	6.65	7.46	8.12	8.69	9.15	9.77	10.34	10.78	11.35
			s.d.	1.45	6.79	6.02	9.76	4.16	5.01	5.68	6.27	6.89	7.34	7.88	8.43	9.13
			rmse	7.12	6.81	6.04	9.91	5.34	5.61	5.99	6.41	6.94	7.34	7.88	8.46	9.23

Table 2.1b

N = 10																
τ	R	1-P(C) x10 ⁴		X1	P	CUE	MLE	HME								
								0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.51	0	3757	exp.	4.37	4.37	6.99	8.35	8.11	8.20	7.95	7.82	8.32	8.31	8.30	8.46	8.26
			s.d.	1.46	1.46	3.52	5.27	2.95	3.26	3.83	4.01	4.34	4.58	4.84	4.93	5.16
			rmse	5.82	5.82	4.63	5.52	3.51	3.73	4.35	4.57	4.65	4.88	5.13	5.16	5.45
	5	293	exp.	4.01	9.27	9.29	11.31	8.07	8.35	8.87	9.27	9.83	10.11	10.42	10.90	10.95
			s.d.	1.55	5.97	4.91	8.19	3.87	4.42	4.99	5.51	5.97	6.36	6.81	7.22	7.85
			rmse	6.18	6.01	4.96	8.30	4.32	4.72	5.12	5.55	5.97	6.36	6.82	7.27	7.90
	10	23	exp.	4.00	9.96	9.92	10.90	7.90	8.43	8.91	9.35	9.66	9.93	10.30	10.56	10.80
			s.d.	1.55	6.00	5.20	7.49	3.92	4.52	4.96	5.37	5.78	6.01	6.37	6.71	7.17
			rmse	6.20	6.00	5.20	7.55	4.45	4.78	5.07	5.41	5.79	6.01	6.37	6.74	7.22
0.69	0	1915	exp.	5.20	5.20	8.34	9.73	9.31	9.41	9.30	9.20	9.58	9.65	9.57	9.74	9.62
			s.d.	1.51	1.51	3.64	5.56	3.10	3.44	3.98	4.24	4.49	4.79	5.10	5.21	5.45
			rmse	5.03	5.03	4.00	5.57	3.17	3.49	4.05	4.31	4.51	4.81	5.12	5.22	5.47
	5	61	exp.	4.99	9.83	9.80	10.85	9.11	9.26	9.58	9.85	10.15	10.32	10.45	10.74	10.63
			s.d.	1.58	5.62	4.41	6.71	3.56	4.00	4.46	4.83	5.15	5.49	5.78	6.08	6.52
			rmse	5.25	5.62	4.42	6.76	3.67	4.07	4.47	4.83	5.15	5.50	5.80	6.12	6.55
	10	2	exp.	4.98	10.01	10.03	10.24	8.86	9.11	9.37	9.61	9.77	9.88	10.08	10.20	10.19
			s.d.	1.58	4.93	4.20	5.21	3.44	3.79	4.03	4.24	4.46	4.55	4.72	4.89	5.14
			rmse	5.26	4.93	4.20	5.22	3.62	3.90	4.07	4.26	4.46	4.55	4.72	4.90	5.15
0.92	0	688	exp.	6.11	6.11	9.33	10.37	10.32	10.32	10.21	10.13	10.31	10.39	10.24	10.37	10.27
			s.d.	1.51	1.51	3.45	5.30	3.05	3.35	3.78	4.09	4.29	4.57	4.88	5.02	5.22
			rmse	4.18	4.18	3.51	5.31	3.07	3.37	3.78	4.09	4.30	4.59	4.89	5.03	5.22
	5	7	exp.	6.02	10.02	9.98	10.24	9.87	9.80	9.90	10.01	10.10	10.15	10.19	10.23	10.09
			s.d.	1.55	4.86	3.58	4.69	3.02	3.31	3.58	3.80	3.94	4.19	4.27	4.47	4.68
			rmse	4.27	4.86	3.58	4.70	3.03	3.31	3.59	3.80	3.95	4.19	4.28	4.47	4.68
	10	0	exp.	6.01	10.03	10.03	9.87	9.57	9.58	9.65	9.75	9.81	9.81	9.91	9.90	9.75
			s.d.	1.55	3.91	3.25	3.59	2.82	3.03	3.16	3.27	3.33	3.39	3.42	3.53	3.65
			rmse	4.28	3.91	3.25	3.59	2.86	3.06	3.18	3.28	3.34	3.39	3.42	3.53	3.66

Table 2.1c

N = 10																
τ	R	1-P(C)		X1	P	CUE	MLE	HME								
		$\times 10^4$						0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.20	0	163	exp.	7.01	7.01	9.82	10.27	10.92	10.76	10.54	10.45	10.42	10.45	10.28	10.30	10.15
			s.d.	1.43	1.43	2.96	4.39	2.76	2.98	3.24	3.53	3.67	3.88	4.11	4.23	4.35
			rmse	3.31	3.31	2.97	4.39	2.91	3.07	3.29	3.56	3.69	3.90	4.12	4.24	4.36
5	0	0	exp.	6.99	10.06	10.02	9.87	10.33	10.10	10.02	10.03	10.02	9.96	9.96	9.83	9.66
			s.d.	1.45	3.96	2.72	3.11	2.43	2.56	2.72	2.83	2.89	3.02	2.99	3.13	3.17
			rmse	3.34	3.96	2.72	3.12	2.45	2.56	2.72	2.83	2.89	3.02	2.99	3.14	3.19
10	0	0	exp.	6.99	10.08	9.99	9.68	10.06	9.86	9.85	9.83	9.86	9.75	9.80	9.69	9.47
			s.d.	1.45	3.08	2.50	2.61	2.24	2.36	2.47	2.53	2.54	2.60	2.59	2.68	2.68
			rmse	3.34	3.08	2.50	2.63	2.24	2.37	2.47	2.54	2.55	2.61	2.59	2.70	2.73
1.61	0	15	exp.	8.00	8.00	9.98	9.82	11.11	10.79	10.49	10.32	10.15	10.12	9.95	9.85	9.68
			s.d.	1.26	1.26	2.17	2.84	2.19	2.27	2.37	2.55	2.58	2.71	2.81	2.87	2.87
			rmse	2.36	2.36	2.17	2.85	2.46	2.41	2.42	2.57	2.58	2.71	2.81	2.87	2.88
5	0	0	exp.	8.00	10.08	10.01	9.65	10.62	10.30	10.11	9.98	9.91	9.75	9.71	9.51	9.40
			s.d.	1.26	2.96	1.91	2.04	1.82	1.84	1.92	1.97	2.03	2.05	2.05	2.13	2.04
			rmse	2.37	2.96	1.91	2.07	1.93	1.86	1.92	1.97	2.04	2.07	2.07	2.19	2.13
10	0	0	exp.	8.00	10.09	9.98	9.57	10.46	10.15	9.99	9.86	9.80	9.63	9.64	9.46	9.33
			s.d.	1.26	2.25	1.82	1.84	1.70	1.72	1.79	1.82	1.89	1.89	1.92	1.99	1.84
			rmse	2.37	2.26	1.82	1.89	1.76	1.73	1.79	1.83	1.90	1.93	1.95	2.06	1.96
2.30	0	0	exp.	9.00	9.00	10.00	9.56	10.89	10.51	10.17	9.98	9.80	9.69	9.59	9.47	9.44
			s.d.	0.95	0.95	1.31	1.39	1.47	1.46	1.44	1.46	1.42	1.44	1.44	1.41	1.38
			rmse	1.38	1.38	1.31	1.46	1.72	1.55	1.45	1.46	1.44	1.48	1.50	1.51	1.49
5	0	0	exp.	9.00	10.06	10.03	9.51	10.64	10.30	10.06	9.82	9.70	9.54	9.42	9.29	9.35
			s.d.	0.95	1.92	1.21	1.25	1.25	1.27	1.27	1.26	1.29	1.24	1.25	1.19	1.18
			rmse	1.38	1.92	1.21	1.34	1.41	1.31	1.27	1.27	1.32	1.33	1.38	1.39	1.35
10	0	0	exp.	9.00	10.04	10.04	9.47	10.57	10.22	9.97	9.78	9.61	9.50	9.36	9.24	9.31
			s.d.	0.95	1.46	1.15	1.21	1.21	1.20	1.20	1.22	1.23	1.22	1.20	1.15	1.15
			rmse	1.38	1.46	1.15	1.32	1.34	1.22	1.20	1.24	1.29	1.32	1.36	1.38	1.34

Table 2.4a

N = 10															
τ	R	$P(Z < -1)$ $\times 10^4$		X1	P	CUE	HME								
							0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.11	0	6990	exp.	2.24	2.24	3.69	4.57	4.65	4.99	5.36	5.91	7.72	10.47	14.27	28.11
			s.d.	0.67	0.67	2.04	1.65	1.98	2.70	3.61	4.73	5.71	8.41	13.46	27.75
			rmse	7.78	7.78	6.63	5.68	5.71	5.69	5.87	6.25	6.15	8.42	14.12	33.14
	5	5089	exp.	1.69	7.26	7.65	5.08	6.03	8.37	9.96	13.14	17.80	24.50	37.68	76.69
			s.d.	0.91	5.56	6.34	4.25	5.94	8.76	11.41	15.98	22.27	32.73	54.40	118.6
			rmse	8.36	6.20	6.77	6.50	7.14	8.91	11.41	16.28	23.60	35.80	61.04	136.0
	10	3546	exp.	1.43	8.97	9.05	4.71	6.49	7.93	10.27	12.76	17.31	23.15	34.51	66.84
			s.d.	0.95	8.47	8.86	5.36	8.41	11.76	15.60	21.47	29.89	44.41	72.94	159.2
			rmse	8.63	8.53	8.91	7.53	9.11	11.94	15.60	21.65	30.77	46.32	76.95	169.1
0.22	0	3546	exp.	2.68	2.68	4.82	5.61	5.85	6.39	7.10	8.20	10.16	13.72	19.36	37.62
			s.d.	0.96	0.96	3.11	2.58	3.22	4.18	5.43	7.17	9.16	13.26	21.29	44.60
			rmse	7.38	7.38	6.05	5.09	5.25	5.52	6.16	7.39	9.16	13.77	23.26	52.46
	5	1586	exp.	2.25	8.15	8.59	6.18	7.10	9.04	10.48	13.21	16.65	21.93	32.05	61.21
			s.d.	1.16	6.20	6.93	4.93	6.81	9.51	12.43	17.04	23.71	34.65	56.96	124.3
			rmse	7.84	6.47	7.07	6.23	7.41	9.56	12.44	17.34	24.62	36.65	61.08	134.4
	10	663	exp.	2.09	9.55	9.59	5.72	7.19	8.39	9.81	11.42	14.14	17.44	23.47	40.06
			s.d.	1.22	8.37	8.45	5.42	7.86	10.63	13.85	18.61	25.46	37.16	60.53	131.1
			rmse	8.01	8.38	8.46	6.91	8.35	10.75	13.86	18.67	25.79	37.89	62.01	134.5
0.36	0	1257	exp.	3.35	3.35	6.36	7.06	7.51	8.17	9.11	10.70	12.81	16.80	23.78	44.97
			s.d.	1.25	1.25	4.12	3.37	4.32	5.57	7.23	9.49	12.60	18.08	29.20	62.03
			rmse	6.77	6.77	5.50	4.47	4.99	5.86	7.28	9.51	12.90	19.32	32.29	71.21
	5	289	exp.	3.10	9.04	9.41	7.43	8.15	9.47	10.55	12.39	14.37	17.47	23.26	39.19
			s.d.	1.41	6.53	6.81	5.02	6.75	9.02	11.66	15.59	21.43	31.02	50.43	109.5
			rmse	7.05	6.60	6.83	5.64	7.00	9.03	11.68	15.77	21.87	31.90	52.14	113.3
	10	61	exp.	3.04	9.92	9.90	6.90	7.87	8.71	9.50	10.27	11.36	12.69	14.65	19.76
			s.d.	1.44	7.44	7.04	4.82	6.34	8.02	9.99	12.77	16.80	23.74	37.82	80.56
			rmse	7.11	7.44	7.04	5.73	6.69	8.12	10.00	12.78	16.86	23.89	38.11	81.15

Table 2.4b

N = 10															
τ	R	P(Z<=1) x10 ⁴		X1	P	CUE	HME								
							0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.51	0	372	exp.	4.12	4.12	7.82	8.46	9.02	9.67	10.59	12.28	14.27	17.87	24.63	44.53
			s.d.	1.44	1.44	4.68	3.77	4.92	6.36	8.30	10.81	14.63	21.02	34.09	73.01
			rmse	6.06	6.06	5.16	4.08	5.01	6.37	8.32	11.05	15.24	22.45	37.10	80.76
	5	41	exp.	4.01	9.62	9.79	8.39	8.87	9.66	10.35	11.35	12.30	13.66	16.26	22.69
			s.d.	1.54	6.35	6.01	4.55	5.85	7.46	9.38	12.13	16.25	23.10	36.95	79.34
			rmse	6.18	6.36	6.02	4.83	5.96	7.47	9.39	12.20	16.41	23.38	37.48	80.35
	10	4	exp.	4.00	10.02	10.00	7.94	8.51	9.03	9.51	9.88	10.24	10.76	11.30	12.42
			s.d.	1.55	6.18	5.52	4.11	4.93	5.70	6.61	7.84	9.54	12.64	19.12	39.35
			rmse	6.20	6.18	5.52	4.60	5.14	5.78	6.63	7.85	9.55	12.67	19.17	39.43
0.69	0	80	exp.	5.02	5.02	9.02	9.68	10.18	10.67	11.34	12.65	14.18	16.63	21.64	36.03
			s.d.	1.54	1.54	4.75	3.83	4.99	6.43	8.39	10.85	14.74	21.19	34.33	73.64
			rmse	5.22	5.22	4.85	3.84	5.00	6.47	8.50	11.17	15.32	22.20	36.25	78.10
	5	4	exp.	4.99	9.92	9.94	9.20	9.40	9.80	10.15	10.57	10.91	11.32	12.18	13.76
			s.d.	1.58	5.76	4.87	3.84	4.61	5.51	6.57	8.01	10.24	13.95	21.63	45.43
			rmse	5.26	5.76	4.87	3.92	4.65	5.52	6.57	8.03	10.28	14.02	21.74	45.59
	10	0	exp.	4.98	10.02	10.04	8.86	9.12	9.38	9.62	9.80	9.91	10.13	10.28	10.36
			s.d.	1.58	4.96	4.25	3.47	3.86	4.15	4.47	4.85	5.27	6.12	8.03	14.64
			rmse	5.26	4.96	4.25	3.65	3.96	4.20	4.48	4.85	5.27	6.12	8.03	14.64
0.92	0	10	exp.	6.02	6.02	9.74	10.57	10.80	10.97	11.28	11.94	12.74	13.82	16.34	23.36
			s.d.	1.54	1.54	4.27	3.54	4.48	5.62	7.21	9.15	12.29	17.49	28.06	59.85
			rmse	4.27	4.27	4.27	3.59	4.55	5.70	7.33	9.35	12.59	17.90	28.77	61.32
	5	0	exp.	6.01	10.03	10.00	9.88	9.82	9.93	10.06	10.16	10.23	10.32	10.44	10.53
			s.d.	1.55	4.89	3.68	3.08	3.45	3.84	4.23	4.70	5.50	6.76	9.62	18.94
			rmse	4.28	4.89	3.68	3.09	3.45	3.84	4.23	4.70	5.51	6.76	9.63	18.95
	10	0	exp.	6.01	10.03	10.03	9.57	9.58	9.65	9.75	9.81	9.82	9.92	9.90	9.76
			s.d.	1.55	3.91	3.26	2.83	3.03	3.17	3.29	3.36	3.44	3.54	3.81	4.78
			rmse	4.28	3.91	3.26	2.86	3.06	3.19	3.30	3.36	3.44	3.54	3.81	4.79

Table 2.4c

N = 10															
τ	R	$P(Z \leq -1)$		X1	P	CUE	HME								
		$\times 10^4$					0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.20	0	1	exp.	6.99	6.99	9.97	11.02	10.93	10.80	10.83	10.95	11.21	11.41	12.17	14.21
			s.d.	1.45	1.45	3.38	2.99	3.54	4.20	5.17	6.28	8.16	11.29	17.69	37.20
			rmse	3.34	3.34	3.38	3.15	3.66	4.28	5.24	6.36	8.25	11.37	17.83	37.44
	5	0	exp.	6.99	10.06	10.02	10.33	10.10	10.02	10.04	10.02	9.97	9.97	9.85	9.69
			s.d.	1.45	3.96	2.73	2.44	2.57	2.75	2.88	2.98	3.19	3.34	3.96	6.09
			rmse	3.34	3.96	2.73	2.46	2.58	2.75	2.88	2.98	3.19	3.34	3.96	6.10
	10	0	exp.	6.99	10.08	9.99	10.06	9.86	9.85	9.83	9.86	9.75	9.80	9.69	9.47
			s.d.	1.45	3.08	2.50	2.24	2.36	2.47	2.53	2.55	2.60	2.59	2.69	2.71
			rmse	3.34	3.08	2.50	2.24	2.37	2.47	2.54	2.55	2.61	2.60	2.70	2.77
1.61	0	0	exp.	8.00	8.00	10.00	11.12	10.82	10.52	10.37	10.22	10.21	10.09	10.07	10.16
			s.d.	1.26	1.26	2.26	2.24	2.40	2.59	2.94	3.25	3.86	4.89	7.12	14.23
			rmse	2.37	2.37	2.26	2.50	2.53	2.64	2.97	3.25	3.87	4.89	7.12	14.23
	5	0	exp.	8.00	10.08	10.01	10.62	10.30	10.11	9.98	9.91	9.75	9.71	9.51	9.40
			s.d.	1.26	2.96	1.91	1.82	1.84	1.92	1.97	2.04	2.06	2.06	2.15	2.15
			rmse	2.37	2.96	1.91	1.93	1.86	1.92	1.97	2.04	2.07	2.08	2.21	2.23
	10	0	exp.	8.00	10.09	9.98	10.46	10.15	9.99	9.86	9.80	9.63	9.64	9.46	9.33
			s.d.	1.26	2.25	1.82	1.70	1.72	1.79	1.82	1.89	1.89	1.92	1.99	1.84
			rmse	2.37	2.26	1.82	1.76	1.73	1.79	1.83	1.90	1.93	1.95	2.06	1.96
2.30	0	0	exp.	9.00	9.00	10.00	10.89	10.51	10.17	9.99	9.80	9.69	9.59	9.47	9.45
			s.d.	0.95	0.95	1.31	1.47	1.46	1.44	1.47	1.45	1.49	1.53	1.64	2.25
			rmse	1.38	1.38	1.31	1.72	1.55	1.45	1.47	1.46	1.52	1.59	1.72	2.31
	5	0	exp.	9.00	10.06	10.03	10.64	10.30	10.06	9.82	9.70	9.54	9.42	9.29	9.35
			s.d.	0.95	1.92	1.21	1.25	1.27	1.27	1.26	1.29	1.24	1.25	1.19	1.18
			rmse	1.38	1.92	1.21	1.41	1.31	1.27	1.27	1.32	1.33	1.38	1.39	1.35
	10	0	exp.	9.00	10.04	10.04	10.57	10.22	9.97	9.78	9.61	9.50	9.36	9.24	9.31
			s.d.	0.95	1.46	1.15	1.21	1.20	1.20	1.22	1.23	1.22	1.20	1.15	1.15
			rmse	1.38	1.46	1.15	1.34	1.22	1.20	1.24	1.29	1.32	1.36	1.38	1.34

§ 2.13 : The Performance of Plant-Capture

When Applied to the Model M_0 : Under Continuous Time Sampling

It has been argued that, conditional on the event $C = \{Z > X_1\}$, the performance of the CUE should be considered superior to that of the HME, and that the performance of the HME should in turn be considered superior to that of the MLE and Peterson-type estimators. In addition to this the distributional properties of the CUE, HME and Peterson-type estimators were investigated conditional on the less restrictive event $A = \{Z > 1\}$. The conclusion drawn from the latter investigation was that, conditional on the event $A = \{Z > 1\}$, one should again consider the CUE to be the estimator whose performance is most desirable. Rather than only discussing the way in which the information gained through plants may improve the performance of the CUE, this section considers how the method of plant-capture can affect the performance of all the estimators described within this chapter. This approach is taken since, in spite of the evidence of the previous sections, it is believed that more traditionally minded practitioners may still prefer to use the MLE. The following discussion is based on an inspection of all the 18 tables of this chapter.

It has previously been mentioned that mean square error is known to reward negative bias and that this characteristic can lead to incorrect conclusions being drawn if one places too much emphasis on mean square error alone. When comparing the performance of estimators, one should always, where possible, consider firstly their mean and standard deviation, and only then should one consider mean square error, or alternative loss functions such as mean absolute deviation. This approach is taken in the following discussion because consideration of mean square error alone can lead to counter intuitive conclusions. For example, consider the performance of the CUE in table 2.2a, where $N = 25$ and $\tau = 0.36$. As R is increased from 0 to 5 to 10 the root mean square error of \hat{N}_u increases from 11.21 to 11.33 to 11.98! However, only when one considers the way in which the bias of \hat{N}_u is being significantly reduced can one see that the extra information gained from the plants is in fact improving the overall performance of the CUE.

This last example is quite typical of the way in which the information gained from plants enhances the performance of the estimators in situations where very little information is gained from the target population. In many of these situations, however, the improvement in bias is accompanied by a reduction in mean square error.

Except for a very few situations where only a small amount of information is available, the CUE \hat{N}_u is usually unbiased, and where not its bias is negligible. In those situations where only a small amount of information is available, \hat{N}_u tends to be negatively biased, with this bias reducing appreciably and monotonically as the number of plants is increased. This behaviour is intuitively very reasonable, since the CUE is

unbiased conditional on the event $Z = Z_1 + Z_2 \geq N$. Now as $Z_2 \sim P(R\lambda\tau)$, the event $Z = Z_1 + Z_2 \geq N$ is more and more likely to occur as R is increased. The standard deviation of \hat{N}_u is generally seen to reduce monotonically as more and more plants are used. Where the standard deviation of the CUE is not reduced by an increase in R , this is always due to its bias being significantly improved.

In the absence of plants, the Peterson-type estimator reduces to X_1 , commonly referred to as the 'enumeration estimator'. For this reason, \hat{N}_p is only considered here when plants are used. In terms of bias, the Peterson-type estimator behaves in a very similar way to the CUE, although \hat{N}_u is on almost all occasions less biased. When the number of plants is small relative to the size of the target population, \hat{N}_p tends to have a relatively large variance. On almost all occasions, however, this is seen to reduce uniformly as R is increased.

The CUE is seen to utilize the information gained through the plants in a very 'smooth' way. That is, as R is increased, usually either the mean of \hat{N}_u is significantly improved at the expense of a slight increase in standard deviation or both the bias and the standard deviation are reduced. The behaviour of the MLE in situations where little information is available is less predictable as more plants are introduced. For an example consider again table 2.2a, in which $N = 25$ and $\tau=0.36$. As the number of plants is increased from 0 to 25 the mean and standard deviation of \hat{N} are respectively 23.62, 14.35; 29.56, 19.81; 29.54, 19.99 and 26.66, 12.94 : the corresponding values taken by \hat{N}_u are respectively 17.36, 8.21; 22.49, 11.05; 24.39, 11.96 and 24.95, 10.64.. The CUE in this example behaves in the manner described above, i.e. as R is increased its' performance improves 'smoothly'. However, as the number of plants is increased from 0 to 5 to 10, the standard deviation of the MLE is seen to increase, whilst its mean either becomes worse or is not significantly improved ! This result appears counter intuitive, that is until one considers the way in which the value of R affects the probability of obtaining a finite MLE. In the above situation, wherein $N = 25$ and $\tau=0.36$, for $R = 0, 5, 10$ and 25 , the probabilities of obtaining an infinite MLE are respectively 0.2690, 0.0445, 0.0074 and 0.0000. In other words the introduction of plants is seen to drastically improve the probability of obtaining a useful MLE. When this advantage is considered along with the performance of the MLE, it can be argued that even in situations where, as in the above example, very little information is obtained from the target population, the presence of plants is beneficial to the overall performance of the MLE. Other than those extreme situations in which very little information is present, an increase in the number of plants is generally seen to improve the performance of the MLE via a reduction in both bias and standard deviation. And where both statistics are not improved, one of the two is.

In conclusion, the introduction of plants can be seen to enhance the performance of all four of the estimators which have been considered within this chapter, this being under the assumption that the planted individuals do indeed behave in an identical manner to members of the target population. In particular the plants are seen to be of most use when only little information has been gained from the target population. Furthermore, whether sampling with or without plants, it is recommended that, conditional on the event $C = \{Z > X_1\}$, the performance of the CUE should be considered superior to that of the HME, MLE and Peterson-type estimators.

Chapter 3 : Estimation Under the Capture-Recapture Model M_h : Discrete Time Sampling Procedure.

§3.1 : Introduction

This chapter introduces, initially in the absence of plants, a new class of estimators for the standard capture-recapture model M_h . The model M_h is one of the set of models discussed in the wildlife monograph by Otis et al. (1978) for capture-recapture data in closed populations. In section 3.7 it is shown how these new estimators can be modified so as to utilize the information gained from planted individuals.

In the previous two chapters consideration was given to a capture-recapture model which assumed that each animal in the target population was equally likely to be caught. Indeed in the large statistical literature on capture-recapture methods the majority of work present is seen to adopt this central assumption, see Seber(1982). However, although the assumption that every animal in the population is equally likely to be caught is convenient from a mathematical point of view, in practice this will rarely be the case. In particular if the population under investigation possesses significant heterogeneity between capture probabilities then, in experiments where the true population size is known, the usual estimators, for example the maximum likelihood estimators as described in chapters 1 and 2, have been shown to become extremely negatively biased, see Edwards & Eberhardt(1973) or Cormack(1966).

The discrete time sampling procedure considered within this chapter is identical to the one considered in chapter 1 : it essentially constitutes what is known in the literature as a Schnabel Census with random sample sizes, see Schnabel(1938) or, for a more comprehensive review, Seber(1982). The sampling procedure considered here is the one most commonly used in practice, and is described in detail in the following section.

§3.2 : Sampling Procedure and Assumptions

A sequence of t sampling experiments is carried out on the target population which is assumed to be closed and of size N . Independently of other animals and independently of its previous capture history animal i ($i = 1, 2, \dots, N$) is captured in sample j ($j = 1, 2, \dots, t$)

with probability p_i . After each sample is taken every animal within that sample which has not previously been marked receives a unique tag before its immediate release so that it may be recognised on subsequent trapping occasions. The experiment generates an N by t matrix A where

$$a_{ij} = \begin{cases} 1 & \text{if animal } i \text{ is caught on sampling occasion } j \\ 0 & \text{if animal } i \text{ is not caught on sampling occasion } j \end{cases}$$

$$i = 1, 2, \dots, N.$$

$$j = 1, 2, \dots, t.$$

The sample space is the set of such matrices.

It is assumed that the p_i , for $i = 1, 2, \dots, N$, are a random sample from some probability distribution $f(p)$, $p \in [0,1]$, with c.d.f. $F(p)$.

§3.3 : A Class of Coverage Adjusted Estimators for the Model M_h

Probabilistic results for the model M_h are covered in detail by Otis et al (1978), Overton(1969) and Pollock & Otto(1983). The following derivation of an approximate maximum likelihood estimator borrows heavily from this previous work : the approach taken here is almost identical to that of Overton(1969) and Pollock & Otto(1983). At this point it is necessary to introduce some notation :

- t \equiv number of sampling occasions.
- X \equiv number of distinct animals seen.
- Z \equiv total number of sightings made.
- p_i \equiv capture probability of animal i , $i = 1, 2, \dots, N$.
- X_i \equiv number of sightings of the i^{th} animal, $i = 1, 2, \dots, N$.

$$f_k = \sum_{i=1}^N I(X_i = k) \quad \equiv \quad \text{number of animals seen exactly } k \text{ times, } k=0, 1, 2, \dots, t.$$

$$C = \frac{\sum_{i=1}^N p_i I(X_i > 0)}{\sum_{i=1}^N p_i} \quad \equiv \quad \text{' Sample Coverage ' .}$$

Let the set $S_x = \{s_k : k = 1, 2, \dots, x\}$, where $s_k \in \{1, 2, 3, \dots, N\}$ for all k , denote the set of the indexes of the x distinct animals seen during the sampling period.

Under this model all of the available information is contained within the vector of capture frequencies (f_1, f_2, \dots, f_t) , see Pollock et al (1990).

The joint probability distribution of the sufficient statistics, $\{f_i; i=1,2,\dots,t\}$, is multinomial:

$$\text{Prob}(f_1, f_2, \dots, f_t | F) = \binom{N}{N-x, f_1, f_2, \dots, f_t} (\pi_0)^{N-x} \prod_{j=1}^t (\pi_j)^{f_j} \quad (3.1)$$

$$\text{where } \pi_j = \int_0^1 \binom{t}{j} p^j (1-p)^{t-j} dF(p)$$

$$j = 0, 1, 2, \dots, t.$$

From equation (3.1), assuming that $F(p)$ is known exactly, the profile likelihood for N can be written as

$$\begin{aligned} L(N) &= \frac{N!}{(N-x)!} \{\pi_0\}^{N-x} \\ &= \frac{N!}{(N-x)!} \left\{ E[(1-p)^t] \right\}^{N-x}, \end{aligned} \quad (3.2)$$

$$\text{since } \pi_0 = \int_0^1 (1-p)^t dF(p) = E[(1-p)^t].$$

If one again assumes that $F(p)$ is known exactly, so that $E[(1-p)^t]$ may be viewed as a known constant, it follows from equation (3.2) that an approximate maximum likelihood estimate may be obtained by equating $L(N)$ to $L(N-1)$:

$$\begin{aligned} L(N) = L(N-1) &\Leftrightarrow \frac{N!}{(N-x)!} \left\{ E[(1-p)^t] \right\}^{N-x} = \frac{(N-1)!}{(N-1-x)!} \left\{ E[(1-p)^t] \right\}^{N-1-x} \\ &\Leftrightarrow N \left\{ E[(1-p)^t] \right\} = (N-x) \\ &\Leftrightarrow N - N \left\{ E[(1-p)^t] \right\} = x \\ &\Leftrightarrow N = \frac{x}{1 - \left\{ E[(1-p)^t] \right\}}. \end{aligned} \quad (3.3)$$

This expression is of no direct use in practice since $E[(1-p)^t]$ is not known. However an estimate of this quantity may be obtained: Overton(1969) used a method based on a theorem of Horvitz & Thompson(1952); Pollock & Otto(1983) obtained exactly the same result using the theory of weighted distributions as follows:

$$\begin{aligned} E[(1-p)^t] &= \pi_0 = \int_0^1 (1-p)^t dF(p) \\ &= \int_0^1 (1-p)^t f(p) dp. \end{aligned}$$

Let $f^w(p)$ be the probability density function of the capture probabilities of all x animals captured during the experiment. When $f^w(p)$ is derived from $f(p)$ as a weighted

distribution, see Patil & Rao(1978), with weight $w(p)=1-(1-p)^t$, the probability of capture at least once,

$$f^w(p) = \frac{w(p)f(p)}{\mu} \\ = \frac{\{1-(1-p)^t\}f(p)}{1-E[(1-p)^t]}.$$

Now one may observe that, using the properties of weighted distributions, an unbiased estimator of $\{1-E[(1-p)^t]\}^{-1}$ is given by $\frac{1}{x} \sum_{i \in S_x} \frac{1}{1-(1-p_i)^t}$.

It then follows from equation (3.3) that

$$\hat{N} = \sum_{i \in S_x} \frac{1}{1-(1-p_i)^t} \tag{3.4}$$

would be an unbiased estimator of population size N if the capture probabilities of the animals seen during the experiment were known exactly. At this point it is also worthwhile to note that, since the probability of animal i being seen at least once during the experiment is $1-(1-p_i)^t$ for $i = 1, 2, \dots, N$, taking an expectation of (3.4), when the p_i for $i \in S_x$ are known, would show \hat{N} to be an unbiased estimator of N . However since these capture probabilities are clearly not known exactly the approach taken here is to estimate the p_i and in doing so obtain an estimator of N by substituting these estimates of capture probability into equation (3.4).

It is now necessary to estimate the capture probabilities of the animals which were seen during the experiment :

Overton(1969) used the fact that under the model $X_i \sim \text{Bin}(t, p_i)$. Based on this distribution the maximum likelihood estimate of the capture probability of animal i is given by $\hat{p}_i^{(1)} = \frac{X_i}{t}$. Overton(1969) then substituted the estimates $\hat{p}_i^{(1)}$ into equation (3.4) to produce the estimator \hat{N}_0 , defined by

$$\hat{N}_0 = \sum_{i \in S_x} \frac{1}{1-[1-\hat{p}_i^{(1)}]^t} \\ = \sum_{i \in S_x} \frac{1}{1-\left(1-\frac{X_i}{t}\right)^t} \\ = \sum_{i=1}^t \frac{f_i}{1-\left(1-\frac{i}{t}\right)^t}.$$

The estimates $\hat{p}_i^{(1)}$ are intuitively reasonable estimates of capture probability - essentially being 'the number of times the animal was seen divided by the number of times the animal could have been seen'. However this method of estimating capture probability does not make full use of all of the available information. In order to obtain better estimates of capture probability one may proceed as follows :

In addition to estimating the capture probability of each animal in turn, via the $\hat{p}_i^{(1)}$, we are also able to obtain independently an estimate of the sum of the capture probabilities of the animals seen during sampling. The approach taken here is to use this latter estimate to scale the $\hat{p}_i^{(1)}$ in an appropriate manner :

Let

$$\hat{p}_i^{(2)} \propto \frac{X_i}{t} \Rightarrow \hat{p}_i^{(2)} = k \frac{X_i}{t}, \quad \text{where } k = \text{constant.} \quad (3.5)$$

It then follows from equation (3.5) that

$$\sum_{i \in S_x} \hat{p}_i^{(2)} = \sum_{i \in S_x} \frac{k \cdot X_i}{t} = \frac{k}{t} \sum_{i \in S_x} X_i = k \frac{Z}{t}. \quad (3.6)$$

For the reasons stated above we now set

$$\sum_{i \in S_x} \hat{p}_i^{(2)} = \sum_{i \in S_x} p_i = \sum_{i=1}^N p_i I(X_i > 0) = C \sum_{i=1}^N p_i, \quad \text{where } C = \frac{\sum_{i=1}^N p_i I(X_i > 0)}{\sum_{i=1}^N p_i}. \quad (3.7)$$

From assumptions,

$$\begin{aligned} X_i &\sim \text{Bin}(t, p_i) \Rightarrow E[X_i] = t \cdot p_i \\ &\Rightarrow E\left[\sum_{i=1}^N X_i\right] = E[Z] = t \sum_{i=1}^N p_i \end{aligned} \quad (3.8)$$

\Rightarrow an estimate of $\sum_{i=1}^N p_i$ is given by $\frac{Z}{t}$. The value of sample coverage C may be estimated by \hat{C}_1 , \hat{C}_2 or \hat{C}_3 : please refer to appendix 4.

Now from equation (3.7) we require that $\sum_{i \in S_x} \hat{p}_i^{(2)} = \hat{C}_j \frac{Z}{t}$, $j = 1, 2 \text{ or } 3$. (3.9)

Combining equations (3.6) and (3.9) enables one to determine the value of the constant k as follows :

$$\sum_{i \in S_x} \hat{p}_i^{(2)} = k \frac{Z}{t} = \hat{C}_j \frac{Z}{t} \Rightarrow k = \hat{C}_j, \quad j = 1, 2 \text{ or } 3,$$

$$\Rightarrow \text{Estimate } p_i \text{ by } \hat{p}_{i,j}^{(2)} = \hat{C}_j \frac{X_i}{t}, \quad i \in S_x, \quad j = 1, 2 \text{ or } 3.$$

Explicitly

$$\begin{aligned}\hat{p}_{i,1}^{(2)} &= \hat{C}_1 \frac{x_i}{t} = \left(1 - \frac{f_1}{z}\right) \frac{x_i}{t}, \\ \hat{p}_{i,2}^{(2)} &= \hat{C}_2 \frac{x_i}{t} = \left(1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z}\right) \frac{x_i}{t} \\ \text{and } \hat{p}_{i,3}^{(2)} &= \hat{C}_3 \frac{x_i}{t} = \left(1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z} - \frac{6}{(t-1)(t-2)} \frac{f_3}{z}\right) \frac{x_i}{t}.\end{aligned}$$

These three estimates of capture probability may each be substituted into equation (3.4) to produce a corresponding Coverage Adjusted Estimator (CAE):

$$\begin{aligned}\hat{N}_{ca1} &= \sum_{i \in S_x} \frac{1}{1 - [1 - \hat{p}_{i,1}^{(2)}]^t} \\ &= \sum_{i \in S_x} \frac{1}{1 - \left(1 - \left(1 - \frac{f_1}{z}\right) \frac{x_i}{t}\right)^t} \\ &= \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \left(1 - \frac{f_1}{z}\right) \frac{i}{t}\right)^t}.\end{aligned}$$

Similarly

$$\hat{N}_{ca2} = \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \left(1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z}\right) \frac{i}{t}\right)^t}$$

and

$$\hat{N}_{ca3} = \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \left(1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z} - \frac{6}{(t-1)(t-2)} \frac{f_3}{z}\right) \frac{i}{t}\right)^t}.$$

§3.4 : Other Estimators for the Model M_h

Within the introduction to this chapter it was stated that the majority of capture-recapture work has been based on the assumption that capture probabilities are equal for all animals in the population being trapped. This is true. However, over the years, a number of authors have considered heterogeneous populations. Overton(1969) introduced the above Horvitz-Thompson type estimator which allows capture probabilities to vary between animals. The performance of this estimator was, however, not considered to be reasonable,

so that, up to this date and beyond, a viable estimation procedure for heterogeneous populations was still not available. Burnham and Overton(1978) sought to rectify this less than ideal situation by introducing a nonparametric jackknife estimator of population size, aimed at heterogeneous populations. The Burnham and Overton(1978) paper stimulated interest in the topic and since its appearance many other authors have proposed estimators for the model M_h . Pollock and Otto(1983) considered a first order jackknife of the estimator proposed by Overton(1969). Smith and van Belle(1984) considered bootstrapping based on the enumeration estimator. Chao(1989), Chao, Lee and Jeng(1992) and, more recently, Norris and Pollock(1996a) have also proposed estimators for the model M_h .

To date, the estimators most commonly favoured have been the jackknife estimators of Burnham and Overton(1978). Lee and Chao(1994), however, assert that the estimators of Chao, Lee and Jeng(1992) are to be preferred, except when the heterogeneity is very mild. In the latter case they recommend the maximum likelihood estimator for the model M_0 , details of which may be found in chapter 1 of this thesis.

§3.5 : Simulation Study

A simulation study was carried out in order to investigate the properties of each estimator. In each simulation the capture probabilities of the N animals were drawn as a random sample from some probability distribution with mean $E(p)$, variance $\text{Var}(p)$ and coefficient of variation $\sqrt{\text{Var}(p)}/E(p)$. Live trapping was then simulated on this population. Each table consists of six cells, with each cell depicting the results for one of $t = 5, 10, 15, 20, 25$ or 30 sampling occasions. For each value of t one thousand simulations were carried out : a different set of capture probabilities was used each time. The values shown in the tables are mostly averages. As many of the estimators are only finite if at least one recapture occurs, any data set not meeting this condition was discarded. The simulation procedure continued until one thousand data sets for which the condition did hold had been generated.

In tables 3.5.1a, 3.5.1b and 3.5.1c the capture probabilities of the animals were drawn from a uniform distribution on the interval $(0, \alpha)$, symbolised by $p \sim U(0, \alpha)$.

In tables 3.5.2a to 3.5.2i the distribution considered is Beta : symbolised by $p \sim \text{Beta}(\alpha, \beta)$. The most comprehensive simulation study to appear in the literature to date was carried out by Burnham and Overton(1979), and they considered mainly Beta distributions. It appeared that they essentially varied the parameters of each Beta

distribution in such a way so as to achieve particular values for the expectation of p . That is prime consideration was given to $E(p)$. It has, however, been known for many years, Cormack(1966) and Edwards and Eberhardt(1973), that the performance of the estimators is dependent not only on the mean but also the coefficient of variation of the distribution of p . Hence within this simulation study prime consideration is given jointly to both the mean and the coefficient of variation of each Beta distribution; the parameters of the distribution leading to these values are considered to be of only secondary importance. It was therefore decided to choose the parameters of each Beta distribution systematically in such a way so as to investigate the dependence of the estimators performance on both mean and coefficient of variation. In practice it is believed that if the model M_h is chosen as an appropriate model to fit the data, using for example the testing procedures described by Otis et al.(1978), then one may expect to see a true coefficient of variation approximately in the range of 0.55 ± 0.25 . Since, if the true coefficient of variation was below say 0.3 one would expect a choice of model M_0 . Whereas if there appeared to be a very large coefficient of variation, say greater than 0.8, it would be reasonable to assume that model M_h would be rejected anyway - in favour of perhaps M_{th} or M_{bh} . It is believed that in most practical situations one may expect the expectation of the distribution of p to vary between 0.04 and 0.20. For the above reasons tables 3.5.2a to 3.5.2i cover the following nine points in the ($E(p)$, $\sqrt{\text{Var}(p)}/E(p)$) plane :

$$\begin{array}{ccc}
 0.04 & & 0.3 \\
 E(p) = 0.12 \times \sqrt{\text{Var}(p)} / E(p) = 0.55. & & \\
 0.20 & & 0.8
 \end{array}$$

At each point in the above grid it necessarily follows that the parameters of the Beta distribution satisfy the equations $\alpha = \frac{1 - ep}{(cv)^2} - ep$ and $\beta = \frac{(1 - ep)^2}{ep(cv)^2} - (1 - ep)$, where $ep = E(p)$ and $cv = \sqrt{\text{Var}(p)} / E(p)$.

The range of detection probabilities covered by the above grid is consistent with the simulation studies of Burnham and Overton(1979), Chao, Lee and Jeng(1992) and Lee and Chao(1994).

In order to further investigate the robustness of the estimators over the entire subset of the ($E(p)$, $\sqrt{\text{Var}(p)}/E(p)$) plane, which is considered here to be appropriate for the model M_h , the simulations of tables 3.5.3a, 3.5.3b and 3.5.3c were carried out in a slightly different way. At the beginning of each simulation the value of $E(p)$ was selected as a random observation from a uniform distribution on the interval (0.04, 0.20). The value of the coefficient of variation was selected as a random observation from a uniform distribution on the interval (0.3, 0.8). The distribution of p was chosen to be Beta(alpha,

beta), and so to achieve the required values of ep and cv it was required that $\alpha = \frac{1-ep}{(cv)^2} - ep$ and $\beta = \frac{(1-ep)^2}{ep(cv)^2} - (1-ep)$. The capture probabilities of the N animals were drawn as a random sample from this beta distribution and live trapping was then simulated in the usual way.

Notation and Estimators

N \equiv population size.

t \equiv number of sampling occasions.

The following estimators are considered within the simulation study. The notation used for each estimator is stated, and where possible a detailed expression for the estimator is given.

x \equiv number of distinct individuals seen.

mle $\equiv \hat{N}$ the maximum likelihood estimator for the model M_0 , for details please refer to chapter 1.

boot $\equiv \hat{N}_B = x + \sum_{i=1}^t f_i \left(1 - \frac{i}{t}\right)^t$ the bootstrap estimator of Smith and van Belle(1984).

dr1 $\equiv \hat{N}_{0,1} = \frac{x}{\hat{C}_1}$ represents the estimator proposed by Darroch &

Ratcliff(1980) for the classical species problem.

$$ac1 = \hat{N}_{ac1} = \frac{x}{\hat{C}_1} + \frac{f_1}{\hat{C}_1} \hat{\gamma}_1^2$$

$$ac2 = \hat{N}_{ac2} = \frac{x}{\hat{C}_2} + \frac{f_1}{\hat{C}_2} \hat{\gamma}_2^2$$

$$ac3 = \hat{N}_{ac3} = \frac{x}{\hat{C}_3} + \frac{f_1}{\hat{C}_3} \hat{\gamma}_3^2$$

the three estimators proposed by Chao, Lee

and Jeng(1992), see also Lee and Chao(1994).

$$\hat{\gamma}_i^2 = \max \left\{ \frac{\hat{N}_{0,i} t \sum_{k=2}^t k(k-1) f_k}{(t-1) \left[\sum_{k=1}^t k f_k \right]^2} - 1, 0 \right\}, \quad i = 1, 2, 3.$$

$$O \equiv \hat{N}_O = \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \frac{i}{t}\right)^t} \quad \text{the estimator of Overton(1969).}$$

$$ca1 \equiv \hat{N}_{ca1} = \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \left(1 - \frac{f_1}{z}\right)\frac{i}{t}\right)^t}$$

$$ca2 \equiv \hat{N}_{ca2} = \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \left(1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z}\right)\frac{i}{t}\right)^t} \quad \text{the CAEs of section 3.3.}$$

$$ca3 \equiv \hat{N}_{ca3} = \sum_{i=1}^t \frac{f_i}{1 - \left(1 - \left(1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z} - \frac{6}{(t-1)(t-2)} \frac{f_3}{z}\right)\frac{i}{t}\right)^t}$$

pojac $\equiv \hat{N}_U$ the first order jackknife of the estimator proposed by Overton(1969), considered by Pollock and Otto(1983). Goudie(1996, pers.comm.) noted a typographical error in that paper. That is \hat{N}_U is in fact explicitly given by

$$\hat{N}_U = f_1 \left\{ t \cdot a_{1,t} - \frac{(t-1)^2}{t} a_{1,t-1} \right\} + \sum_{i=2}^t f_i \left[t \cdot a_{i,t} - \frac{(t-1)}{t} \{ i \cdot a_{i-1,t-1} + (t-i) a_{i,t-1} \} \right],$$

$$\text{where } a_{i,t} = \left\{ 1 - \left(1 - \frac{i}{t}\right)^t \right\}^{-1}$$

jac1 $\equiv \hat{N}_{J1} = x + \frac{(t-1)}{t} f_1$ the first order jackknife estimator of Burnham & Overton(1978).

jacseq $\equiv \hat{N}_{Jk}$ the jackknife of order k - where k is chosen from the set (1,2,3,4) according to the procedure of Burnham & Overton(1978).

jacint $\equiv \hat{N}_J$ the interpolated jackknife estimator of Burnham & Overton(1979).

N.B. The selection procedure proposed by Burnham and Overton(1978) is not entirely objective. That is if the fourth order jackknife is rejected their recommendation is to select the jackknife of whichever of the first three orders appears most appropriate. In order to avoid this subjectivity it was decided that, in the event of the selection procedure rejecting the fourth order jackknife, the estimator \hat{N}_{Jk} should be made equivalent to \hat{N}_{J1} . That is the selection procedure has been slightly modified, but only to the extent of always selecting the first order jackknife when the fourth order is rejected. Alternative modifications of the selection procedure

were investigated. However, the above version of the sequential jackknife is the one whose performance was best. Some of the alternative modifications of the sequential jackknife that were considered may merit future consideration.

When the selection procedure chooses the first order jackknife then the interpolated jackknife \hat{N}_J is equal to \hat{N}_{J1} . When the selection procedure chooses the jackknife of order k , for $k = 2, 3$ or 4 , then \hat{N}_J is a weighted average of the jackknives of orders k and $k-1$. When the selection procedure rejects the fourth order jackknife, \hat{N}_J is equal to \hat{N}_{J1} .

s.d. \equiv standard deviation.

rmse \equiv root mean square error.

$\Pr(\text{inf mle}) \equiv$ the probability that an estimator in the set $\{\text{mle}, \text{dr1}, \text{ac1}, \text{ca1}\}$ is infinite : each estimator in this set is infinite if and only if $1 - \frac{f_1}{Z} = 0$.

N.B. All results are given conditional on $1 - \frac{f_1}{Z} > 0$.

$$c \equiv C = \frac{\sum_{i=1}^N p_i I(X_i > 0)}{\sum_{i=1}^N p_i} \equiv \text{'Sample Coverage'}$$

ch1 $\equiv \hat{C}_1$

ch2 $\equiv \hat{C}_2$

ch3 $\equiv \hat{C}_3$

the three estimators of sample coverage proposed by Chao, Lee and Jeng(1992), see section 3.3.

cv1 $\equiv \hat{\gamma}_1$

cv2 $\equiv \hat{\gamma}_2$

cv3 $\equiv \hat{\gamma}_3$

the three estimators of the coefficient of variation proposed by Chao, Lee and Jeng(1992), where $\hat{\gamma}_i$, for $i = 1, 2, 3$, are as above.

N.B. Two of the estimators which were discussed in section 3.4 have not been included in the simulation study. The estimator introduced by Chao(1989) produces infinite estimates when no animals are seen exactly twice during the sampling experiment, and so to avoid imposing additional constraints on each simulation this estimator was omitted from the study. The nonparametric maximum likelihood estimator of Norris and Pollock(1996a) was also omitted : Norris and Pollock(1996a) noted that although possessing small bias, the variance of their estimator was usually large when compared to the estimators of Chao, Lee and Jeng(1992).

Discussion

The following discussion is based on an inspection of all fifteen tables, but with particular attention being paid to tables 3.5.3a, 3.5.3b and 3.5.3c, which are believed to give the best overall view of how the estimators perform in practice.

Firstly, as one would expect, the estimators designed for the model M_0 , namely the maximum likelihood estimator \hat{N} and the Darroch and Ratcliff estimator $\hat{N}_{0,1}$, perform well in situations when the heterogeneity is mild. However these estimators are seen to become negatively biased when the coefficient of variation becomes significantly large, and it is these situations in which their performance is unsatisfactory.

The bootstrap estimator of Smith and van Belle(1984) does not perform well. In almost all situations it is negatively biased, even when a large proportion of the population is seen during the experiment. And when a small proportion of the population is seen, its negative bias is extreme. The performance of the first order jackknife estimator \hat{N}_{J1} generally dominates that of the bootstrap estimator \hat{N}_B . This is an intuitively reasonable outcome : both \hat{N}_{J1} and \hat{N}_B are based on the enumeration estimator x . But whereas x is an ideal estimator to jackknife, being biased with small variance, it is not ideal for bootstrapping. For point estimation, it would be best to bootstrap an estimator with small bias and large variance. Within this context, however, it is believed that bootstrapping would be of most use in obtaining confidence intervals.

In terms of bias, the Coverage Adjusted estimators clearly perform better than the Overton estimator \hat{N}_O - particularly when sample coverage is small. The reason for this being that the estimators $\hat{p}_i^{(1)}$ tend to overestimate capture probability and consequently the estimator \hat{N}_O , which directly incorporates the $\hat{p}_i^{(1)}$, has a tendency to always underestimate population size. When sample coverage is small, the estimators $\hat{p}_i^{(1)}$ are particularly positively biased and so \hat{N}_O is particularly negatively biased. The estimators $\hat{p}_{i,j}^{(2)}$ are reasonable estimators of capture probability for most values of sample coverage and as a result of this \hat{N}_{ca1} , \hat{N}_{ca2} and \hat{N}_{ca3} generally possess an acceptable mean for all values of sample coverage. As one would expect, the Coverage Adjusted estimators perform in a very similar way to the Overton estimator \hat{N}_O when a large proportion of the population is seen during sampling. This is easily explained by the fact that, for each i and j , the value of $\hat{p}_{i,j}^{(2)}$ tends towards that of $\hat{p}_i^{(1)}$ as t is increased - since $\hat{C}_j \rightarrow 1$ as $t \rightarrow \infty$, for $j=1,2,3$. As a consequence of the significant improvement in mean which \hat{N}_{ca1} , \hat{N}_{ca2} and \hat{N}_{ca3} have over \hat{N}_O , the Coverage Adjusted estimators, although having a larger variance than \hat{N}_O , usually possess a much smaller mean square error.

Consider the relative performance of the Coverage Adjusted estimators and the estimators of Chao, Lee and Jeng(1992). It is, as a first step, worthwhile to note that, in most situations, the standard deviations of \hat{N}_{ca1} , \hat{N}_{ca2} and \hat{N}_{ca3} are less than those of \hat{N}_{ac1} , \hat{N}_{ac2} and \hat{N}_{ac3} respectively. With reference to tables 3.5.2a, 3.5.2d and 3.5.2g one may observe that in situations wherein the coefficient of variation is small, the coverage adjusted estimators can become positively biased, whereas in contrast \hat{N}_{ac1} , \hat{N}_{ac2} and \hat{N}_{ac3} each tend to possess a very good mean value. Despite this, in situations of this type wherein less than about 60% of the population is seen during the experiment, the coverage adjusted estimators, owing to their smaller variance, are seen to be performing best in terms of mean square error. With reference to tables 3.5.2b,c, 3.5.2e,f and 3.5.2h,i one may observe that in situations wherein a moderate to large coefficient of variation is present, the coverage adjusted estimators tend to perform well : they generally possess a good mean value and a relatively small variance. Consequently in these situations the coverage adjusted estimators tend to perform better than the estimators of Chao, Lee and Jeng(1992) in terms of both mean and variance. Hence for the majority of situations in which the model M_h would seem to be the most appropriate choice it is seen that the coverage adjusted estimators tend to perform better than the estimators of Chao, Lee and Jeng(1992). Tables 3.5.3a,b and c support this conclusion.

It is clear from tables 3.5.1a, 3.5.1b and 3.5.1c that the estimators of Chao, Lee and Jeng(1992) do not respond well when the heterogeneity results from the capture probabilities having a uniform distribution. In situations of this type, even when a large proportion of the population is seen during the experiment, the estimators of Chao, Lee and Jeng(1992) can be very negatively biased. In contrast to this the coverage adjusted estimators perform particularly well in tables 3.5.1a, 3.5.1b and 3.5.1c.

When considering the problem of estimating sample coverage it is seen that the estimators \hat{C}_2 and \hat{C}_3 are to be preferred to \hat{C}_1 . This is particularly true when the number of sampling occasions is small. For this reason, along with observing the performance of the estimators in the tables, it is recommended that \hat{N}_{ca2} , \hat{N}_{ca3} be preferred to \hat{N}_{ca1} . In a similar way it is believed that \hat{N}_{ac2} , \hat{N}_{ac3} should be preferred to \hat{N}_{ac1} .

The comparison between \hat{N}_{ca2} , \hat{N}_{ca3} and the first order jackknife estimator \hat{N}_{j1} is seen to depend mainly upon the value of sample coverage - or equivalently upon our estimate of sample coverage since this quantity may be estimated very well. If sample coverage is greater than say 0.7 then \hat{N}_{ca2} , \hat{N}_{ca3} and \hat{N}_{j1} all perform very well in that they have small bias and relatively small variance. However if sample coverage is less than 0.7 the first order jackknife tends to be very negatively biased whereas \hat{N}_{ca2} , \hat{N}_{ca3} continue to achieve a good mean value. The first order jackknife estimator has a very small variance

and other than x , the enumeration estimator, and \hat{N}_B , the bootstrap estimator, usually has the smallest variance of all of the estimators. For this reason, even though \hat{N}_{ca2} , \hat{N}_{ca3} have a better mean than \hat{N}_{j1} , when sample coverage is less than 0.7, the first order jackknife estimator can have a smaller mean square error. Even though this is true, it is believed that overall the Coverage Adjusted estimators \hat{N}_{ca2} , \hat{N}_{ca3} are to be preferred to the first order jackknife estimator.

The sequential jackknife estimator is generally seen to be superior to the first order jackknife in terms of mean - particularly when sample coverage is small. This is due to the fact that the sequential selection procedure developed by Burnham & Overton(1978) generally works well : when sample coverage is low high order jackknives are usually chosen whereas when sample coverage is high low order jackknives are usually chosen. However even with a lot of data the sequential selection procedure used to determine the sequential jackknife can be unpredictable. Rosenberg, Overton and Anthony(1995) stated that, when capture probabilities are low and heterogeneous, the selection procedure should be 'treated with caution'. The performance of the selection procedure results in the sequential jackknife estimator having a high variance even with good data. A good indication as to which jackknife estimator would be most appropriate is sample coverage. That is if our estimate of sample coverage is high, say above 0.7, then the first order, or second order, jackknife estimator should be considered most appropriate. If however sample coverage is smaller then the sequential selection procedure of Burnham & Overton(1978) is to date the best way of deciding which order jackknife to choose.

Consider the interpolated jackknife estimator. Due to the imprecise and often unpredictable nature of the sequential selection procedure, it is believed that the interpolated jackknife does not differ greatly enough from the sequential jackknife to warrant consideration as an estimator in its own right. In other words the difference between the interpolated jackknife and the sequential jackknife, in any one given situation, is believed to be insignificant when compared with the variance of the sequential jackknife estimator.

Of the jackknife estimators then, the first order jackknife is to be preferred if sample coverage is high, say above 0.7, whereas if sample coverage is small then the sequential jackknife should be considered a more appropriate choice.

Our conclusion above was that \hat{N}_{ca2} , \hat{N}_{ca3} are to be preferred to the first order jackknife. The above discussion now also implies that \hat{N}_{ca2} , \hat{N}_{ca3} are to be preferred to the sequential jackknife estimator when sample coverage is above 0.7. It remains to consider how \hat{N}_{ca2} , \hat{N}_{ca3} compare to the sequential jackknife when sample coverage is small, or

rather less than about 0.7. When sample coverage is very small, usually for $t = 5$ sampling occasions, the sequential jackknife estimator, although being negatively biased, can have a smaller mean square error than $\hat{N}_{ca2}, \hat{N}_{ca3}$ - although in this situation $\hat{N}_{ca2}, \hat{N}_{ca3}$ tend to have a much better mean. As sample coverage becomes larger, or as t is increased, the sequential jackknife estimator becomes less biased but, as mentioned above, does have a relatively large variance. As a result of this the Coverage Adjusted estimators $\hat{N}_{ca2}, \hat{N}_{ca3}$ generally perform far better than the sequential jackknife estimator in terms of mean square error - whilst at the same time performing in very much the same way in terms of bias.

Table 3.5.1c

$N = 100 : p \sim U(0, 0.40) : E(p) = 0.20 : \sqrt{\text{Var}(p)}/E(p) = 0.5774$
 Number of simulations = 1000

Number of sampling occasions, t = ... 5					Number of sampling occasions, t = ... 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	60.25	-39.75	5.004	40.061	x	77.20	-22.80	4.149	23.175
mle	77.62	-22.38	8.244	23.854	mle	81.89	-18.11	4.661	18.698
dr1	89.21	-10.79	10.410	14.994	dr1	86.96	-13.04	5.332	14.088
boot	72.21	-27.79	6.059	28.446	boot	87.50	-12.50	4.867	13.410
ac1	98.17	-1.83	14.773	14.887	ac1	91.05	-8.95	6.796	11.237
ac2	81.09	-18.91	11.519	22.144	ac2	88.16	-11.84	6.492	13.507
ac3	87.08	-12.92	13.519	18.702	ac3	89.01	-10.99	6.575	12.803
O	77.42	-22.58	6.584	23.524	O	91.88	-8.12	5.336	9.716
ca1	96.00	-4.00	11.048	11.752	ca1	95.82	-4.18	6.199	7.475
ca2	88.81	-11.19	9.750	14.842	ca2	94.96	-5.04	6.069	7.890
ca3	91.37	-8.63	10.252	13.400	ca3	95.20	-4.80	6.093	7.754
pojac	106.00	6.00	11.642	13.098	pojac	104.73	4.73	10.596	11.604
jac1	85.66	-14.34	7.623	16.245	jac1	97.03	-2.97	6.352	7.010
jacseq	97.95	-2.05	14.057	14.206	jacseq	98.36	-1.64	9.042	9.190
jacint	93.02	-6.98	13.689	15.366	jacint	97.62	-2.38	7.623	7.986
Pr(inf mle), mean jackknife order = 0.000, 2.242 c, ch1, ch2, ch3 = 0.749, 0.680, 0.775, 0.738 cvh1, cvh2, cvh3 = 0.398147, 0.175534, 0.259839					Pr(inf mle), mean jackknife order = 0.000, 1.110 c, ch1, ch2, ch3 = 0.905, 0.889, 0.911, 0.904 cvh1, cvh2, cvh3 = 0.381294, 0.339151, 0.352298				
Number of sampling occasions, t = ... 15					Number of sampling occasions, t = ... 20				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	84.52	-15.48	3.611	15.900	x	88.14	-11.86	3.175	12.276
mle	85.70	-14.30	3.689	14.771	mle	88.18	-11.82	3.182	12.244
dr1	89.19	-10.81	4.010	11.532	dr1	90.91	-9.09	3.413	9.708
boot	92.51	-7.49	4.108	8.540	boot	94.52	-5.48	3.628	6.571
ac1	92.35	-7.65	4.845	9.054	ac1	93.69	-6.31	4.105	7.524
ac2	91.48	-8.52	4.758	9.757	ac2	93.32	-6.68	4.067	7.818
ac3	91.68	-8.32	4.782	9.600	ac3	93.39	-6.61	4.080	7.768
O	95.80	-4.20	4.430	6.106	O	97.17	-2.83	3.958	4.868
ca1	97.11	-2.89	4.723	5.536	ca1	97.75	-2.25	4.116	4.689
ca2	96.91	-3.09	4.675	5.603	ca2	97.69	-2.31	4.101	4.710
ca3	96.96	-3.04	4.690	5.588	ca3	97.70	-2.30	4.108	4.708
pojac	101.85	1.85	9.160	9.345	pojac	101.01	1.01	8.395	8.456
jac1	99.02	-0.98	5.401	5.489	jac1	99.47	-0.53	4.634	4.664
jacseq	99.63	-0.37	7.360	7.369	jacseq	99.97	-0.03	6.082	6.082
jacint	99.41	-0.59	6.521	6.547	jacint	99.79	-0.21	5.485	5.489
Pr(inf mle), mean jackknife order = 0.000, 1.048 c, ch1, ch2, ch3 = 0.955, 0.948, 0.955, 0.953 cvh1, cvh2, cvh3 = 0.420097, 0.408829, 0.411348					Pr(inf mle), mean jackknife order = 0.000, 1.043 c, ch1, ch2, ch3 = 0.973, 0.970, 0.973, 0.972 cvh1, cvh2, cvh3 = 0.456086, 0.451684, 0.452397				
Number of sampling occasions, t = ... 25					Number of sampling occasions, t = ... 30				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	90.43	-9.57	2.898	9.997	x	91.80	-8.20	2.725	8.644
mle	90.43	-9.57	2.898	9.997	mle	91.80	-8.20	2.725	8.644
dr1	92.25	-7.75	3.045	8.326	dr1	93.04	-6.96	2.778	7.496
boot	95.68	-4.32	3.270	5.418	boot	96.20	-3.80	2.992	4.838
ac1	94.64	-5.36	3.552	6.433	ac1	95.09	-4.91	3.144	5.830
ac2	94.44	-5.56	3.547	6.597	ac2	94.98	-5.02	3.126	5.912
ac3	94.47	-5.53	3.555	6.577	ac3	94.99	-5.01	3.128	5.908
O	97.84	-2.16	3.571	4.174	O	97.98	-2.02	3.220	3.799
ca1	98.17	-1.83	3.689	4.119	ca1	98.14	-1.86	3.245	3.740
ca2	98.14	-1.86	3.681	4.123	ca2	98.13	-1.87	3.237	3.739
ca3	98.15	-1.85	3.684	4.124	ca3	98.13	-1.87	3.239	3.739
pojac	100.71	0.71	7.480	7.513	pojac	99.95	-0.05	6.717	6.718
jac1	99.97	-0.03	4.209	4.209	jac1	99.75	-0.25	3.890	3.898
jacseq	100.48	0.48	5.974	5.993	jacseq	100.17	0.17	5.324	5.326
jacint	100.27	0.27	5.091	5.099	jacint	100.00	0.00	4.658	4.658
Pr(inf mle), mean jackknife order = 0.000, 1.043 c, ch1, ch2, ch3 = 0.982, 0.980, 0.982, 0.982 cvh1, cvh2, cvh3 = 0.474867, 0.472717, 0.472992					Pr(inf mle), mean jackknife order = 0.000, 1.041 c, ch1, ch2, ch3 = 0.987, 0.987, 0.987, 0.987 cvh1, cvh2, cvh3 = 0.486945, 0.485736, 0.485865				

Table 3.5.2b

N = 100 : p ~ Beta(alpha,beta) : E(p) = 0.04 : sqrt[Var(p)]/E(p) = 0.55
alpha = 3.1336 : beta = 75.2053 : Number of simulations = 1000

Number of sampling occasions, t = ... 5					Number of sampling occasions, t = ... 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	18.22	-81.78	3.865	81.869	x	32.10	-67.90	4.472	68.044
mle	84.19	-15.81	47.739	50.289	mle	87.35	-12.65	36.223	38.369
dr1	109.19	9.19	62.120	62.796	dr1	98.65	-1.35	41.271	41.293
boot	23.64	-76.36	5.081	76.526	boot	41.60	-58.40	5.839	58.690
ac1	130.47	30.47	79.963	85.570	ac1	110.49	10.49	51.213	52.277
ac2	94.81	-5.19	59.216	59.443	ac2	98.37	-1.63	45.155	45.184
ac3	98.89	-1.11	71.136	71.145	ac3	99.19	-0.81	45.578	45.586
O	26.19	-73.81	5.633	74.029	O	46.41	-53.59	6.570	53.996
ca1	111.65	11.65	62.526	63.601	ca1	104.55	4.55	41.639	41.886
ca2	92.20	-7.80	50.324	50.925	ca2	97.39	-2.61	37.659	37.749
ca3	93.20	-6.80	50.960	51.412	ca3	97.76	-2.24	37.702	37.768
pojac	45.52	-54.48	10.278	55.436	pojac	78.36	-21.64	12.261	24.874
jac1	31.10	-68.90	6.781	69.234	jac1	55.01	-44.99	7.912	45.678
jacseq	47.96	-52.04	11.612	53.318	jacseq	70.80	-29.20	16.902	33.736
jacint	45.14	-54.86	10.999	55.952	jacint	66.65	-33.35	15.480	36.766
Pr(inf mle) , mean jackknife order = 0.168, 3.860 c, ch1, ch2, ch3 = 0.231, 0.211 , 0.259 , 0.256 cvh1, cvh2, cvh3 = 0.400661, 0.090000, 0.103173					Pr(inf mle) , mean jackknife order = 0.000, 2.228 c, ch1, ch2, ch3 = 0.396, 0.359 , 0.390 , 0.388 cvh1, cvh2, cvh3 = 0.271022, 0.191507, 0.198043				
Number of sampling occasions, t = ... 15					Number of sampling occasions, t = ... 20				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	43.17	-56.83	4.901	57.038	x	51.60	-48.40	4.932	48.648
mle	82.04	-17.96	18.550	25.821	mle	82.31	-17.69	13.525	22.269
dr1	89.06	-10.94	20.892	23.584	dr1	88.09	-11.91	14.888	19.063
boot	55.06	-44.94	6.332	45.388	boot	64.87	-35.13	6.357	35.699
ac1	97.78	-2.22	28.030	28.117	ac1	95.78	-4.22	19.940	20.381
ac2	91.42	-8.58	25.737	27.128	ac2	91.72	-8.28	18.774	20.520
ac3	91.78	-8.22	25.903	27.177	ac3	91.95	-8.05	18.850	20.499
O	61.07	-38.93	7.095	39.571	O	71.50	-28.50	7.137	29.378
ca1	97.45	-2.55	21.618	21.767	ca1	98.13	-1.87	15.800	15.909
ca2	94.16	-5.84	20.537	21.351	ca2	96.22	-3.78	15.283	15.743
ca3	94.34	-5.66	20.574	21.338	ca3	96.31	-3.69	15.295	15.734
pojac	96.84	-3.16	13.602	13.965	pojac	107.66	7.66	14.091	16.038
jac1	71.05	-28.95	8.701	30.232	jac1	81.91	-18.09	8.598	20.028
jacseq	93.33	-6.67	20.541	21.596	jacseq	101.75	1.75	20.643	20.717
jacint	84.95	-15.05	19.713	24.801	jacint	92.08	-7.92	18.771	20.372
Pr(inf mle) , mean jackknife order = 0.000, 2.314 c, ch1, ch2, ch3 = 0.524, 0.501 , 0.525 , 0.523 cvh1, cvh2, cvh3 = 0.273519, 0.214161, 0.217915					Pr(inf mle) , mean jackknife order = 0.000, 2.085 c, ch1, ch2, ch3 = 0.613, 0.595 , 0.613 , 0.612 cvh1, cvh2, cvh3 = 0.290091, 0.246089, 0.248383				
Number of sampling occasions, t = ... 25					Number of sampling occasions, t = ... 30				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	58.96	-41.04	4.881	41.329	x	64.52	-35.48	4.726	35.790
mle	83.18	-16.82	9.752	19.444	mle	83.87	-16.13	8.123	18.057
dr1	88.20	-11.80	10.808	16.005	dr1	88.51	-11.49	8.921	14.548
boot	73.00	-27.00	6.080	27.681	boot	78.76	-21.24	5.870	22.038
ac1	95.16	-4.84	14.980	15.743	ac1	95.14	-4.86	12.624	13.529
ac2	92.39	-7.61	14.363	16.255	ac2	93.14	-6.86	12.235	14.030
ac3	92.53	-7.47	14.393	16.217	ac3	93.23	-6.77	12.261	14.005
O	79.85	-20.15	6.764	21.254	O	85.61	-14.39	6.535	15.805
ca1	99.43	-0.57	11.784	11.797	ca1	100.42	0.42	9.998	10.007
ca2	98.23	-1.77	11.557	11.691	ca2	99.64	-0.36	9.823	9.829
ca3	98.28	-1.72	11.572	11.699	ca3	99.69	-0.31	9.829	9.834
pojac	113.70	13.70	13.381	19.151	pojac	116.30	16.30	13.270	21.022
jac1	90.17	-9.83	8.144	12.768	jac1	95.24	-4.76	8.044	9.345
jacseq	104.69	4.69	19.744	20.293	jacseq	105.53	5.53	17.556	18.407
jacint	96.24	-3.76	16.388	16.813	jacint	99.03	-0.97	13.355	13.390
Pr(inf mle) , mean jackknife order = 0.000, 1.803 c, ch1, ch2, ch3 = 0.689, 0.674 , 0.688 , 0.687 cvh1, cvh2, cvh3 = 0.307286, 0.274258, 0.275945					Pr(inf mle) , mean jackknife order = 0.000, 1.583 c, ch1, ch2, ch3 = 0.743, 0.732 , 0.743 , 0.743 cvh1, cvh2, cvh3 = 0.326660, 0.301499, 0.302791				

Table 3.5.2c

$N = 100$: $p \sim \text{Beta}(\alpha, \beta)$: $E(p) = 0.04$: $\sqrt{\text{Var}(p)}/E(p) = 0.80$
 $\alpha = 1.4600$: $\beta = 35.0400$: Number of simulations = 1000

Number of sampling occasions, t = ... 5					Number of sampling occasions, t = ... 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	17.87	-82.13	3.825	82.220	x	30.63	-69.37	4.531	69.518
mle	70.25	-29.75	42.584	51.945	mle	70.16	-29.84	26.266	39.756
dr1	90.87	-9.13	55.401	56.149	dr1	79.98	-20.02	30.484	36.471
boot	23.06	-76.94	5.008	77.106	boot	39.36	-60.64	5.852	60.925
ac1	108.53	8.53	72.010	72.513	ac1	92.11	-7.89	41.191	41.939
ac2	80.48	-19.52	54.294	57.696	ac2	82.62	-17.38	37.183	41.043
ac3	84.47	-15.53	64.876	66.709	ac3	83.44	-16.56	37.827	41.293
O	25.54	-74.47	5.539	74.671	O	43.76	-56.24	6.560	56.621
ca1	93.21	-6.79	55.756	56.168	ca1	85.54	-14.46	30.928	34.141
ca2	77.55	-22.45	44.917	50.214	ca2	80.41	-19.59	28.276	34.398
ca3	78.56	-21.44	45.492	50.289	ca3	80.81	-19.19	28.399	34.277
pojac	43.77	-56.23	10.049	57.121	pojac	72.13	-27.87	12.220	30.428
jac1	30.20	-69.80	6.657	70.114	jac1	51.47	-48.53	7.911	49.172
jacseq	45.65	-54.35	11.452	55.546	jacseq	66.53	-33.47	16.585	37.356
jacint	43.05	-56.95	10.839	57.976	jacint	61.97	-38.03	15.637	41.120
Pr(inf mle) , mean jackknife order = 0.101, 3.764 c, ch1, ch2, ch3 = 0.279, 0.247, 0.302, 0.298 cvh1, cvh2, cvh3 = 0.389301, 0.116463, 0.133231					Pr(inf mle) , mean jackknife order = 0.000, 2.241 c, ch1, ch2, ch3 = 0.454, 0.416, 0.448, 0.446 cvh1, cvh2, cvh3 = 0.324446, 0.244188, 0.252078				
Number of sampling occasions, t = ... 15					Number of sampling occasions, t = ... 20				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	40.38	-59.62	4.854	59.817	x	48.08	-51.92	5.010	52.163
mle	68.18	-31.82	14.435	34.941	mle	68.94	-31.06	9.913	32.605
dr1	74.94	-25.06	16.400	29.949	dr1	74.95	-25.05	11.203	27.442
boot	51.02	-48.98	6.231	49.373	boot	59.79	-40.21	6.301	40.705
ac1	84.67	-15.33	23.747	28.263	ac1	84.79	-15.21	17.384	23.101
ac2	79.59	-20.41	22.100	30.081	ac2	81.64	-18.36	16.684	24.805
ac3	79.95	-20.05	22.261	29.957	ac3	81.85	-18.15	16.734	24.684
O	56.37	-43.63	6.987	44.190	O	65.58	-34.42	7.005	35.127
ca1	82.66	-17.34	17.225	24.442	ca1	84.17	-15.83	12.168	19.969
ca2	80.28	-19.72	16.424	25.667	ca2	82.89	-17.11	11.883	20.833
ca3	80.45	-19.55	16.486	25.576	ca3	82.96	-17.04	11.897	20.782
pojac	87.23	-12.77	13.359	18.481	pojac	96.21	-3.79	13.253	13.784
jac1	65.00	-35.00	8.485	36.009	jac1	74.67	-25.33	8.312	26.663
jacseq	83.38	-16.62	19.042	25.276	jacseq	89.81	-10.19	19.565	22.057
jacint	75.77	-24.23	17.788	30.059	jacint	82.05	-17.95	17.261	24.905
Pr(inf mle) , mean jackknife order = 0.000, 2.200 c, ch1, ch2, ch3 = 0.577, 0.554, 0.577, 0.576 cvh1, cvh2, cvh3 = 0.348919, 0.295317, 0.299306					Pr(inf mle) , mean jackknife order = 0.000, 1.929 c, ch1, ch2, ch3 = 0.662, 0.649, 0.665, 0.664 cvh1, cvh2, cvh3 = 0.397515, 0.362089, 0.364475				
Number of sampling occasions, t = ... 25					Number of sampling occasions, t = ... 30				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	54.44	-45.56	5.106	45.845	x	59.33	-40.67	4.739	40.949
mle	71.24	-28.76	8.341	29.942	mle	72.57	-27.43	7.309	28.391
dr1	76.93	-23.07	9.398	24.907	dr1	77.94	-22.06	8.305	23.571
boot	66.78	-33.22	6.308	33.811	boot	71.80	-28.20	5.865	28.808
ac1	86.66	-13.34	14.950	20.038	ac1	87.45	-12.55	12.928	18.020
ac2	84.50	-15.50	14.483	21.212	ac2	85.91	-14.09	12.649	18.934
ac3	84.62	-15.38	14.517	21.146	ac3	86.00	-14.00	12.656	18.873
O	72.80	-27.20	6.964	28.079	O	77.79	-22.21	6.535	23.151
ca1	87.17	-12.83	10.464	16.553	ca1	88.78	-11.22	9.418	14.652
ca2	86.35	-13.65	10.311	17.109	ca2	88.22	-11.78	9.324	15.020
ca3	86.38	-13.62	10.314	17.082	ca3	88.25	-11.75	9.328	15.000
pojac	102.32	2.32	13.080	13.285	pojac	104.80	4.80	13.099	13.951
jac1	81.82	-18.18	8.383	20.020	jac1	86.14	-13.86	8.055	16.028
jacseq	94.84	-5.16	18.529	19.233	jacseq	95.88	-4.12	18.433	18.888
jacint	87.83	-12.17	15.494	19.702	jacint	90.20	-9.80	14.496	17.496
Pr(inf mle) , mean jackknife order = 0.000, 1.785 c, ch1, ch2, ch3 = 0.729, 0.712, 0.724, 0.723 cvh1, cvh2, cvh3 = 0.431791, 0.406248, 0.407800					Pr(inf mle) , mean jackknife order = 0.000, 1.582 c, ch1, ch2, ch3 = 0.775, 0.765, 0.774, 0.773 cvh1, cvh2, cvh3 = 0.464550, 0.447307, 0.448278				

Table 3.5.2d

$N = 100$: $p \sim \text{Beta}(\alpha, \beta)$: $E(p) = 0.12$: $\sqrt{\text{Var}(p)}/E(p) = 0.30$
 $\alpha = 9.6578$: $\beta = 70.8237$: Number of simulations = 1000

Number of sampling occasions, $t = \dots 5$					Number of sampling occasions, $t = \dots 10$				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	46.24	-53.76	4.736	53.970	x	70.09	-29.91	4.451	30.235
mle	95.65	-4.35	22.571	22.986	mle	93.91	-6.09	8.598	10.536
dr1	115.87	15.87	29.166	33.205	dr1	100.91	0.91	9.949	9.990
boot	58.36	-41.64	6.024	42.070	boot	85.27	-14.73	5.559	15.749
ac1	134.23	34.23	40.991	53.403	ac1	106.33	6.33	13.201	14.638
ac2	101.96	1.96	29.620	29.684	ac2	98.32	-1.68	11.668	11.789
ac3	106.89	6.89	33.012	33.724	ac3	99.44	-0.56	11.952	11.965
O	63.93	-36.07	6.642	36.672	O	92.26	-7.74	6.168	9.895
ca1	121.90	21.90	29.583	36.810	ca1	111.95	11.95	10.778	16.093
ca2	105.64	5.64	24.374	25.018	ca2	108.31	8.31	10.175	13.135
ca3	107.95	7.95	25.065	26.295	ca3	108.81	8.81	10.247	13.515
pojac	102.05	2.05	12.207	12.378	pojac	125.75	25.75	12.721	28.717
jac1	73.89	-26.11	7.912	27.278	jac1	102.54	2.54	7.487	7.905
jacseq	102.80	2.80	14.878	15.139	jacseq	113.91	13.91	16.183	21.337
jacint	98.35	-1.65	14.351	14.445	jacint	106.66	6.66	13.236	14.815
Pr(inf mle) , mean jackknife order = 0.000, 3.716 c, ch1, ch2, ch3 = 0.494, 0.417, 0.499, 0.486 cvh1, cvh2, cvh3 = 0.399709, 0.151290, 0.191827					Pr(inf mle) , mean jackknife order = 0.000, 1.708 c, ch1, ch2, ch3 = 0.735, 0.698, 0.739, 0.733 cvh1, cvh2, cvh3 = 0.259418, 0.163912, 0.177511				
Number of sampling occasions, $t = \dots 15$					Number of sampling occasions, $t = \dots 20$				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	82.68	-17.32	3.697	17.706	x	89.72	-10.28	3.078	10.732
mle	95.11	-4.89	5.163	7.112	mle	96.06	-3.94	3.666	5.379
dr1	99.23	-0.77	5.877	5.928	dr1	98.94	-1.06	4.156	4.288
boot	96.83	-3.17	4.530	5.529	boot	101.59	1.59	3.732	4.057
ac1	102.38	2.38	7.616	7.981	ac1	101.00	1.00	5.190	5.284
ac2	99.11	-0.89	7.180	7.235	ac2	99.51	-0.49	5.068	5.092
ac3	99.55	-0.45	7.231	7.245	ac3	99.69	-0.31	5.070	5.080
O	103.01	3.01	5.053	5.880	O	106.43	6.43	4.233	7.700
ca1	111.21	11.21	6.787	13.103	ca1	109.96	9.96	5.076	11.177
ca2	109.99	9.99	6.610	11.981	ca2	109.51	9.51	4.996	10.743
ca3	110.16	10.16	6.634	12.138	ca3	109.57	9.57	5.016	10.802
pojac	121.99	21.99	11.734	24.922	pojac	113.13	13.13	10.752	16.973
jac1	110.14	10.14	6.394	11.991	jac1	110.68	10.68	5.618	12.068
jacseq	111.76	11.76	9.266	14.973	jacseq	110.79	10.79	6.572	12.634
jacint	110.55	10.55	7.148	12.741	jacint	110.72	10.72	6.050	12.310
Pr(inf mle) , mean jackknife order = 0.000, 1.110 c, ch1, ch2, ch3 = 0.854, 0.835, 0.855, 0.852 cvh1, cvh2, cvh3 = 0.245586, 0.195656, 0.202314					Pr(inf mle) , mean jackknife order = 0.000, 1.019 c, ch1, ch2, ch3 = 0.917, 0.907, 0.918, 0.917 cvh1, cvh2, cvh3 = 0.244734, 0.218949, 0.222307				
Number of sampling occasions, $t = \dots 25$					Number of sampling occasions, $t = \dots 30$				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	93.84	-6.16	2.429	6.617	x	96.10	-3.90	1.999	4.383
mle	97.01	-2.99	2.729	4.048	mle	97.58	-2.42	2.127	3.222
dr1	99.13	-0.87	3.028	3.151	dr1	99.28	-0.72	2.319	2.429
boot	103.31	3.31	2.971	4.449	boot	103.57	3.57	2.409	4.307
ac1	100.55	0.55	3.662	3.703	ac1	100.28	0.28	2.682	2.696
ac2	99.80	-0.20	3.609	3.614	ac2	99.86	-0.14	2.660	2.664
ac3	99.91	-0.09	3.635	3.636	ac3	99.90	-0.10	2.665	2.667
O	106.92	6.92	3.442	7.730	O	106.24	6.24	2.781	6.835
ca1	108.50	8.50	3.873	9.341	ca1	106.97	6.97	2.996	7.589
ca2	108.30	8.30	3.821	9.140	ca2	106.88	6.88	2.985	7.503
ca3	108.33	8.33	3.831	9.164	ca3	106.89	6.89	2.992	7.513
pojac	105.83	5.83	9.427	11.084	pojac	100.92	0.92	7.620	7.675
jac1	108.81	8.81	4.506	9.900	jac1	107.28	7.28	3.634	8.140
jacseq	108.80	8.80	5.814	10.546	jacseq	107.21	7.21	3.716	8.114
jacint	108.88	8.88	5.234	10.311	jacint	107.25	7.25	3.647	8.119
Pr(inf mle) , mean jackknife order = 0.000, 1.023 c, ch1, ch2, ch3 = 0.953, 0.947, 0.953, 0.952 cvh1, cvh2, cvh3 = 0.252211, 0.237973, 0.239814					Pr(inf mle) , mean jackknife order = 0.000, 1.006 c, ch1, ch2, ch3 = 0.972, 0.968, 0.972, 0.971 cvh1, cvh2, cvh3 = 0.263439, 0.255651, 0.256595				

Table 3.5.2e

N = 100 : $p \sim \text{Beta}(\alpha, \beta)$: $E(p) = 0.12$: $\sqrt{\text{Var}(p)}/E(p) = 0.55$
 $\alpha = 2.7891$: $\beta = 20.4533$: Number of simulations = 1000

Number of sampling occasions, t = 5					Number of sampling occasions, t = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	44.38	-55.62	4.957	55.838	x	66.19	-33.81	4.667	34.132
mle	82.29	-17.71	18.973	25.955	mle	84.08	-15.92	7.696	17.687
dr1	99.86	-0.14	24.817	24.817	dr1	91.69	-8.31	9.041	12.281
boot	55.59	-44.41	6.250	44.849	boot	79.97	-20.03	5.754	20.843
ac1	117.32	17.32	36.357	40.273	ac1	99.80	-0.20	13.156	13.158
ac2	90.36	-9.64	26.688	28.377	ac2	92.87	-7.13	11.977	13.940
ac3	95.64	-4.36	29.838	30.155	ac3	94.02	-5.98	12.200	13.589
O	60.74	-39.26	6.881	39.858	O	86.33	-13.67	6.364	15.074
ca1	105.59	5.59	25.275	25.887	ca1	102.00	2.00	9.926	10.125
ca2	92.85	-7.15	21.018	22.202	ca2	99.24	-0.76	9.457	9.487
ca3	95.19	-4.81	21.700	22.227	ca3	99.69	-0.31	9.519	9.524
pojac	95.02	-4.98	12.392	13.354	pojac	116.77	16.77	12.503	20.920
jac1	69.78	-30.22	8.117	31.286	jac1	95.67	-4.33	7.612	8.759
jacseq	94.60	-5.40	15.220	16.150	jacseq	106.02	6.02	15.629	16.749
jacint	90.41	-9.59	14.890	17.713	jacint	99.74	-0.26	12.584	12.587
Pr(inf mle) , mean jackknife order = 0.000, 3.524 c, ch1, ch2, ch3 = 0.541, 0.463, 0.547, 0.529 cvh1, cvh2, cvh3 = 0.432891, 0.193849, 0.242735					Pr(inf mle) , mean jackknife order = 0.000, 1.693 c, ch1, ch2, ch3 = 0.764, 0.725, 0.762, 0.755 cvh1, cvh2, cvh3 = 0.376647, 0.292608, 0.306845				
Number of sampling occasions, t = 15					Number of sampling occasions, t = 20				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	77.55	-22.45	4.182	22.832	x	84.17	-15.83	3.608	16.231
mle	86.60	-13.40	5.447	14.467	mle	88.76	-11.24	4.125	11.970
dr1	91.99	-8.01	6.236	10.153	dr1	93.09	-6.91	4.623	8.314
boot	90.52	-9.48	5.094	10.761	boot	95.54	-4.46	4.316	6.209
ac1	98.45	-1.55	9.055	9.187	ac1	98.39	-1.61	6.372	6.572
ac2	95.74	-4.26	8.695	9.683	ac2	97.10	-2.90	6.253	6.894
ac3	96.11	-3.89	8.748	9.572	ac3	97.26	-2.74	6.283	6.856
O	96.30	-3.70	5.655	6.756	O	100.42	0.42	4.804	4.822
ca1	103.28	3.28	7.265	7.972	ca1	103.86	3.86	5.586	6.793
ca2	102.37	2.37	7.116	7.501	ca2	103.51	3.51	5.529	6.549
ca3	102.52	2.52	7.122	7.556	ca3	103.56	3.56	5.545	6.591
pojac	116.46	16.46	12.258	20.520	pojac	112.34	12.34	10.963	16.510
jac1	103.39	3.39	7.000	7.777	jac1	105.64	5.64	6.082	8.292
jacseq	106.56	6.56	12.026	13.700	jacseq	106.49	6.49	8.239	10.491
jacint	104.50	4.50	9.165	10.209	jacint	105.96	5.96	6.825	9.058
Pr(inf mle) , mean jackknife order = 0.000, 1.215 c, ch1, ch2, ch3 = 0.863, 0.844, 0.862, 0.859 cvh1, cvh2, cvh3 = 0.409232, 0.376264, 0.381227					Pr(inf mle) , mean jackknife order = 0.000, 1.064 c, ch1, ch2, ch3 = 0.914, 0.905, 0.914, 0.913 cvh1, cvh2, cvh3 = 0.440186, 0.425102, 0.427043				
Number of sampling occasions, t = 25					Number of sampling occasions, t = 30				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	88.46	-11.54	3.183	11.972	x	91.22	-8.78	2.889	9.241
mle	90.72	-9.28	3.372	9.869	mle	92.24	-7.76	2.989	8.317
dr1	94.28	-5.72	3.726	6.825	dr1	95.19	-4.81	3.219	5.788
boot	98.21	-1.79	3.760	4.163	boot	99.52	-0.48	3.384	3.418
ac1	98.78	-1.22	5.082	5.226	ac1	98.96	-1.04	4.208	4.335
ac2	98.09	-1.91	5.047	5.398	ac2	98.55	-1.45	4.203	4.447
ac3	98.17	-1.83	5.052	5.373	ac3	98.59	-1.41	4.198	4.429
O	102.28	2.28	4.182	4.763	O	102.85	2.85	3.750	4.708
ca1	104.12	4.12	4.611	6.182	ca1	103.89	3.89	3.975	5.561
ca2	103.94	3.94	4.601	6.057	ca2	103.81	3.81	3.971	5.500
ca3	103.96	3.96	4.599	6.071	ca3	103.81	3.81	3.968	5.502
pojac	108.72	8.72	9.874	13.175	pojac	105.69	5.69	8.806	10.486
jac1	105.72	5.72	5.261	7.769	jac1	105.54	5.54	4.469	7.120
jacseq	106.17	6.17	7.302	9.563	jacseq	105.89	5.89	6.711	8.926
jacint	105.99	5.99	6.358	8.735	jacint	105.80	5.80	5.830	8.222
Pr(inf mle) , mean jackknife order = 0.000, 1.034 c, ch1, ch2, ch3 = 0.944, 0.938, 0.944, 0.943 cvh1, cvh2, cvh3 = 0.461711, 0.454075, 0.454955					Pr(inf mle) , mean jackknife order = 0.000, 1.026 c, ch1, ch2, ch3 = 0.961, 0.959, 0.962, 0.961 cvh1, cvh2, cvh3 = 0.478040, 0.473689, 0.474141				

Table 3.5.2f

N = 100 : $p \sim \text{Beta}(\alpha, \beta)$: $E(p) = 0.12$: $\sqrt{\text{Var}(p)}/E(p) = 0.80$
 $\alpha = 1.2550$: $\beta = 9.2033$: Number of simulations = 1000

Number of sampling occasions, t = 5					Number of sampling occasions, t = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	41.41	-58.59	4.907	58.794	x	59.92	-40.08	4.822	40.369
mle	65.67	-34.33	12.967	36.694	mle	71.02	-28.98	6.699	29.742
dr1	79.63	-20.37	17.179	26.647	dr1	78.49	-21.51	7.916	22,920
boot	51.26	-48.74	6.155	49.132	boot	71.57	-28.43	5.834	29.026
ac1	94.82	-5.18	26.984	27.477	ac1	88.64	-11.36	12.303	16.746
ac2	75.08	-24.92	20.942	32.555	ac2	83.57	-16.43	11.461	20.030
ac3	80.31	-19.69	23.870	30.941	ac3	84.51	-15.49	11.630	19.368
O	55.69	-44.31	6.754	44.822	O	76.95	-23.05	6.387	23.919
ca1	84.86	-15.14	17.754	23.333	ca1	87.65	-12.35	8.820	15.178
ca2	76.27	-23.73	15.299	28.234	ca2	85.86	-14.14	8.508	16.506
ca3	78.47	-21.53	16.045	26.853	ca3	86.18	-13.82	8.567	16.261
pojac	84.49	-15.51	12.072	19.653	pojac	102.35	2.35	11.965	12.193
jac1	63.42	-36.58	7.977	37.440	jac1	84.80	-15.20	7.490	16.948
jacseq	82.48	-17.52	15.060	23.106	jacseq	93.77	-6.23	15.055	16.293
jacint	78.58	-21.42	15.046	26.176	jacint	88.75	-11.25	12.396	16.737

Pr(inf mle), mean jackknife order = 0.000, 3.191
c, ch1, ch2, ch3 = 0.602, 0.534, 0.620, 0.595
cvh1, cvh2, cvh3 = 0.477044, 0.263912, 0.324121

Number of sampling occasions, t = 15					Number of sampling occasions, t = 20				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	70.36	-29.64	4.569	29.990	x	76.16	-23.84	4.372	24.238
mle	75.77	-24.23	5.344	24.814	mle	78.81	-21.19	4.698	21.702
dr1	81.77	-18.23	6.144	19.241	dr1	83.86	-16.14	5.265	16.979
boot	81.59	-18.41	5.463	19.199	boot	86.43	-13.57	5.137	14.507
ac1	91.26	-8.74	9.449	12.873	ac1	92.77	-7.23	7.917	10.724
ac2	89.21	-10.79	9.128	14.136	ac2	91.73	-8.27	7.798	11.369
ac3	89.51	-10.49	9.173	13.938	ac3	91.84	-8.16	7.806	11.289
O	86.70	-13.30	6.005	14.588	O	91.03	-8.97	5.624	10.589
ca1	91.97	-8.03	7.198	10.784	ca1	93.99	-6.01	6.271	8.686
ca2	91.36	-8.64	7.102	11.182	ca2	93.72	-6.28	6.231	8.844
ca3	91.46	-8.54	7.116	11.116	ca3	93.75	-6.25	6.229	8.824
pojac	106.26	6.26	11.855	13.406	pojac	105.64	5.64	11.327	12.653
jac1	93.29	-6.71	7.119	9.784	jac1	96.50	-3.50	6.860	7.702
jacseq	97.81	-2.19	13.200	13.380	jacseq	99.14	-0.86	11.773	11.805
jacint	95.25	-4.75	10.512	11.534	jacint	97.68	-2.32	9.217	9.504

Pr(inf mle), mean jackknife order = 0.000, 1.321
c, ch1, ch2, ch3 = 0.877, 0.862, 0.875, 0.873
cvh1, cvh2, cvh3 = 0.552546, 0.532779, 0.535884

Number of sampling occasions, t = 25					Number of sampling occasions, t = 30				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x	80.54	-19.46	4.055	19.879	x	83.61	-16.39	3.739	16.815
mle	81.67	-18.33	4.202	18.802	mle	83.94	-16.06	3.817	16.512
dr1	85.84	-14.16	4.568	14.879	dr1	87.55	-12.45	4.067	13.102
boot	89.69	-10.31	4.686	11.325	boot	91.86	-8.14	4.289	9.198
ac1	93.82	-6.18	6.739	9.143	ac1	94.88	-5.12	5.821	7.755
ac2	93.27	-6.73	6.661	9.471	ac2	94.54	-5.46	5.767	7.940
ac3	93.33	-6.67	6.657	9.426	ac3	94.58	-5.42	5.774	7.922
O	93.69	-6.31	5.149	8.142	O	95.42	-4.58	4.643	6.523
ca1	95.43	-4.57	5.510	7.159	ca1	96.53	-3.47	4.900	6.004
ca2	95.28	-4.72	5.485	7.234	ca2	96.46	-3.54	4.893	6.038
ca3	95.30	-4.70	5.496	7.230	ca3	96.47	-3.53	4.892	6.035
pojac	104.24	4.24	10.192	11.037	pojac	103.30	3.30	9.243	9.812
jac1	97.84	-2.16	6.096	6.467	jac1	98.99	-1.01	5.291	5.386
jacseq	99.20	-0.80	9.596	9.629	jacseq	99.98	-0.02	8.044	8.044
jacint	98.53	-1.47	8.083	8.216	jacint	99.49	-0.51	6.495	6.515

Pr(inf mle), mean jackknife order = 0.000, 1.097
c, ch1, ch2, ch3 = 0.942, 0.939, 0.943, 0.942
cvh1, cvh2, cvh3 = 0.622433, 0.617217, 0.617767

Pr(inf mle), mean jackknife order = 0.000, 1.072
c, ch1, ch2, ch3 = 0.957, 0.955, 0.958, 0.958
cvh1, cvh2, cvh3 = 0.646150, 0.642928, 0.643218

§3.6 : Some Standard Data Sets

The aim of this section is to assess how the estimators perform on various standard data sets which have previously been studied in the literature. Of particular interest in this section is the fact that some of the following data sets were obtained from experiments carried out on populations of known size. It is also of interest to see how the results of this section, being obtained from real life populations, compare with the results of the simulation study of section 3.5. It is noted however that it would be unwise to draw firm conclusions from a small number of data sets.

The notation used in the following tables is identical to that of the previous section, with the following additions :

N \equiv population size (if known).

f_1 \equiv the number of animals seen exactly once.

f_2 \equiv the number of animals seen exactly twice.

.

.

f_t \equiv the number of animals seen exactly t times.

x \equiv number of distinct individuals seen.

z \equiv total number of sightings.

Carothers (1973) carried out a capture-recapture experiment on a population of 420 taxicabs working in Edinburgh. The whole data set, as illustrated in table 3.6.1, was presented along with various subsets. Tables 3.6.2a and 3.6.2b show how the estimators perform on all 42 subsets. A number of the subsets intersect and so not all of the estimates are independent. For details of how the different subsets were obtained please refer to Carothers(1973). The whole taxicab data set, along with the snowshoe hare data of Burnham and Cushwa, appeared in the Otis et al. (1978) wildlife monograph, wherein for both sets of data the model selection procedure of Otis et al. (1978) was shown to choose the model M_h . The model selection procedure of Otis et al. (1978) also selects the model M_h as being most appropriate for the meadow vole trapping data of Pollock et al. (1990). The model M_h has been judged appropriate by Norris and Pollock(1996) for the eastern chipmunk data of Mares et al. (1981). It is impossible to apply the model selection procedure of Otis et al. (1978) to the mud turtle data of Chao(1989) : this data set is included so as to illustrate through a real data example how the estimators perform in a situation wherein sample coverage is very small.

The performance and relative performance of the estimators in connection with the standard data sets is almost entirely consistent with the simulation study of section 3.5. One may firstly observe that the standard data sets illustrate the tendency of the maximum

likelihood estimator \hat{N} and Darroch and Ratcliff estimator $dr1$ to underestimate in the presence of heterogeneity. The bootstrap estimator of Smith and van Belle(1984) is negatively biased, and particularly so when sample coverage is small. Also immediately apparent is the way in which the coverage adjusted estimators clearly perform much better than the Overton estimator. The Overton estimator \hat{N}_o can be severely negatively biased and consequently its performance tends to be unacceptable, unless that is, as with the data of Mares et al. (1981) and Pollock et al. (1990), an extremely large proportion of the population is seen during the experiment. \hat{N}_o is particularly negatively biased when sample coverage is small whereas in this situation, for the reasons discussed in the previous section, the coverage adjusted estimators \hat{N}_{ca1} , \hat{N}_{ca2} and \hat{N}_{ca3} are still able to provide reasonable estimates of population size.

The estimators $ac1$ and $ca1$ both make use of $ch1$ as an estimator of sample coverage. In the same way $ac2$, $ca2$ and $ac3$, $ca3$ incorporate $ch2$ and $ch3$ respectively. In tables 3.6.2a and 3.6.2b it is therefore reasonable to compare $ca1$ to $ac1$, $ca2$ to $ac2$ and $ca3$ to $ac3$. In order to highlight this comparison, the cell containing the better estimate is shaded in each case. Looking at the shaded cells in tables 3.6.2a and 3.6.2b shows quite clearly that, for this particular population, the coverage adjusted estimators for each given estimate of sample coverage, tend to perform better than the corresponding $ac1$, $ac2$ or $ac3$. Explicitly, on 25 of the 42 subsets, the estimate provided by $ca1$ is closer to the true value of 420 than that given by $ac1$. Similarly on 35 of these 42 subsets, the estimators $ca2$ and $ca3$ improve on $ac2$ and $ac3$ respectively. This is consistent with the simulation study. However, for the complete taxicab data, as shown in table 3.6.1, $ac1$ is actually more accurate than $ca1$. For this data set, $ca2$ and $ca3$ are respectively more accurate than $ac2$ and $ac3$, so it is perhaps worthwhile to recall at this point that the simulation study recommended use of either $ca2$ or $ca3$ in favour of $ca1$.

The complete taxicab data of table 3.6.1 also illustrates how, even with good data, the sequential selection procedure of Burnham and Overton(1978) can mislead. That is, for the taxicab data, the sequential selection procedure chooses the third order jackknife as being the most appropriate - whereas this estimator clearly overestimates by a relatively large amount. The meadow vole data appears to provide a similar example. The discussion of the previous section recommended that, if estimates of sample coverage were above 0.7, then one should consider simply ignoring the sequential selection procedure of Burnham and Overton(1978) and always choose the first order jackknife. The standard data sets do lend support to this suggestion.

The mud turtle data set illustrates how, when sample coverage is very small, the jackknife estimators can possess extreme negative bias - for example, for the mud turtle data, even the fifth order jackknife estimator provides an estimate of 491 (Chao(1989) concluded that, if equal catchability were a reasonable assumption, then there were about 800 turtles in the habitat). This data set, therefore, also highlights the extreme negative bias of the bootstrap and Overton estimators when sample coverage is very small.

It has been noted that it would be unwise to draw firm conclusions from a small number of data sets. With this in mind, it is still, however, pleasing to observe that the performance of the estimators in connection with the real life data is consistent with their performance within the simulation study. And that furthermore, the recommendations of the previous section are supported by the performance of the estimators on the standard data sets.

Table 3.6.1 : Standard Data Sets

Source	Mares et al.(1981)	Carothers (1973)	Burnham and Cushwa. (see text)	Pollock et al. (1990)	A.Chao (1989)
Description	Eastern Chipmunks	Taxicab Data	Snowshoe Hare	Meadow Vole	Mud Turtle
N	82	420	-	-	-
t	13	10	6	5	40
f1	14	142	25	29	94
f2	13	81	22	15	5
f3	18	49	13	15	0
f4	12	7	5	16	0
f5	7	3	1	27	0
f6	5	1	2	-	0
f7	1	0	-	-	0
f8	1	0	-	-	0
f9 to f(t)	0	0	-	-	0
x	71	283	68	102	99
z	222	500	145	303	104
mle	73	368	75	103	1011
dr1	76	395	82	113	1030
boot	78	343	79	113	134
ac1	77	416	89	123	1030
ac2	76	386	81	118	1004
ac3	77	393	84	123	1004
O	81	370	83	118	153
ca1	82	439	90	121	1053
ca2	82	427	87	120	1028
ca3	82	429	88	121	1028
pojac	86	504	100	144	294
jac1	84	411	89	125	191
jacseq	84	495	89	142	191
jacint	84	469	89	138	191
order jac	1	3	1	3	1
ch1	0.937	0.716	0.828	0.904	0.096
ch2	0.947	0.752	0.888	0.929	0.099
ch3	0.943	0.744	0.861	0.904	0.099
cvh1	0.331	0.327	0.478	0.555	0.000
cvh2	0.313	0.232	0.380	0.523	0.000
cvh3	0.320	0.256	0.425	0.555	0.000

Table 3.6.2b : Taxicab Data : N = 420

Subset	Sampling Scheme A							Sampling Scheme B						
	Data Subset gamma							Data Subset gamma						
	a	b	c	d	e	f	g	a	b	c	d	e	f	g
t	10	10	10	10	10	10	10	10	10	10	10	10	10	10
f1	101	115	110	143	127	145	142	100	103	98	110	112	113	104
f2	33	27	27	62	73	67	81	30	26	29	67	62	61	67
f3	4	3	4	19	19	19	49	6	6	6	20	24	19	51
f4	0	0	0	3	3	1	7	0	0	1	3	8	7	12
f5	0	0	0	0	1	2	3	0	0	0	3	0	2	6
f6	0	0	0	0	0	0	1	0	0	0	0	0	0	1
f7 to f10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x	138	145	141	227	223	234	283	136	135	134	203	206	202	241
z	179	178	176	336	347	350	500	178	173	178	331	340	330	475
mle	299	380	347	368	335	373	368	287	304	272	287	287	285	288
dr1	317	410	376	395	352	400	395	310	334	298	304	307	307	309
boot	177	188	182	284	276	292	343	174	174	171	249	252	249	286
ac1	317	421	391	412	357	421	416	321	357	322	317	320	329	324
ac2	289	374	345	369	328	376	386	286	310	281	287	292	299	306
ac3	291	375	346	372	330	380	393	288	313	284	290	296	302	311
O	196	210	203	312	300	320	370	193	193	190	271	274	270	306
ca1	341	436	402	434	388	439	439	335	358	322	337	340	340	343
ca2	319	405	375	414	371	419	427	314	337	304	324	328	328	336
ca3	320	406	376	416	373	421	429	316	338	305	325	330	329	338
pojac	316	354	340	467	425	475	504	313	320	308	375	384	383	395
jac1	229	249	240	356	337	365	411	226	228	222	302	307	304	335
jacseq	320	249	240	463	387	465	495	347	228	342	342	352	351	370
jacint	312	249	240	435	345	436	469	331	228	326	306	323	319	349
order jac	3	1	1	3	2	3	3	4	1	4	2	2	2	2
ch1	0.436	0.354	0.375	0.574	0.634	0.586	0.716	0.438	0.405	0.449	0.668	0.671	0.658	0.781
ch2	0.477	0.388	0.409	0.615	0.681	0.628	0.752	0.476	0.438	0.486	0.713	0.711	0.699	0.812
ch3	0.475	0.386	0.407	0.611	0.676	0.624	0.744	0.473	0.435	0.483	0.708	0.705	0.694	0.803
cov1	0.000	0.186	0.228	0.256	0.160	0.295	0.327	0.212	0.300	0.329	0.281	0.273	0.358	0.340
cov2	0.000	0.000	0.000	0.000	0.000	0.116	0.232	0.000	0.083	0.160	0.105	0.116	0.249	0.269
cov3	0.000	0.000	0.000	0.049	0.000	0.144	0.256	0.000	0.117	0.178	0.135	0.148	0.263	0.290

§ 3.7 : Plant-Capture Applied to the Model M_h

§ 3.7.1 : Introduction

Within the previous sections of this chapter consideration was given to the standard problem of estimating population size from capture-recapture data, in the absence of plants. The aim of this section is to show how an initial insertion of planted individuals into the population prior to the beginning of the capture-recapture experiment can enhance point estimation of population size. This is done by deriving a Peterson-type estimator in section 3.7.4, and in section 3.7.5 it is shown how the coverage adjusted estimators of section 3.3 can be modified so as to enable them to utilize the information gained from planted individuals.

§ 3.7.2 : Sampling Procedure, Assumptions and Some Additional Notation

The sampling procedure considered within this section is almost identical to the one described in section 3.2. The only difference being that prior to the beginning of the experiment it is assumed that a known number of R pre-marked individuals have been mixed with the target population, of size N . It is assumed that the planted individuals behave in an identical manner to those of the target population.

Once the planted animals have mixed with the target population a sequence of t sampling experiments are carried out on the augmented population which is assumed to be closed and of size $N+R$. Independently of other animals and independently of its previous capture history animal i ($i=1,2,\dots,N+R$) is captured in sample j ($j=1,2,\dots,t$) with probability p_i . After each sample is taken every animal within that sample which has not previously been marked receives a unique tag before its immediate release so that it may be recognised on subsequent trapping occasions. The experiment generates an $(N+R)$ by t matrix A where

$$a_{ij} = \begin{cases} 1 & \text{if animal } i \text{ is caught on sampling occasion } j \\ 0 & \text{if animal } i \text{ is not caught on sampling occasion } j \end{cases}$$
$$i = 1, 2, \dots, N+R.$$
$$j = 1, 2, \dots, t.$$

The sample space is the set of such matrices.

It is assumed that the p_i , for $i= 1, 2, \dots, N+R$, are a random sample from some probability distribution $f(p)$, $p \in [0,1]$, with c.d.f. $F(p)$.

At this point it is necessary to introduce some additional notation .

As in chapters 1 and 2, let

$X_1 \equiv$ number of distinct animals seen from target population,

$X_2 \equiv$ number of distinct animals seen from planted population

and let $X = X_1 + X_2 \equiv$ number of distinct animals seen from augmented population.

Now the frequencies for the target, planted and augmented populations are written as

$f_k \equiv$ number of animals from target population seen exactly k times,

$f_k^* \equiv$ number of animals from planted population seen exactly k times

and $f_k^a = f_k + f_k^* \equiv$ number of animals from augmented population seen exactly k times.

Finally let

$$Z_a = \sum_{k=1}^t k f_k^a \equiv \text{total number of sightings made.}$$

§ 3.7.3 : Some Distribution Theory

By the independence of the target and planted populations, it follows from equation (3.1) that the joint probability distribution function of $\{X_1, X_2, f_1, f_2, \dots, f_t, f_1^*, f_2^*, \dots, f_t^*\}$ may be written as

$$\begin{aligned} \text{Prob}\{X_1, X_2, f_1, f_2, \dots, f_t, f_1^*, f_2^*, \dots, f_t^*\} = \\ \binom{N}{N - x_1, f_1, f_2, \dots, f_t} (\pi_0)^{N - x_1} \prod_{j=1}^t (\pi_j)^{f_j} \cdot \binom{R}{R - x_2, f_1^*, f_2^*, \dots, f_t^*} (\pi_0)^{R - x_2} \prod_{j=1}^t (\pi_j)^{f_j^*}, \quad (3.10) \end{aligned}$$

where $\pi_j = \int_0^1 \binom{t}{j} p^j (1-p)^{t-j} dF(p),$
 $j = 0, 1, 2, \dots, t.$

It can be shown (K. Pollock pers.com.) that this probability function may be decomposed as follows :

$$\text{Prob}\{X_1, X_2, f_1, f_2, \dots, f_t, f_1^*, f_2^*, \dots, f_t^*\} = P_1 \cdot P_2 \cdot P_3 \cdot P_4, \quad (3.11)$$

where $P_1 = \frac{\binom{N}{\sum f_i} \binom{R}{\sum f_i^*}}{\binom{N+R}{\sum f_i + \sum f_i^*}},$

$$P_2 = \left(\frac{N+R}{\sum f_i + \sum f_i^*} \right) (1 - \pi_0)^{\sum f_i + \sum f_i^*} (\pi_0)^{N+R - \sum f_i - \sum f_i^*},$$

$$P_3 = \left(\sum_{f_1 + f_1^*, \dots, f_t + f_t^*} \sum f_i + \sum f_i^* \right) \prod_{i=1}^t \left(\frac{\pi_i}{1 - \pi_0} \right)^{f_i + f_i^*}$$

and

$$P_4 = \frac{\binom{\sum f_i}{f_1, f_2, \dots, f_t} \binom{\sum f_i^*}{f_1^*, f_2^*, \dots, f_t^*}}{\binom{\sum f_i + \sum f_i^*}{f_1 + f_1^*, \dots, f_t + f_t^*}}.$$

N.B. Each summation is to be evaluated over the range $i = 1, 2, \dots, t$.

§ 3.7.4 : A Peterson-Type Estimator

The Peterson-type estimator proposed here is derived from the P_1 component of the probability function (3.11) : this P_1 component is in fact a hypergeometric density function. Let $L(N)$ denote the likelihood function for N based on this hypergeometric density. By equating $L(N)$ to $L(N - 1)$ one can show that the likelihood function for N , based on the hypergeometric distribution P_1 , is in fact maximised by the Peterson-type estimator $\tilde{N}_p = \frac{RX_1}{X_2}$. In order to avoid introducing an estimator which becomes infinite when $X_2 = 0$, and in view of the fact that population size N is integer valued, \tilde{N}_p is slightly modified : from this point consideration is given to the estimator $\hat{N}_p = \left[0.5 + \frac{(R+1)X_1}{X_2 + 1} \right]$, where $[.]$ denotes the integer part.

§ 3.7.5 : Plant-Capture Versions of the Overton and Coverage Adjusted Estimators

Of the estimators considered within this chapter that were initially designed for the standard problem of estimating population size in the absence of plants, other than the nonparametric maximum likelihood estimator of Norris and Pollock(1996), only the Overton estimator and coverage adjusted estimators of section 3.3 can be extended naturally in such a way that allows them to be able to utilize the extra information gained from the planted individuals. The plant-capture versions of the Overton and coverage adjusted estimators are obtained via an approach almost identical to the one used to derive them in the absence of plants : hence the following derivation is explained only briefly.

From equation 3.10 it follows that the likelihood function for N may be written as

$$\begin{aligned} L(N) &\propto \frac{N!}{(N-x_1)!} (\pi_0)^{N-x_1} \\ &= \frac{N!}{(N-x_1)!} \left\{ E[(1-p)^t] \right\}^{N-x_1}. \end{aligned}$$

If one assumes for the moment that $F(p)$ is known exactly, so that $E[(1-p)^t]$ can be viewed as a known constant, then, by equating $L(N)$ to $L(N-1)$, one can show that an approximate maximum likelihood estimate is given by

$$\hat{N} = \frac{x_1}{1 - E[(1-p)^t]}. \quad (3.12)$$

Using the theory of weighted distributions, and assuming that the capture probabilities of the animals seen during the experiment are known exactly, one can show that an unbiased estimator of $\left\{ 1 - E[(1-p)^t] \right\}^{-1}$ is given by $\frac{1}{x} \sum_{i \in S_x^a} \frac{1}{1 - (1-p_i)^t}$: the set

$S_x^a = \{s_k : k = 1, 2, \dots, x\}$, where $s_k \in \{1, 2, 3, \dots, N+R\}$ for all k , represents the set of the indices of the x distinct animals seen during the sampling period.

Now assuming that the p_i for all $i \in S_x^a$ are known, use of equation (3.12) suggests the estimator

$$\hat{N} = \frac{x_1}{x} \sum_{i \in S_x^a} \frac{1}{1 - (1-p_i)^t}. \quad (3.13)$$

A plant-capture version of the Overton estimator can now be obtained by substituting into equation (3.13) the maximum likelihood estimates of the p_i :

$$\hat{N}_0^a = \frac{x_1}{x} \sum_{i=1}^t \frac{f_i^a}{1 - \left(1 - \frac{i}{t}\right)^t}.$$

Similarly substituting coverage-adjusted estimates of the capture-probabilities into equation (3.13) yields the three following plant-capture versions of the coverage adjusted estimators of section 3.3 :

$$\hat{N}_{ca1}^a = \frac{x_1}{x} \sum_{i=1}^t \frac{f_i^a}{1 - \left(1 - \left(1 - \frac{f_1^a}{z_a}\right) \frac{i}{t}\right)^t},$$

$$\hat{N}_{ca2}^a = \frac{x_1}{x} \sum_{i=1}^t \frac{f_i^a}{1 - \left(1 - \left(1 - \frac{f_1^a}{z_a} + \frac{2}{(t-1)} \frac{f_2^a}{z_a}\right) \frac{i}{t}\right)^t}$$

and
$$\hat{N}_{ca3}^a = \frac{x_1}{x} \sum_{i=1}^t \frac{f_i^a}{1 - \left(1 - \left(1 - \frac{f_1^a}{z_a} + \frac{2}{(t-1)} \frac{f_2^a}{z_a} - \frac{6}{(t-1)(t-2)} \frac{f_3^a}{z_a}\right) \frac{i}{t}\right)^t}.$$

N.B. In the absence of plants, i.e. when $R = 0$, the estimators \hat{N}_0^a , \hat{N}_{ca1}^a , \hat{N}_{ca2}^a and \hat{N}_{ca3}^a do in fact reduce to \hat{N}_0 , \hat{N}_{ca1} , \hat{N}_{ca2} and \hat{N}_{ca3} respectively.

§ 3.7.6 : Plant-Capture Simulation Study

A simulation study was carried out in order to investigate the performance of the plant-capture estimators. The simulations were carried out in an almost identical manner to those of tables 3.5.3a, b and c. In each simulation the capture probabilities of the N+R animals were drawn as a random sample from some probability distribution p with mean $E(p)$, variance $\text{Var}(p)$ and coefficient of variation $\sqrt{\text{Var}(p)}/E(p)$. At the beginning of each simulation the value of $E(p)$ was selected as a random observation from a uniform distribution on some interval $(. , .)$. Similarly the value of the coefficient of variation was selected as a random observation from a uniform distribution on some interval $(. , .)$. The distribution of p was chosen to be Beta(alpha, beta), where, as in section 3.3, $\alpha = \frac{1 - ep}{(cv)^2} - ep$ and $\beta = \frac{(1 - ep)^2}{ep(cv)^2} - (1 - ep)$. The capture probabilities of the N+R animals were drawn as a random sample from this beta distribution and live trapping was then simulated in the usual way. Each table is split into two columns : each column depicting the results for a certain number of sampling occasions. Each column is split into four cells : the first cell shows how the estimators perform in the absence of planted individuals ; the second, third and fourth cells illustrate respectively the performance of the estimators in the presence of 10, 25 and 50 plants. For each value of t and R one thousand simulations were carried out : a different set of capture probabilities was used each time. The values shown in the tables are mostly averages. As many of the estimators are only finite if at least one recapture occurs, any data set not meeting this condition was discarded. The simulation procedure continued until one thousand data sets for which the condition did hold had been generated. This simulation study considers various target populations of size $N = 100$. The results of the simulations are presented in tables 3.7.1a,b, 3.7.2a,b, 3.7.3a,b, 3.7.4a,b, 3.7.5a,b and 3.7.6a,b. The notation used within these tables is identical to the notation of section 3.5, with the following additions :

$$x1 \quad \equiv \quad X_1.$$

$$x2 \quad \equiv \quad X_2.$$

$$\text{pet} \quad \equiv \quad \hat{N}_p.$$

mlea \equiv the maximum likelihood estimator for the model M_0 , for details

please refer to chapter 1.

$$\text{dr1a} \quad \equiv \quad \hat{N}_{0,1}^a = \frac{X_1}{1 - \frac{f_1^a}{Z_a}}.$$

$$Oa \equiv \hat{N}_0^a.$$

$$ca1a \equiv \hat{N}_{ca1}^a.$$

$$ca2a \equiv \hat{N}_{ca2}^a.$$

$$ca3a \equiv \hat{N}_{ca3}^a.$$

$$ca \equiv C^a = \frac{\sum_{i \in S_x^a} p_i}{\sum_{i=1}^{N+R} p_i}.$$

$$ch1a \equiv 1 - \frac{f_1^a}{z_a}.$$

$$ch2a \equiv 1 - \frac{f_1^a}{z_a} + \frac{2}{(t-1)} \frac{f_2^a}{z_a}.$$

$$ch3a \equiv 1 - \frac{f_1^a}{z_a} + \frac{2}{(t-1)} \frac{f_2^a}{z_a} - \frac{6}{(t-1)(t-2)} \frac{f_3^a}{z_a}.$$

Discussion

An important characteristic of the Peterson-type estimator \hat{N}_p is that it remains virtually unbiased in the presence of heterogeneity : consistent unbiasedness in the presence of heterogeneity is something no other existing estimator is able to achieve. However a major disadvantage of the Peterson-type estimator is that it can have a very large variance when the number of plants is small relative to the size of the target population.

The plant-capture versions of the Overton and coverage adjusted estimators can be seen to improve as the number of plants is increased from $R = 0$: the beneficial effect of the plants is most noticeable when the number of sampling occasions is small. However it is disappointing to observe that this improvement due to the plants is on the whole only marginal. There is scope for more work here since it is believed that the plants may be used in a more efficient way to improve point estimation in this situation, perhaps through considering an estimator in the form of a weighted average of the Peterson-type and coverage adjusted estimators.

Table 3.7.3b

N = 100 : p ~ Beta(alpha,beta) : ep ~ U(0.08, 0.12) : cv ~ U(0.30, 0.80)
 Number of simulations = 1000

t = 15					t = 20				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	71.70	-28.30	6.528	29.045	x1	79.35	-20.65	6.281	21.588
mle	84.94	-15.06	8.768	17.431	mle	87.11	-12.89	7.556	14.938
dr1	90.67	-9.33	9.164	13.078	dr1	91.64	-8.36	7.466	11.206
boot	85.25	-14.75	7.544	16.568	boot	91.83	-8.17	6.966	10.734
ac1	97.64	-2.36	11.379	11.621	ac1	97.50	-2.50	8.449	8.811
ac2	94.25	-5.75	10.762	12.204	ac2	95.76	-4.24	8.147	9.185
ac3	94.67	-5.33	10.825	12.067	ac3	95.95	-4.05	8.168	9.115
O	91.47	-8.53	8.088	11.756	O	97.35	-2.65	7.357	7.820
ca1	101.95	1.95	10.280	10.464	ca1	103.02	3.02	8.500	9.020
ca2	100.69	0.69	10.008	10.031	ca2	102.46	2.46	8.355	8.710
ca3	100.85	0.85	10.031	10.067	ca3	102.52	2.52	8.360	8.731
pojac	116.79	16.79	13.402	21.481	pojac	114.19	14.19	12.240	18.737
jac1	99.74	-0.26	9.109	9.112	jac1	103.83	3.83	8.057	8.920
jacseq	105.99	5.99	15.362	16.489	jacseq	105.96	5.96	11.286	12.762
jacint	101.92	1.92	12.158	12.309	jacint	104.58	4.58	9.517	10.562
Pr(inf mle), mean jackknife order = 0.000, 1.401 c, ch1, ch2, ch3 = 0.813, 0.793, 0.813, 0.811					Pr(inf mle), mean jackknife order = 0.000, 1.141 c, ch1, ch2, ch3 = 0.879, 0.867, 0.878, 0.877				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	7.22				x2	7.91			
pet	99.21	-0.79	20.766	20.782	pet	100.14	0.14	16.522	16.523
mlea	85.54	-14.46	8.519	16.779	mlea	87.45	-12.55	7.409	14.570
dr1a	90.56	-9.44	8.968	13.020	dr1a	91.62	-8.38	7.395	11.178
Oa	91.46	-8.54	8.018	11.717	Oa	97.34	-2.66	7.365	7.829
cala	101.86	1.86	10.058	10.228	cala	102.97	2.97	8.414	8.923
ca2a	100.60	0.60	9.772	9.790	ca2a	102.40	2.40	8.278	8.618
ca3a	100.75	0.75	9.808	9.836	ca3a	102.45	2.45	8.300	8.655
ca, ch1a, ch2a, ch3a = 0.814, 0.794, 0.814, 0.811					ca, ch1a, ch2a, ch3a = 0.879, 0.867, 0.879, 0.878				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	17.90				x2	19.68			
pet	99.99	-0.01	14.464	14.464	pet	100.18	0.18	11.612	11.613
mlea	86.27	-13.73	8.640	16.219	mlea	87.50	-12.50	7.350	14.499
dr1a	90.32	-9.68	9.155	13.325	dr1a	91.14	-8.86	7.397	11.540
Oa	91.23	-8.77	8.042	11.898	Oa	96.88	-3.12	7.374	8.007
cala	101.60	1.60	10.279	10.402	cala	102.50	2.50	8.375	8.739
ca2a	100.31	0.31	9.945	9.950	ca2a	101.92	1.92	8.228	8.449
ca3a	100.47	0.47	9.981	9.992	ca3a	101.99	1.99	8.247	8.483
ca, ch1a, ch2a, ch3a = 0.814, 0.794, 0.815, 0.812					ca, ch1a, ch2a, ch3a = 0.876, 0.867, 0.878, 0.877				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	35.73				x2	39.71			
pet	99.29	-0.71	11.155	11.178	pet	100.26	0.26	9.184	9.188
mlea	86.14	-13.86	8.132	16.074	mlea	88.35	-11.65	6.725	13.454
dr1a	89.29	-10.71	8.831	13.879	dr1a	91.56	-8.44	6.842	10.865
Oa	90.38	-9.62	8.049	12.541	Oa	97.49	-2.51	6.923	7.364
cala	100.43	0.43	9.886	9.896	cala	102.91	2.91	7.729	8.260
ca2a	99.21	-0.79	9.626	9.658	ca2a	102.36	2.36	7.598	7.956
ca3a	99.34	-0.66	9.642	9.664	ca3a	102.41	2.41	7.598	7.971
ca, ch1a, ch2a, ch3a = 0.813, 0.797, 0.817, 0.814					ca, ch1a, ch2a, ch3a = 0.880, 0.870, 0.881, 0.880				

Table 3.7.4a

N = 100 : p ~ Beta(alpha,beta) : ep ~ U(0.04, 0.08) : cv ~ U(0.55, 0.80)
 Number of simulations = 1000

t = 5					t = 10				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	25.13	-74.87	5.646	75.081	x1	41.52	-58.48	7.378	58.942
mle	83.32	-16.68	44.193	47.234	mle	76.73	-23.27	22.101	32.090
dr1	106.12	6.12	57.681	58.005	dr1	86.02	-13.98	25.424	29.016
boot	32.26	-67.74	7.137	68.111	boot	52.48	-47.52	9.017	48.366
ac1	128.60	28.60	78.620	83.662	ac1	96.75	-3.25	32.547	32.708
ac2	96.33	-3.67	61.388	61.498	ac2	87.31	-12.69	28.624	31.311
ac3	101.87	1.87	75.572	75.595	ac3	88.26	-11.74	29.039	31.322
O	35.62	-64.38	7.869	64.863	O	57.88	-42.12	9.786	43.243
ca1	109.48	9.48	57.940	58.710	ca1	93.33	-6.67	25.933	26.778
ca2	92.24	-7.76	46.918	47.556	ca2	88.60	-11.40	23.676	26.276
ca3	93.90	-6.10	48.028	48.414	ca3	89.03	-10.97	23.782	26.190
pojac	60.00	-40.00	13.165	42.111	pojac	90.61	-9.39	15.454	18.084
jac1	41.88	-58.12	9.174	58.845	jac1	66.98	-33.02	11.147	34.848
jacseq	62.17	-37.83	14.297	40.440	jacseq	84.95	-15.05	20.771	25.652
jacint	58.88	-41.12	13.681	43.338	jacint	78.85	-21.15	19.455	28.735
Pr(inf mle), mean jackknife order = 0.031, 3.808 c, ch1, ch2, ch3 = 0.347, 0.286, 0.345, 0.338					Pr(inf mle), mean jackknife order = 0.000, 2.256 c, ch1, ch2, ch3 = 0.546, 0.507, 0.542, 0.539				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	2.54				x2	4.11			
pet	94.87	-5.13	54.642	54.883	pet	100.88	0.88	48.008	48.016
mlea	89.98	-10.02	46.376	47.446	mlea	80.12	-19.88	19.218	27.649
dr1a	104.75	4.75	57.335	57.532	dr1a	85.19	-14.81	23.460	27.742
Oa	35.62	-64.38	7.871	64.856	Oa	57.88	-42.12	9.785	43.240
cala	108.20	8.20	57.569	58.150	cala	92.52	-7.48	24.000	25.138
ca2a	91.15	-8.85	46.319	47.157	ca2a	87.88	-12.12	21.985	25.103
ca3a	92.76	-7.24	46.960	47.514	ca3a	88.29	-11.71	22.087	24.998
ca, ch1a, ch2a, ch3a = 0.347, 0.286, 0.346, 0.339					ca, ch1a, ch2a, ch3a = 0.546, 0.509, 0.544, 0.540				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	6.26				x2	10.45			
pet	100.41	0.41	47.373	47.375	pet	99.42	-0.58	28.889	28.895
mlea	90.31	-9.69	31.996	33.431	mlea	82.43	-17.57	16.469	24.080
dr1a	100.96	0.96	54.172	54.181	dr1a	83.85	-16.15	18.917	24.874
Oa	35.57	-64.43	7.736	64.895	Oa	57.87	-42.13	10.015	43.300
cala	104.40	4.40	54.346	54.524	cala	91.19	-8.81	19.623	21.511
ca2a	87.95	-12.05	43.902	45.525	ca2a	86.60	-13.40	18.155	22.563
ca3a	89.43	-10.57	44.728	45.961	ca3a	87.01	-12.99	18.225	22.382
ca, ch1a, ch2a, ch3a = 0.349, 0.293, 0.355, 0.348					ca, ch1a, ch2a, ch3a = 0.545, 0.509, 0.545, 0.541				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	12.63				x2	21.00			
pet	99.24	-0.76	32.789	32.798	pet	99.10	-0.90	19.977	19.997
mlea	91.20	-8.80	28.550	29.877	mlea	84.37	-15.63	14.276	21.165
dr1a	97.46	-2.54	52.389	52.451	dr1a	82.96	-17.04	17.657	24.541
Oa	35.48	-64.53	8.121	65.034	Oa	58.06	-41.94	9.709	43.049
cala	100.89	0.89	52.600	52.607	cala	90.32	-9.68	18.386	20.778
ca2a	85.32	-14.68	42.695	45.148	ca2a	85.91	-14.09	17.058	22.122
ca3a	86.96	-13.04	43.620	45.528	ca3a	86.31	-13.69	17.138	21.936
ca, ch1a, ch2a, ch3a = 0.348, 0.298, 0.359, 0.352					ca, ch1a, ch2a, ch3a = 0.548, 0.515, 0.551, 0.547				

Table 3.7.4b

N = 100 : p ~ Beta(alpha,beta) : ep ~ U(0.04, 0.08) : cv ~ U(0.55, 0.80)
 Number of simulations = 1000

t = 15					t = 20				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	53.06	-46.94	7.651	47.555	x1	61.93	-38.07	7.829	38.869
mle	76.08	-23.92	11.391	26.496	mle	78.22	-21.78	9.050	23.586
dr1	82.63	-17.37	12.800	21.579	dr1	83.92	-16.08	10.079	18.978
boot	65.60	-34.40	8.879	35.529	boot	74.98	-25.02	8.763	26.511
ac1	91.40	-8.60	17.943	19.897	ac1	92.30	-7.70	14.540	16.451
ac2	86.73	-13.27	16.738	21.360	ac2	89.56	-10.44	13.988	17.456
ac3	87.12	-12.88	16.855	21.214	ac3	89.78	-10.22	14.058	17.383
O	71.68	-28.32	9.476	29.864	O	81.19	-18.81	9.251	20.963
ca1	92.20	-7.80	13.735	15.793	ca1	94.75	-5.25	11.230	12.395
ca2	90.19	-9.81	13.219	16.465	ca2	93.72	-6.28	11.032	12.697
ca3	90.36	-9.64	13.293	16.421	ca3	93.80	-6.20	11.043	12.667
pojac	103.49	3.49	14.214	14.635	pojac	109.63	9.63	14.275	17.217
jac1	81.01	-18.99	10.484	21.695	jac1	90.07	-9.93	9.996	14.087
jacseq	96.13	-3.87	18.914	19.305	jacseq	100.50	0.50	18.338	18.344
jacint	88.51	-11.49	16.990	20.512	jacint	94.56	-5.44	15.393	16.328
Pr(inf mle), mean jackknife order = 0.000, 1.929 c, ch1, ch2, ch3 = 0.671, 0.650, 0.672, 0.670					Pr(inf mle), mean jackknife order = 0.000, 1.620 c, ch1, ch2, ch3 = 0.758, 0.742, 0.757, 0.755				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	5.35				x2	6.22			
pet	99.64	-0.36	36.714	36.716	pet	99.51	-0.49	26.940	26.945
mlea	78.30	-21.70	11.045	24.347	mlea	79.65	-20.35	8.838	22.183
dr1a	82.42	-17.58	12.375	21.502	dr1a	83.86	-16.14	9.682	18.826
Oa	71.67	-28.33	9.456	29.870	Oa	81.22	-18.78	9.206	20.916
cala	91.98	-8.02	13.328	15.554	cala	94.69	-5.31	10.803	12.040
ca2a	89.95	-10.05	12.864	16.326	ca2a	93.62	-6.38	10.629	12.396
ca3a	90.12	-9.88	12.892	16.241	ca3a	93.69	-6.31	10.638	12.369
ca, ch1a, ch2a, ch3a = 0.671, 0.651, 0.673, 0.671					ca, ch1a, ch2a, ch3a = 0.758, 0.742, 0.757, 0.756				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	13.32				x2	15.43			
pet	99.94	-0.06	22.562	22.562	pet	99.98	-0.02	18.181	18.181
mlea	80.92	-19.08	10.596	21.829	mlea	81.03	-18.97	8.450	20.769
dr1a	83.02	-16.98	11.600	20.561	dr1a	83.68	-16.32	9.198	18.734
Oa	71.94	-28.06	9.693	29.692	Oa	81.07	-18.93	9.037	20.980
cala	92.65	-7.35	12.590	14.579	cala	94.56	-5.44	10.291	11.639
ca2a	90.61	-9.39	12.164	15.368	ca2a	93.48	-6.52	10.112	12.034
ca3a	90.78	-9.22	12.206	15.298	ca3a	93.56	-6.44	10.120	11.994
ca, ch1a, ch2a, ch3a = 0.671, 0.646, 0.668, 0.666					ca, ch1a, ch2a, ch3a = 0.756, 0.740, 0.755, 0.754				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	26.74				x2	30.65			
pet	99.48	-0.52	16.772	16.780	pet	99.98	-0.02	13.508	13.508
mlea	82.32	-17.68	10.042	20.329	mlea	82.31	-17.69	8.528	19.635
dr1a	82.15	-17.85	11.490	21.231	dr1a	83.12	-16.88	9.332	19.287
Oa	71.83	-28.17	9.601	29.759	Oa	80.52	-19.48	9.272	21.572
cala	91.76	-8.24	12.455	14.934	cala	93.91	-6.09	10.458	12.102
ca2a	89.78	-10.22	12.066	15.814	ca2a	92.84	-7.16	10.301	12.542
ca3a	89.95	-10.05	12.104	15.731	ca3a	92.93	-7.07	10.319	12.508
ca, ch1a, ch2a, ch3a = 0.674, 0.652, 0.675, 0.673					ca, ch1a, ch2a, ch3a = 0.754, 0.740, 0.755, 0.754				

Table 3.7.5a

N = 100 : p ~ Beta(alpha,beta) : ep ~ U(0.08, 0.12) : cv ~ U(0.30, 0.55)
 Number of simulations = 1000

t = 5					t = 10				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	39.50	-60.50	5.928	60.793	x1	61.49	-38.51	6.317	39.025
mle	91.54	-8.46	33.475	34.528	mle	88.85	-11.15	11.354	15.912
dr1	112.01	12.01	43.487	45.114	dr1	97.02	-2.98	13.174	13.506
boot	50.05	-49.95	7.381	50.488	boot	75.80	-24.20	7.515	25.337
ac1	130.78	30.78	59.479	66.970	ac1	104.96	4.96	18.056	18.725
ac2	98.74	-1.26	43.382	43.400	ac2	96.15	-3.85	15.919	16.377
ac3	103.28	3.28	48.258	48.370	ac3	97.24	-2.76	16.232	16.465
O	54.94	-45.06	8.079	45.777	O	82.57	-17.43	8.112	19.225
ca1	117.25	17.25	43.769	47.046	ca1	107.17	7.17	14.022	15.750
ca2	100.63	0.63	35.581	35.586	ca2	103.15	3.15	13.117	13.490
ca3	102.60	2.60	36.305	36.398	ca3	103.63	3.63	13.234	13.724
pojac	88.86	-11.14	13.502	17.507	pojac	118.47	18.47	13.855	23.091
jac1	63.77	-36.23	9.354	37.420	jac1	93.10	-6.90	9.277	11.562
jacseq	90.09	-9.91	15.305	18.234	jacseq	109.47	9.47	18.321	20.625
jacint	85.96	-14.04	14.820	20.415	jacint	101.08	1.08	16.728	16.763
Pr(inf mle) , mean jackknife order = 0.000, 3.736 c, ch1, ch2, ch3 = 0.451, 0.384, 0.461, 0.449					Pr(inf mle) , mean jackknife order = 0.000, 2.001 c, ch1, ch2, ch3 = 0.680, 0.640, 0.680, 0.674				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	3.94				x2	6.19			
pet	98.41	-1.59	40.982	41.013	pet	99.43	-0.57	28.421	28.427
mlea	92.29	-7.71	24.523	25.706	mlea	89.71	-10.29	10.538	14.727
dr1a	110.54	10.54	37.685	39.132	dr1a	96.78	-3.22	12.609	13.013
Oa	54.95	-45.05	8.050	45.762	Oa	82.57	-17.43	8.126	19.228
ca1a	115.76	15.76	38.019	41.155	ca1a	106.92	6.92	13.507	15.176
ca2a	99.44	-0.56	30.738	30.743	ca2a	102.93	2.93	12.685	13.019
ca3a	101.38	1.38	31.222	31.252	ca3a	103.43	3.43	12.762	13.215
ca, ch1a, ch2a, ch3a = 0.451, 0.384, 0.461, 0.450					ca, ch1a, ch2a, ch3a = 0.681, 0.641, 0.681, 0.675				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	9.89				x2	15.47			
pet	99.69	-0.31	37.212	37.213	pet	99.28	-0.72	16.695	16.711
mlea	94.41	-5.59	21.601	22.313	mlea	90.48	-9.52	9.721	13.606
dr1a	112.11	12.11	38.864	40.706	dr1a	96.20	-3.80	11.777	12.375
Oa	54.76	-45.24	8.097	45.964	Oa	82.52	-17.48	7.932	19.200
ca1a	117.35	17.35	39.245	42.909	ca1a	106.30	6.30	12.621	14.108
ca2a	100.75	0.75	32.068	32.077	ca2a	102.35	2.35	11.829	12.060
ca3a	102.80	2.80	32.761	32.880	ca3a	102.84	2.84	11.935	12.268
ca, ch1a, ch2a, ch3a = 0.449, 0.375, 0.450, 0.439					ca, ch1a, ch2a, ch3a = 0.681, 0.645, 0.685, 0.679				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	19.66				x2	30.84			
pet	99.25	-0.75	21.666	21.679	pet	99.92	-0.08	13.132	13.132
mlea	94.42	-5.58	17.742	18.599	mlea	91.79	-8.21	9.219	12.343
dr1a	108.84	8.84	30.455	31.712	dr1a	96.46	-3.54	11.542	12.072
Oa	54.50	-45.50	8.044	46.202	Oa	82.75	-17.25	8.042	19.029
ca1a	114.07	14.07	30.822	33.882	ca1a	106.62	6.62	12.391	14.047
ca2a	98.04	-1.96	25.133	25.210	ca2a	102.64	2.64	11.642	11.939
ca3a	99.97	-0.03	25.582	25.582	ca3a	103.12	3.12	11.726	12.134
ca, ch1a, ch2a, ch3a = 0.446, 0.377, 0.452, 0.441					ca, ch1a, ch2a, ch3a = 0.681, 0.644, 0.683, 0.678				

Table 3.7.5b

N = 100 : p ~ Beta(alpha,beta) : ep ~ U(0.08, 0.12) : cv ~ U(0.30, 0.55)
 Number of simulations = 1000

t = 15					t = 20				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	74.72	-25.28	5.570	25.889	x1	82.41	-17.59	5.161	18.331
mle	90.47	-9.53	6.899	11.761	mle	91.71	-8.29	5.494	9.942
drl	95.68	-4.32	7.707	8.835	drl	95.66	-4.34	5.876	7.305
boot	89.09	-10.91	6.190	12.540	boot	95.51	-4.49	5.476	7.084
ac1	100.92	0.92	10.348	10.388	ac1	99.68	-0.32	7.404	7.411
ac2	97.15	-2.85	9.738	10.146	ac2	97.74	-2.26	7.158	7.506
ac3	97.58	-2.42	9.814	10.107	ac3	97.95	-2.05	7.214	7.499
O	95.68	-4.32	6.587	7.876	O	101.20	1.20	5.783	5.907
ca1	107.47	7.47	8.673	11.449	ca1	107.40	7.40	6.822	10.068
ca2	105.99	5.99	8.428	10.337	ca2	106.71	6.71	6.739	9.507
ca3	106.15	6.15	8.445	10.448	ca3	106.77	6.77	6.744	9.558
pojac	121.64	21.64	12.742	25.110	pojac	117.08	17.08	11.998	20.876
jac1	104.28	4.28	7.594	8.719	jac1	107.61	7.61	6.666	10.116
jacseq	110.08	10.08	13.984	17.236	jacseq	109.05	9.05	9.280	12.960
jacint	106.30	6.30	10.537	12.277	jacint	108.02	8.02	7.389	10.908
Pr(inf mle) , mean jackknife order = 0.000, 1.355 c, ch1, ch2, ch3 = 0.806, 0.783, 0.805, 0.802					Pr(inf mle) , mean jackknife order = 0.000, 1.097 c, ch1, ch2, ch3 = 0.875, 0.863, 0.875, 0.874				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	7.48				x2	8.27			
pet	99.88	-0.12	19.983	19.983	pet	99.73	-0.27	15.544	15.546
mlea	90.89	-9.11	6.628	11.268	mlea	91.99	-8.01	5.399	9.659
drla	95.59	-4.41	7.444	8.654	drla	95.69	-4.31	5.762	7.198
Oa	95.66	-4.34	6.512	7.823	Oa	101.26	1.26	5.737	5.874
cala	107.41	7.41	8.376	11.180	cala	107.44	7.44	6.674	9.996
ca2a	105.89	5.89	8.144	10.051	ca2a	106.78	6.78	6.583	9.451
ca3a	106.07	6.07	8.152	10.165	ca3a	106.84	6.84	6.592	9.499
ca, ch1a, ch2a, ch3a = 0.806, 0.784, 0.805, 0.803					ca, ch1a, ch2a, ch3a = 0.875, 0.862, 0.875, 0.873				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	18.48				x2	20.50			
pet	99.94	-0.06	13.677	13.677	pet	100.81	0.81	10.543	10.575
mlea	90.74	-9.26	7.049	11.639	mlea	92.31	-7.69	5.269	9.319
drla	94.64	-5.36	7.798	9.461	drla	95.60	-4.40	5.572	7.102
Oa	94.60	-5.40	6.863	8.731	Oa	101.39	1.39	5.659	5.826
cala	106.32	6.32	8.750	10.795	cala	107.32	7.32	6.433	9.745
ca2a	104.81	4.81	8.509	9.775	ca2a	106.69	6.69	6.344	9.217
ca3a	104.99	4.99	8.515	9.869	ca3a	106.76	6.76	6.358	9.279
ca, ch1a, ch2a, ch3a = 0.800, 0.782, 0.804, 0.802					ca, ch1a, ch2a, ch3a = 0.877, 0.866, 0.878, 0.877				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	37.37				x2	41.14			
pet	99.77	-0.23	9.723	9.726	pet	100.14	0.14	8.147	8.148
mlea	91.78	-8.22	6.279	10.346	mlea	92.44	-7.56	5.150	9.148
drla	95.33	-4.67	6.983	8.400	drla	95.57	-4.43	5.410	6.990
Oa	95.50	-4.50	6.599	7.990	Oa	101.24	1.24	5.561	5.696
cala	107.11	7.11	7.860	10.600	cala	107.34	7.34	6.241	9.634
ca2a	105.64	5.64	7.660	9.512	ca2a	106.69	6.69	6.152	9.091
ca3a	105.81	5.81	7.679	9.631	ca3a	106.75	6.75	6.167	9.142
ca, ch1a, ch2a, ch3a = 0.805, 0.784, 0.806, 0.803					ca, ch1a, ch2a, ch3a = 0.875, 0.863, 0.875, 0.874				

Table 3.7.6a

$N = 100 : p \sim U(0, 0.16) : E(p) = 0.08 : \sqrt{Var(p)}/E(p) = 0.5774$
 Number of simulations = 1000

t = 5					t = 10				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	32.44	-67.56	4.510	67.706	x1	51.70	-48.30	5.086	48.572
mle	83.09	-16.91	35.492	39.317	mle	78.57	-21.43	11.588	24.363
dr1	102.80	2.80	46.371	46.455	dr1	86.01	-13.99	13.362	19.343
boot	41.33	-58.67	5.789	58.952	boot	64.19	-35.81	6.389	36.374
ac1	120.86	20.86	63.997	67.312	ac1	92.88	-7.12	17.699	19.079
ac2	91.23	-8.77	48.615	49.400	ac2	84.67	-15.33	15.402	21.731
ac3	95.56	-4.44	58.350	58.519	ac3	85.65	-14.35	15.681	21.259
O	45.49	-54.51	6.427	54.888	O	70.19	-29.81	7.100	30.649
ca1	107.11	7.11	46.654	47.193	ca1	94.69	-5.31	14.201	15.161
ca2	91.29	-8.71	38.065	39.049	ca2	90.80	-9.20	13.240	16.121
ca3	92.98	-7.02	39.085	39.711	ca3	91.26	-8.74	13.323	15.936
pojac	74.64	-25.36	11.693	27.925	pojac	103.01	3.01	13.154	13.494
jac1	52.99	-47.01	7.659	47.628	jac1	79.66	-20.34	8.460	22.028
jacseq	76.15	-23.85	13.895	27.606	jacseq	95.57	-4.43	17.518	18.070
jacint	72.40	-27.60	13.428	30.696	jacint	87.79	-12.21	16.026	20.147
Pr(inf mle), mean jackknife order = 0.001, 3.723 c, ch1, ch2, ch3 = 0.421, 0.352, 0.425, 0.416					Pr(inf mle), mean jackknife order = 0.000, 2.056 c, ch1, ch2, ch3 = 0.648, 0.609, 0.649, 0.644				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	3.14				x2	5.05			
pet	103.40	3.40	58.831	58.929	pet	101.92	1.92	33.834	33.888
mlea	87.47	-12.53	28.244	30.897	mlea	81.32	-18.68	11.408	21.885
dr1a	101.37	1.37	43.967	43.988	dr1a	86.00	-14.00	12.963	19.079
Oa	45.48	-54.52	6.412	54.898	Oa	70.21	-29.79	7.103	30.628
cala	105.75	5.75	44.266	44.638	cala	94.69	-5.31	13.819	14.806
ca2a	90.22	-9.78	36.486	37.775	ca2a	90.81	-9.19	12.915	15.851
ca3a	91.96	-8.04	37.809	38.655	ca3a	91.25	-8.75	12.994	15.663
ca, ch1a, ch2a, ch3a = 0.420, 0.354, 0.427, 0.417					ca, ch1a, ch2a, ch3a = 0.647, 0.609, 0.648, 0.644				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	8.10				x2	12.91			
pet	99.88	-0.12	33.742	33.743	pet	99.50	-0.50	23.299	23.305
mlea	89.82	-10.18	23.650	25.749	mlea	82.33	-17.67	10.864	20.741
dr1a	101.24	1.24	42.967	42.985	dr1a	84.53	-15.47	11.978	19.565
Oa	45.45	-54.55	6.520	54.941	Oa	69.55	-30.45	6.835	31.212
cala	105.62	5.62	43.250	43.613	cala	93.10	-6.90	12.756	14.504
ca2a	90.09	-9.91	35.090	36.464	ca2a	89.29	-10.71	11.935	16.039
ca3a	91.71	-8.29	35.623	36.575	ca3a	89.74	-10.26	11.994	15.782
ca, ch1a, ch2a, ch3a = 0.420, 0.350, 0.422, 0.412					ca, ch1a, ch2a, ch3a = 0.647, 0.613, 0.653, 0.648				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	16.19				x2	25.76			
pet	99.31	-0.69	24.791	24.801	pet	99.60	-0.40	16.670	16.675
mlea	90.48	-9.52	19.390	21.600	mlea	84.83	-15.17	10.135	18.240
dr1a	97.25	-2.75	33.498	33.611	dr1a	84.81	-15.19	10.997	18.750
Oa	45.20	-54.80	6.424	55.178	Oa	69.70	-30.30	6.424	30.974
cala	101.58	1.58	33.847	33.884	cala	93.43	-6.57	11.751	13.462
ca2a	86.85	-13.15	27.855	30.803	ca2a	89.61	-10.39	11.034	15.157
ca3a	88.52	-11.48	28.619	30.834	ca3a	90.06	-9.94	11.087	14.892
ca, ch1a, ch2a, ch3a = 0.418, 0.354, 0.427, 0.417					ca, ch1a, ch2a, ch3a = 0.646, 0.610, 0.650, 0.646				

Table 3.7.6b

N = 100 : p ~ U(0, 0.16) : E(p) = 0.08 : sqrt[Var(p)]/E(p) = 0.5774

Number of simulations = 1000

t = 15					t = 20				
R = 0					R = 0				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x1	62.94	-37.06	5.041	37.405	x1	71.08	-28.92	4.479	29.268
mle	78.77	-21.23	7.575	22.540	mle	80.82	-19.18	5.664	19.997
dr1	83.72	-16.28	8.525	18.374	dr1	85.05	-14.95	6.230	16.199
boot	75.82	-24.18	6.172	24.958	boot	83.42	-16.58	5.330	17.419
ac1	88.59	-11.41	11.476	16.182	ac1	89.41	-10.59	8.233	13.415
ac2	84.97	-15.03	10.803	18.508	ac2	87.52	-12.48	8.021	14.837
ac3	85.38	-14.62	10.882	18.226	ac3	87.72	-12.28	8.050	14.682
O	81.78	-18.22	6.803	19.449	O	88.93	-11.07	5.844	12.514
ca1	94.05	-5.95	9.554	11.255	ca1	95.86	-4.14	7.232	8.333
ca2	92.56	-7.44	9.305	11.915	ca2	95.19	-4.81	7.140	8.607
ca3	92.74	-7.26	9.336	11.829	ca3	95.26	-4.74	7.146	8.576
pojac	107.36	7.36	12.925	14.874	pojac	107.75	7.75	11.872	14.178
jac1	89.92	-10.08	8.139	12.952	jac1	95.78	-4.22	6.958	8.138
jacseq	96.28	-3.72	14.260	14.737	jacseq	98.52	-1.48	11.372	11.469
jacint	91.92	-8.08	10.884	13.557	jacint	96.78	-3.22	8.764	9.336
Pr(inf mle) , mean jackknife order = 0.000, 1.429 c, ch1, ch2, ch3 = 0.772, 0.755, 0.777, 0.775					Pr(inf mle) , mean jackknife order = 0.000, 1.181 c, ch1, ch2, ch3 = 0.850, 0.837, 0.850, 0.849				
R = 10					R = 10				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	6.35				x2	7.09			
pet	99.40	-0.60	27.989	27.996	pet	100.02	0.02	21.859	21.859
mlea	80.07	-19.93	7.725	21.374	mlea	81.56	-18.44	5.611	19.271
dr1a	83.66	-16.34	8.465	18.402	dr1a	85.04	-14.96	6.111	16.158
Oa	81.78	-18.22	6.815	19.449	Oa	88.95	-11.05	5.777	12.466
cala	94.00	-6.00	9.493	11.232	cala	95.85	-4.15	7.067	8.197
ca2a	92.50	-7.50	9.231	11.893	ca2a	95.18	-4.82	6.963	8.469
ca3a	92.68	-7.32	9.242	11.790	ca3a	95.23	-4.77	6.983	8.455
ca, ch1a, ch2a, ch3a = 0.772, 0.755, 0.778, 0.775					ca, ch1a, ch2a, ch3a = 0.850, 0.837, 0.850, 0.849				
R = 25					R = 25				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	15.82				x2	17.83			
pet	100.55	0.55	17.111	17.120	pet	99.77	-0.23	14.789	14.790
mlea	82.61	-17.39	7.202	18.821	mlea	82.55	-17.45	6.138	18.494
dr1a	84.74	-15.26	7.818	17.144	dr1a	85.06	-14.94	6.701	16.377
Oa	82.76	-17.24	6.232	18.334	Oa	88.98	-11.02	6.375	12.727
cala	95.19	-4.81	8.786	10.017	cala	95.88	-4.12	7.713	8.743
ca2a	93.66	-6.34	8.522	10.622	ca2a	95.22	-4.78	7.625	9.001
ca3a	93.85	-6.15	8.548	10.532	ca3a	95.29	-4.71	7.633	8.968
ca, ch1a, ch2a, ch3a = 0.776, 0.753, 0.776, 0.773					ca, ch1a, ch2a, ch3a = 0.849, 0.837, 0.850, 0.849				
R = 50					R = 50				
Estimator	mean	bias	s.d.	rmse	Estimator	mean	bias	s.d.	rmse
x2	31.74				x2	35.76			
pet	99.89	-0.11	13.293	13.293	pet	99.23	-0.77	10.608	10.636
mlea	83.67	-16.33	7.210	17.854	mlea	83.33	-16.67	5.672	17.606
dr1a	84.35	-15.65	7.592	17.397	dr1a	85.03	-14.97	6.039	16.144
Oa	82.43	-17.57	6.186	18.625	Oa	88.91	-11.09	5.719	12.480
cala	94.79	-5.21	8.514	9.983	cala	95.87	-4.13	6.944	8.081
ca2a	93.28	-6.72	8.290	10.674	ca2a	95.17	-4.83	6.883	8.406
ca3a	93.45	-6.55	8.316	10.586	ca3a	95.25	-4.75	6.886	8.367
ca, ch1a, ch2a, ch3a = 0.776, 0.753, 0.776, 0.773					ca, ch1a, ch2a, ch3a = 0.850, 0.836, 0.849, 0.847				

Chapter 4 : Estimation Under The Capture-Recapture Model M_h : Continuous Time Sampling Procedure

§ 4.1 : Introduction

Within this chapter a new estimator of population size is proposed for a continuous time sampling procedure. The population in question is assumed to behave according to a continuous time analogue of the standard capture-recapture model M_h . This chapter concentrates on the standard estimation problem when no plants are used.

The sampling procedure considered here is identical to the one of chapter 2. But whereas within chapter 2 it was assumed that each animal in the population behaved in exactly the same way, this strong condition is no longer imposed. As was the case in chapter 3 however, a link between the behaviour of the animals is still needed and this is described below.

The estimation problem considered here has previously been studied in a software reliability context by Chao, Ma and Yang(1993) and more recently, in a capture-recapture context, by Yip and Chao(1996).

§ 4.2 : Sampling Procedure, Assumptions, Some Notation and the Sufficient Statistics

It is assumed that one animal is seen at a time and that animals seen for the first time receive a unique tag so that they may be recognised if subsequently recaptured. Sampling stops after a fixed predetermined amount of time τ . It is assumed that there are N animals in the population and that each animal is seen according to a Poisson process with rate λ_i , $i = 1, 2, \dots, N$. It is further assumed that $\lambda_i = k \cdot u_i$, for $i = 1, 2, \dots, N$, where k is a constant and the u_i form a random sample from some probability distribution with c.d.f. $F(u)$, $u \in [0,1]$. Detection times for different animals are assumed to be independent.

Let

$X \equiv$ number of distinct animals seen in time $(0, \tau)$.

$Z \equiv$ total number of sightings made in time $(0, \tau)$.

$X_i \equiv$ number of sightings of the i th animal, $i = 1, 2, \dots, N$.

$f_k = \sum_{i=1}^N I(X_i = k) \equiv$ number of animals seen exactly k times, $k = 0, 1, 2, \dots$

$m \equiv$ the most number of times any one particular animal was seen.

$$\bar{\lambda} = \frac{1}{N} \sum_{i=1}^N \lambda_i$$

$$\gamma = \frac{1}{\bar{\lambda}} \left\{ \frac{1}{N} \sum_{i=1}^N (\lambda_i - \bar{\lambda})^2 \right\}^{1/2} \equiv \text{the coefficient of variation .}$$

$$C = \frac{1}{N\bar{\lambda}} \sum_{i=1}^N \lambda_i I(X_i > 0) \equiv \text{sample coverage.}$$

$$p_i = \frac{\lambda_i}{\sum_{j=1}^N \lambda_j} \equiv \text{the probability of capture of the } i^{\text{th}} \text{ animal on any trapping}$$

occasion, $i = 1, 2, \dots, N$.

The set $S_x = \{s_k : k = 1, 2, \dots, x\}$, where $s_k \in \{1, 2, 3, \dots, N\}$ for all k , is used to denote the set of the indices of the x distinct animals seen during the sampling period.

Nayak(1991) showed, in his proposition 2.2, that $(x ; X_1, X_2, \dots, X_x)$ or equivalently $(m; f_1, f_2, \dots, f_m)$ are complete and sufficient for the parameters, namely $\lambda_1, \lambda_2, \dots, \lambda_N$ and N .

§ 4.3 : A New Coverage Adjusted Estimator for the Model M_h

Under this sampling procedure, no usable information is gained from observing the value of z alone, i.e. even if one were able to know the value of $\tau \sum_{i=1}^N \lambda_i$ exactly this would not aid estimation of population size N . (This is also true when looking at the model M_0 : under the model M_0 one obtains the same likelihood function for N from both the probability distribution of x given z and from the joint probability distribution of x and z , see chapter 2). For this reason the approach taken here is to derive an estimate of population size from the conditional distribution of the frequencies given Z . This is essentially equivalent to treating the continuous time data set as if it were obtained from a discrete time sampling experiment with z sampling occasions, where z is viewed as a known constant.

Consider the estimator

$$\hat{N} = \sum_{i \in S_x} \frac{1}{1 - (1 - p_i)^z} \tag{4.1}$$

Under the assumptions described above, it is straightforward to show that \hat{N} would be an unbiased estimator of population size N if the capture probabilities of the animals seen during the experiment were known exactly. However since these capture probabilities are clearly not known exactly the approach taken here is to estimate the p_i

and by doing so obtain an estimator of N by substituting these estimates of capture probability into equation (4.1).

It is now required to estimate the capture probabilities of the animals which were seen during the experiment :

One could use the same approach that Overton(1969) did in the discrete time version of the problem, and use the fact that, under the model, $X_i|Z \sim \text{Bin}(z, p_i)$. Based on this distribution the maximum likelihood estimate of the capture probability of animal i is given by $\hat{p}_i^{(1)} = \frac{X_i}{z}$. On substituting the estimates $\hat{p}_i^{(1)}$ into equation (4.1) one may obtain the estimator \hat{N}_0 , defined by

$$\begin{aligned}\hat{N}_0 &= \sum_{i \in S_x} \frac{1}{1 - [1 - \hat{p}_i^{(1)}]^z} \\ &= \sum_{i \in S_x} \frac{1}{1 - \left(1 - \frac{X_i}{z}\right)^z} \\ &= \sum_{i=1}^z \frac{f_i}{1 - \left(1 - \frac{i}{z}\right)^z}.\end{aligned}$$

The $\hat{p}_i^{(1)}$ are intuitively reasonable estimates of capture probability. This method of estimating capture probability, however, as was the case in chapter 3, does not make full use of all of the available information. The main problem with using the $\hat{p}_i^{(1)}$ is that the sum of these estimates of capture probability is always equal to 1. Whereas they should sum to 1 only if the entire population was seen during the experiment, that is if $x = N$. If $x < N$ then the sum of the capture probabilities of the animals seen during the sampling period must clearly be less than 1. Hence the estimates $\hat{p}_i^{(1)}$ always tend to overestimate capture probability. In order to obtain better estimates of capture probability one may proceed as follows :

The above discussion implies that, firstly, one can argue on intuitive grounds that the estimates of capture probability should be proportional to $\frac{X_i}{z}$, but that the sum of these estimates of capture probability should be strictly less than 1, unless $x = N$. The sum of the capture probabilities of the x animals seen during the experiment is in fact the quantity which has previously been referred to in the literature as 'sample

coverage' : defined by $C = \frac{\sum_{i=1}^N \lambda_i I(X_i > 0)}{\sum_{i=1}^N \lambda_i}$. This quantity is well known in the capture-

recapture literature, see Seber(1982). The fact that the sum of the capture probabilities of the animals seen during the experiment is given by C , sample coverage, may be seen directly .

A good estimator of sample coverage was presented in a software reliability context by Chao, Ma and Jeng(1993) : a derivation of the estimator is outlined in appendix 5. Explicitly a point estimate of sample coverage is given by $\hat{C} = 1 - \frac{f_1}{z}$.

Now returning to the question of how to estimate the p_i for $i \in S_x$, one may proceed as follows :

It is required that $\hat{p}_i \propto \frac{x_i}{z}$

$$\Rightarrow \hat{p}_i = k \frac{x_i}{z}, \quad \text{where } k \text{ is a constant.} \quad (4.2)$$

For the reasons stated above we now set

$$\begin{aligned} \sum_{i \in S_x} \hat{p}_i &= \hat{C} = 1 - \frac{f_1}{z} \\ \Rightarrow k \sum_{i \in S_x} \frac{x_i}{z} &= 1 - \frac{f_1}{z} && \text{using (4.2)} \\ \Rightarrow k &= 1 - \frac{f_1}{z} \\ \Rightarrow \text{estimate the } p_i \text{ by } \hat{p}_i^{(2)} &= \left(1 - \frac{f_1}{z}\right) \frac{x_i}{z}, \quad \text{for } i \in S_x. \end{aligned}$$

These estimates of capture probability are now substituted into equation (4.1) to produce the coverage adjusted estimator \hat{N}_{ca} : defined by

$$\begin{aligned} \hat{N}_{ca} &= \sum_{i \in S_x} \frac{1}{1 - [1 - \hat{p}_i^{(2)}]^z} \\ &= \sum_{i \in S_x} \frac{1}{1 - \left[1 - \left(1 - \frac{f_1}{z}\right) \frac{x_i}{z}\right]^z} \\ &= \sum_{i=1}^z \frac{f_i}{1 - \left[1 - \left(1 - \frac{f_1}{z}\right) \frac{i}{z}\right]^z}. \end{aligned}$$

§ 4.4 : Simulation Study

The results presented in this simulation study concentrate on situations in which the capture probabilities of the animals are obtained via a scaled random sample from a Beta distribution. For each value of N, results are also given for the situation in which each animal in the population is equally likely to be caught. Explicitly results are

presented for situations in which the capture probabilities of the animals are obtained as follows :

Case 1 : $\lambda_i = c \quad i = 1, 2, 3, \dots, N.$

Case 2 : $\lambda_i = c \cdot u_i$ where $u_1, u_2, u_3, \dots, u_N$ are a random sample from $B(\alpha, \beta)$: $\alpha > 0, \beta > 0.$

The constant c in each case is a normalising constant used to ensure that $\sum_{i=1}^N \lambda_i = 1.$

So that for case 1 $\sum_{i=1}^N \lambda_i = 1 \Rightarrow Nc = 1 \Rightarrow c = \frac{1}{N} \Rightarrow p_i = \lambda_i = \frac{1}{N}.$

Similarly for case 2 $p_i = \frac{\lambda_i}{\sum_{j=1}^N \lambda_j} = \lambda_i = c \cdot u_i \Rightarrow c = \frac{1}{\sum_{j=1}^N u_j}.$

For each case 1000 simulations were generated where each simulation ended when the fixed predetermined stopping time τ was reached. On each resulting data set the estimates produced by each of the estimators were calculated and at the end of the 1000 simulations the mean and root mean square error of each estimator was determined. For each version of case 2, a different random sample was used to obtain a different set of capture probabilities in each of the 1000 simulations. Results are given for populations of size 100 and 400, for various stopping times. For each value of N four stopping times were considered. For $N=100$ results are given for $\tau = 41, 69, 139$ and 230 : these are the times for which in the homogeneous case, case 1, one would expect to see the proportions 0.33, 0.5, 0.75 and 0.9 of the population. For $N=400$ results are given for $\tau = 89, 143, 277$ and 644 : these are the times for which in the homogeneous case, case 1, one would expect to see the proportions 0.2, 0.3, 0.5 and 0.8 of the population. Results are conditional upon seeing at least one animal more than once during sampling, i.e. results are conditional upon the maximum likelihood estimator \hat{N} , of chapter 2, producing a finite estimate. In each table 'cv' represents the coefficient of variation of the population ; 'average D' gives the average number of distinct individuals seen. The stopping time τ is denoted by $t.$

Results are given for the following estimators :

mle $\equiv \hat{N}$ the maximum likelihood estimator described in chapter 2.

chao $\equiv \hat{N}_1 = \frac{x}{\hat{C}} + \frac{f_1}{\hat{C}} \hat{\gamma}^2$ is the estimator introduced by Chao, Ma and Yang(1993),

where $\hat{\gamma}^2 = \max \left\{ \frac{\hat{N}_D}{z^2} \sum_i i(i-1)f_i - 1, 0 \right\}.$

$\hat{N}_D = \frac{x}{\hat{C}}$ is the estimator proposed by Darroch

and Ratcliff(1980).

$$nj1 \equiv \hat{N}_{j1} = x + \left(\frac{z-1}{z}\right)f_1 \quad \text{a first order jackknife estimator - when } z \text{ is assumed to be a}$$

constant, this estimator is equivalent to the first order jackknife estimator of Burnham and Overton(1978, 1979).

$$bov \equiv \hat{N}_j \quad \text{again under the assumption that } z \text{ is a pre-chosen constant, this estimator is}$$

equivalent to the interpolated jackknife estimator of Burnham and Overton(1979). This interpolated jackknife is obtained as follows. When the selection procedure of Burnham and Overton(1978) chooses the first order jackknife then \hat{N}_j is equal to \hat{N}_{j1} . When the selection procedure chooses the jackknife of order k , for $k = 2, 3$ or 4 , then \hat{N}_j is a weighted average of the jackknives of orders k and $k-1$. When the selection procedure rejects the fourth order jackknife, \hat{N}_j is equal to \hat{N}_{j1} .

$$O \equiv \hat{N}_O = \sum_{i=1}^z \frac{f_i}{1 - \left(1 - \frac{i}{z}\right)^z}$$

$$\text{and } ca \equiv \hat{N}_{ca} = \sum_{i=1}^z \frac{f_i}{1 - \left[1 - \left(1 - \frac{f_i}{z}\right)\frac{i}{z}\right]^z} \quad \text{represents the coverage adjusted estimator.}$$

Discussion

In case 1 the maximum likelihood estimator \hat{N} for homogeneous populations, from chapter 2, is seen to perform best when both mean and mean square error are considered together. However \hat{N} becomes unacceptably negatively biased in most heterogeneous situations.

The estimators $\hat{p}_i^{(1)}$, as mentioned above, tend to produce estimates which overestimate the capture probabilities of the animals seen during the sampling period. For this reason the estimator \hat{N}_O , which directly incorporates the $\hat{p}_i^{(1)}$, has a tendency to underestimate population size. When sampling for a small amount of time the $\hat{p}_i^{(1)}$ are particularly positively biased. The reason for this is that the sum of the $\hat{p}_i^{(1)}$ is always equal to 1, whereas, as proved above, the sum of the capture probabilities of the animals seen during the experiment should sum to the random quantity C , sample coverage, which is equal to 1 only if $x = N$. For small sampling times one would expect the sum of the capture probabilities of the sighted animals, or equivalently the value of C , to be small - 'a lot less than 1'. Conversely for longer sampling times one would expect the value of C to be large - 'a lot closer to 1'. This is why the $\hat{p}_i^{(1)}$ overestimate more for

small sampling times and less for longer sampling times. A direct consequence of the behaviour of the $\hat{p}_i^{(1)}$ is that the estimator \hat{N}_O tends to be extremely negatively biased for small sampling times and less negatively biased for longer sampling times. Since \hat{N}_{ca} incorporates the estimates $\hat{p}_i^{(2)}$, which are much more reasonable estimates of capture probability over all sampling times, this estimator is able to perform well, notably in terms of mean, for sampling times both long and short.

Given the above discussion, it is not surprising to observe from the following tables that the bias of \hat{N}_{ca} is usually much smaller than that of \hat{N}_O : particularly for the smaller sampling times considered. As one might also have expected \hat{N}_O and \hat{N}_{ca} are seen to perform in a very similar way for the longer sampling times. This is explained by the fact that, for each i , the value of $\hat{p}_i^{(2)}$ tends towards that of $\hat{p}_i^{(1)}$ as sampling time is increased \sim since $1 - \frac{f_1}{z} \rightarrow 1$ as $\tau, z \rightarrow \infty$. In terms of root mean square error, due to the fact that \hat{N}_{ca} generally possesses a much more realistic mean, the coverage adjusted estimator \hat{N}_{ca} is seen on the whole to clearly out perform the Overton-type estimator \hat{N}_O .

The comparison between the first order jackknife estimator \hat{N}_{j1} and the interpolated jackknife estimator \hat{N}_j is also quite straightforward. As one would expect \hat{N}_j generally has a better mean than \hat{N}_{j1} whilst possessing a larger variance. This results in \hat{N}_j being a far better alternative to \hat{N}_{j1} for the smaller stopping times - but that for the longer stopping times \hat{N}_{j1} , due to its smaller variance, can occasionally improve upon \hat{N}_j in terms of mean square error. In terms of overall performance though the interpolated jackknife estimator \hat{N}_j is seen to be preferable to the first order jackknife estimator \hat{N}_{j1} .

For the smaller sampling times \hat{N}_{ca} generally possesses a very good mean value whereas a feature of \hat{N}_j is that in this situation it can be extremely negatively biased. For longer sampling times \hat{N}_j and \hat{N}_{ca} , in situations where the level of heterogeneity is not extreme, have similar bias. However in situations where the coefficient of variation is very large, the negative bias of \hat{N}_j tends to be slightly less than that of \hat{N}_{ca} . The comparison between \hat{N}_j and \hat{N}_{ca} in terms of mean square error is confused somewhat by the way in which mean square error works as a loss function. It has previously been discussed, see chapter 3, that since mean square error rewards negative bias to quite a large extent it is not an ideal loss function - and should not be used on its own. That is in deciding between which of the two estimators \hat{N}_j and \hat{N}_{ca} is performing best overall one must consider both mean and mean square error - or equivalently both mean and variance. For this reason the choice between the estimators \hat{N}_j and \hat{N}_{ca} is not straightforward and is particularly difficult due to the fact that these two estimators each

behave in a very different way. For the smaller stopping times considered in the tables it is seen that \hat{N}_{ca} usually has a very good mean but possesses a relatively large variance whereas \hat{N}_j tends to be extremely negatively biased whilst possessing a small variance. For these small stopping times, the estimator \hat{N}_{ca} , in terms of mean square error, is generally seen to be a better alternative to \hat{N}_j . In situations of this type where \hat{N}_{ca} has a larger mean square error than \hat{N}_j it almost always exhibits a much more realistic mean value. For the longer sampling times, when the coefficient of variation is very large, \hat{N}_j generally possesses a slightly better mean than \hat{N}_{ca} and consequently can also have a slightly smaller mean square error.

For long stopping times, when the heterogeneity is mild, an important feature of \hat{N}_1 is that it tends to possess a good mean value. This is in contrast to the behaviour of the other estimators designed for heterogeneous populations. That is the estimators \hat{N}_{j1} , \hat{N}_j and \hat{N}_{ca} have a tendency to overestimate when the sampling time is long and the coefficient of variation is less than about 0.4. However, for reasonably large values of τ , when the coefficient of variation is above about 0.4, the negative bias of \hat{N}_1 tends to be greater than that of \hat{N}_{j1} , \hat{N}_j and \hat{N}_{ca} .

The comparison between the coverage adjusted estimator \hat{N}_{ca} and the estimator of Chao and Yang(1993) is quite straightforward. Only when the coefficient of variation is very small and the sampling time is long does \hat{N}_1 perform better than \hat{N}_{ca} : in this situation \hat{N}_{ca} is positively biased and its variance is larger than that of the almost unbiased \hat{N}_1 . For small to moderate sampling times, the variance of \hat{N}_{ca} is smaller than that of \hat{N}_1 and consequently, even though when the heterogeneity is mild \hat{N}_1 can be less biased, the coverage adjusted \hat{N}_{ca} tends to possess a mean square error smaller than that of \hat{N}_1 . Finally, for large values of τ , when the coefficient of variation is above about 0.4, the coverage adjusted estimator \hat{N}_{ca} is generally less biased and also tends to have a mean square error smaller than that of \hat{N}_1 .

Given the above evidence it is clear that the coverage adjusted estimator \hat{N}_{ca} may at least be considered as a viable alternative to the estimator proposed by Chao, Ma and Yang(1993) and to the jackknife estimators. Furthermore it is believed that, particularly due to the performance of the estimators for the smaller stopping times, one should always use the coverage adjusted estimator \hat{N}_{ca} in preference to either the jackknife estimators or the estimator of Chao, Ma and Yang(1993).

Chapter 5 : Conclusions

§ 5.1 : Initial Objectives and Results Obtained

The initial aim of this thesis was to begin a systematic investigation into the method of plant-capture when applied to populations behaving according to the closed capture-recapture models of Otis et al. (1978) and their continuous time analogues.

Chapter 1 considers the addition of plants to target populations which behave according to the most basic Otis et al. (1978) model M_0 . A plant-capture generalisation of the standard maximum likelihood estimator, that was discussed by Otis et al. (1978), is derived. Other new estimators are also introduced. A near-unbiased estimator, described as a conditionally unbiased estimator (CUE), which was originally considered in a rather simpler urn theory context, is shown to be more satisfactory than the maximum likelihood estimator and the Peterson-type estimator. This latter conclusion holds either in the presence or absence of planted individuals, and hence for the standard capture-recapture model it is recommended that the CUE be preferred to the usual maximum likelihood estimator of Otis et al. (1978).

In chapter 2 consideration is given to the use of plants in connection with a continuous time analogue of the Otis et al. (1978) model M_0 . In the absence of plants this model is equivalent to the recapture debugging model of Nayak(1988). The maximum likelihood estimator of Nayak(1988), which was designed for situations in which no plants are used, is generalised and other new estimators are introduced. As well as estimators corresponding to those of chapter 1, harmonic mean estimators are considered. Again, however, in situations where plants are deployed and in those which no plants are used, it is the near-unbiased estimator, described as a conditionally unbiased estimator, which is seen to give the best results.

Difficulties arose when investigating the usefulness of plants when used in connection with populations behaving according to the important heterogeneity model M_h . The main problem was that the most commonly preferred estimators, which for the model M_h have been the jackknife estimators of Burnham and Overton(1978, 1979) and the estimators of Chao, Lee and Jeng(1992), cannot easily be generalised in a way which allows them to use the extra information gained from the plants. The approach taken to overcome this problem was to seek new estimators for the standard capture-recapture problem with a view to finding an estimator that could be generalised for use in a plant-capture scenario. This approach proved fruitful and led to a number of new estimators and estimation procedures for the standard capture-recapture model M_h . Of the new estimators that were obtained, the most satisfactory were found to be the

coverage adjusted estimators, which are presented in chapter 3. Also within chapter 3 it is shown how these coverage adjusted estimators can be modified in a way which allows them to utilize the information gained from the planted individuals. In the absence of plants, the coverage adjusted estimators are shown, through simulation and real data, to compare favourably with other estimators.

It was recognised that the approach taken in chapter 3 could be modified to produce a new estimator for the continuous time analogue of the Otis et al. (1978) model M_h . Within chapter 4, this new coverage adjusted estimator was shown using a simulation study to perform more satisfactorily than other estimators that have been proposed for this model.

The emphasis throughout this thesis has been on point estimation. Most practitioners, however, would also require reasonable variance estimates. An appropriate extension of the work contained in this thesis would, therefore, be to develop estimators of variance. This could be done analytically or using computer-intensive methods.

Throughout this thesis it has been concluded that the use of plants is beneficial. It is important to keep in mind that this conclusion has been reached under the assumption that the planted individuals behave in an identical manner to those of the target population. Further work should, therefore, investigate the robustness of the conclusions reached in the case where the behaviour of the planted individuals is different to those of the target population. The development of procedures to test whether this central assumption does indeed hold would also be appropriate. Some work has already been done in this area : in a personal communication K. Pollock showed that in the component P_4 of the probability function 3.11 of section 3.7.3 chapter 3, we have a multiple hypergeometric distribution which gives rise to a contingency table test of the assumption that the planted individuals behave in an identical manner to those of the target population. This test could be used in connection with target populations behaving according to the Otis et al. (1978) models M_0 and M_h . Similar contingency table tests would be appropriate for the models M_b and M_{bh} , K. Pollock(pers. com.).

Within each of chapters 1, 2 and 3 a Peterson-type estimator was derived. In each chapter it has exactly the same functional form, and is based only on the number of distinct individuals seen from the planted and target populations. An important feature of this estimator is that, when the plants do behave as members of the target population, it is on the whole unbiased. Within chapters 1 and 2, however, the Peterson-type estimator did not perform well in comparison to the other estimators : although its bias tended to be low, its variance, particularly when the number of plants used was small relative to the size of the target population, was relatively large. In chapter 3 the performance of the Peterson-type estimator improved in relation to that of the other

estimators; the consistent near unbiasedness of the Peterson-type estimator becoming more useful in the presence of heterogeneity. Within chapters 1 and 2 each of the models considered had considerable structure which benefited the other estimators and enabled them to perform better than the Peterson-type estimator. This Peterson-type estimator is in fact suitable for use in plant-capture scenarios when the target population behaves according to any of the eight closed capture-recapture models of Otis et al. (1978) or their continuous time analogues. In each of these situations, so long as the planted individuals do behave in an identical manner to those of the target population, the Peterson-type estimator would remain nearly unbiased. Hence for the more complicated, less structured, models the Peterson-type estimator should prove to be a much stronger candidate.

Appendix 1 : The Classical Occupancy Distribution

Suppose Z balls are dropped independently into N urns in such a way that the probability of any one ball being allocated to any one urn is $1/N$. Let X denote the number of occupied urns after the Z balls have been dropped, i.e. X denotes the number of urns containing at least one ball after the Z balls have been dropped. The distribution of X given the value of Z is known as the Classical Occupancy distribution, see Johnson and Kotz(1977) p.110. The conditional probability function of X given Z can be written as

$$P(X = x|Z = z) = N^{-z} \binom{N}{x} x! S(x, z), \quad x=1,2,3,\dots,\min(N,z),$$

where $S(x,z)$ is a Stirling number of the second kind defined by

$$S(x,z) = \frac{1}{x!} \sum_{k=0}^x \binom{x}{k} (-1)^k (x-k)^z.$$

Proof :

As a first step in this proof consider the following formula :

Boole's Formula : The Inclusion Exclusion Principle

$$P\left[\bigcup_{i=1}^x A_i\right] = \sum_{k=1}^x (-1)^{k-1} S_k,$$

$$\text{where } S_k = \sum_{\substack{\text{all subsets} \\ \text{of size } k}} P(A_{i_1}, A_{i_2}, A_{i_3}, \dots, A_{i_k}).$$

Now define the events B and A_i in the following way

$B \equiv$ a particular $N-x$ urns remain empty

(i.e. z balls are restricted to the other x urns)

$A_i \equiv$ a particular $N-x$ urns remain empty but urn i of the remaining x urns also remains empty.

$$P[B, \text{ all } x \text{ urns are occupied }] = P[B] - P[B, \text{ at least one of the } x \text{ urns is empty }]$$

$$\begin{aligned} &= P[B] - P\left[\bigcup_{i=1}^x A_i\right] \\ &= P[B] - \sum_{k=1}^x (-1)^{k-1} S_k, \quad \text{using Boole's formula.} \\ &= \left(\frac{x}{N}\right)^z - \sum_{k=1}^x (-1)^{k-1} \left(\frac{x-k}{N}\right)^z \binom{x}{k}, \end{aligned} \quad (\text{A1.1})$$

$$\text{since } S_k = \sum_{\substack{\text{all subsets} \\ \text{of size } k}} P(A_{i_1}, A_{i_2}, A_{i_3}, \dots, A_{i_k}) = \left(\frac{x-k}{N}\right)^z \binom{x}{k}.$$

From (A1.1) it follows that

$$\begin{aligned} P[\text{B, all } x \text{ urns are occupied}] &= \left(\frac{1}{N}\right)^z \left[x^z + \sum_{k=1}^x \binom{x}{k} (-1)^k (x-k)^z \right] \\ &= N^{-z} \sum_{k=0}^x \binom{x}{k} (-1)^k (x-k)^z \\ &= N^{-z} x! S(x, z). \end{aligned}$$

But a particular subset of size x can be chosen in $\binom{N}{x}$ ways. Hence

$$P(X = x | Z = z) = N^{-z} \binom{N}{x} x! S(x, z), \quad x = 1, 2, \dots, \min(N, z).$$

Appendix 2 : The Stirling distribution of the second kind

Suppose that X_i has a Poisson distribution with mean λ . Then the random variable $Y = \sum_{i=1}^n X_i | X_i > 0$ is said to have the Stirling distribution of the second kind, see Patil et al(1984). That is a random variable is said to have the Stirling distribution of the second kind if it can be represented as the sum of a number of zero truncated Poisson random variables.

The probability function of X_i is given by $P(X_i = x_i) = \frac{\lambda^{x_i} \exp(-\lambda)}{x_i!}$, $x_i = 0, 1, 2, \dots$

It follows that $P(X_i = x_i | x_i > 0) = \frac{\lambda^{x_i} \exp(-\lambda)}{x_i!(1 - \exp(-\lambda))}$, $x_i = 1, 2, 3, \dots$

The probability generating function of $X_i | X_i > 0$ is given by

$$\begin{aligned} G_{X_i | X_i > 0}(t) &= E(t^{X_i | X_i > 0}) \\ &= \sum_{x_i=1}^{\infty} t^{x_i} \cdot \frac{\lambda^{x_i} \exp(-\lambda)}{x_i!(1 - \exp(-\lambda))} \\ &= \frac{\exp(-\lambda)}{1 - \exp(-\lambda)} \cdot \left[\sum_{x_i=0}^{\infty} \frac{(t\lambda)^{x_i}}{x_i!} - 1 \right] \\ &= \frac{\exp(-\lambda)}{1 - \exp(-\lambda)} \cdot [\exp(\lambda t) - 1] \\ &= \frac{\exp(\lambda t) - 1}{\exp(\lambda) - 1}. \end{aligned}$$

The probability generating function of Y is then

$$\begin{aligned} G_Y(t) &= G_{\sum_{i=1}^n X_i | X_i > 0}(t) = E\left(t^{\sum_{i=1}^n X_i | X_i > 0}\right) = \prod_{i=1}^n G_{X_i | X_i > 0}(t) \\ &= \left[\frac{\exp(\lambda t) - 1}{\exp(\lambda) - 1} \right]^n \\ &= \frac{1}{(\exp(\lambda) - 1)^n} \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} \exp(k\lambda t) \\ &= \frac{1}{(\exp(\lambda) - 1)^n} \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} \sum_{r=0}^{\infty} \frac{(k\lambda)^r t^r}{r!}. \end{aligned}$$

The probability of Y taking the value r is equal to the coefficient of t^r in the expansion of $G_Y(t)$. Hence

$$\begin{aligned} P(Y = r) &= \frac{\lambda^r}{r!(\exp(\lambda) - 1)^n} \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} k^r \\ &= \frac{n!}{r!} \frac{\lambda^r S(n, r)}{(\exp(\lambda) - 1)^n}, \quad r = n, n+1, n+2, \dots \end{aligned}$$

where $S(n, r)$ is a Stirling number of the second kind.

Appendix 3 : The Distribution Function of A Sum of Zero Truncated Binomial
Random Variables

Suppose that Y_j , for $j = 1, 2, \dots, x$, are independently and identically distributed Binomial random variables, each having parameters t and p . Here consideration is given to the distribution of $Q = \sum_{j=1}^x Y_j | Y_j > 0$. This distribution was first derived by Ahuja(1970) and, more recently, was inspected via a power series approach by Charalambides and Singh(1988). The approach taken here is to obtain the distribution of Q using probability generating functions :

Firstly, given that $Y_j \sim \text{Bin}(t, p)$,

$$\text{Prob}(Y_j = y_j) = \binom{t}{y_j} p^{y_j} (1-p)^{t-y_j}, \quad y_j = 0, 1, 2, \dots, t.$$

$$\text{Prob}(Y_j = y_j | Y_j > 0) = \frac{\binom{t}{y_j} p^{y_j} (1-p)^{t-y_j}}{1 - (1-p)^t}, \quad y_j = 1, 2, \dots, t.$$

The probability generating function of $Y_j | Y_j > 0$ is then given by

$$\begin{aligned} G_{Y_j | Y_j > 0}(m) &= E(m^{Y_j | Y_j > 0}) \\ &= \sum_{y_j=1}^t m^{y_j} \cdot \frac{\binom{t}{y_j} p^{y_j} (1-p)^{t-y_j}}{1 - (1-p)^t} \\ &= \frac{1}{1 - (1-p)^t} \sum_{y_j=1}^t \binom{t}{y_j} (mp)^{y_j} (1-p)^{t-y_j} \\ &= \frac{1}{1 - (1-p)^t} \left[\sum_{y_j=0}^t \binom{t}{y_j} (mp)^{y_j} (1-p)^{t-y_j} - (1-p)^t \right] \\ &= \frac{1}{1 - (1-p)^t} \left[(1-p + mp)^t - (1-p)^t \right]. \end{aligned} \tag{A3.1}$$

The probability generating function of $Q = \sum_{j=1}^x Y_j | Y_j > 0$ may now be obtained as follows :

$$\begin{aligned} G_Q(m) &= E(m^Q) = E\left(m^{\sum_{j=1}^x Y_j | Y_j > 0}\right) = E\left(\prod_{j=1}^x m^{Y_j | Y_j > 0}\right) \\ &= \prod_{j=1}^x E\left(m^{Y_j | Y_j > 0}\right) \quad \text{by independence} \\ &= \prod_{j=1}^x G_{Y_j | Y_j > 0}(m) \end{aligned}$$

$$\begin{aligned}
&= \prod_{j=1}^x \frac{1}{1-(1-p)^t} \left[(1-p+mp)^t - (1-p)^t \right], && \text{from (A3.1)} \\
&= \frac{1}{[1-(1-p)^t]^x} \left[(1-p+mp)^t - (1-p)^t \right]^x \\
&= \frac{1}{[1-(1-p)^t]^x} \sum_{r=0}^x \binom{x}{r} (1-p+mp)^{tr} (-1)^{x-r} (1-p)^{tx-tr} \\
&= \frac{1}{[1-(1-p)^t]^x} \sum_{r=0}^x \binom{x}{r} (-1)^{x-r} (1-p)^{tx-tr} \sum_{q=0}^{tr} \binom{tr}{q} (1-p)^{tr-q} p^q m^q.
\end{aligned}$$

It then follows that

$$\begin{aligned}
\text{Prob}(Q = q) &= \frac{1}{[1-(1-p)^t]^x} \sum_{r=0}^x \binom{x}{r} (-1)^{x-r} (1-p)^{tx-tr} \binom{tr}{q} (1-p)^{tr-q} p^q \\
&= \frac{p^q (1-p)^{tx-q}}{[1-(1-p)^t]^x} \sum_{r=0}^x \binom{x}{r} \binom{tr}{q} (-1)^{x-r}, \quad q = x, x+1, x+2, \dots, tx.
\end{aligned}$$

Appendix 4 : Estimation of Sample Coverage : Model M_h : Discrete Time Sampling

Procedure

This appendix describes how one may estimate the quantity referred to as 'sample coverage' within chapter 3. For notation and relevant background please refer to chapter 3. The three estimators of sample coverage that are described below were considered by Chao, Lee and Jeng(1992).

Sample Coverage (C), is defined as follows :

$$C = \frac{\sum_{i=1}^N p_i I(X_i > 0)}{\sum_{i=1}^N p_i}, \quad \text{where} \quad I(X_i > 0) = \begin{cases} 1 & \text{w.p. } 1 - (1 - p_i)^t \\ 0 & \text{w.p. } (1 - p_i)^t \end{cases}$$

As a first step to estimating this quantity, consider its expectation :

$$\begin{aligned} E(C) &= E \left[\frac{\sum_{i=1}^N p_i I(X_i > 0)}{\sum_{i=1}^N p_i} \right] \\ &= 1 - \frac{\sum_{i=1}^N p_i (1 - p_i)^t}{\sum_{i=1}^N p_i}. \end{aligned} \tag{A4.1}$$

An estimate of C may be obtained from equation (A4.1) by substituting in estimates of $\sum_{i=1}^N p_i (1 - p_i)^t$ and $\sum_{i=1}^N p_i$.

As in section 3.3 , using equation (3.8), an estimate of $\sum_{i=1}^N p_i$ is given by $\frac{Z}{t}$.

To obtain an estimate of $\sum_{i=1}^N p_i (1 - p_i)^t$ consider the expansion :

$$\begin{aligned} \sum_{i=1}^N p_i (1 - p_i)^t &= \sum_{i=1}^N p_i (1 - p_i)^{t-1} - \sum_{i=1}^N p_i^2 (1 - p_i)^{t-1} \\ &= \frac{1}{\binom{t}{1}} E(f_1) - \sum_{i=1}^N p_i^2 (1 - p_i)^{t-2} + \sum_{i=1}^N p_i^3 (1 - p_i)^{t-2} \\ &= \frac{1}{\binom{t}{1}} E(f_1) - \frac{1}{\binom{t}{2}} E(f_2) + \sum_{i=1}^N p_i^3 (1 - p_i)^{t-3} - \sum_{i=1}^N p_i^4 (1 - p_i)^{t-3} \\ &= \frac{1}{\binom{t}{1}} E(f_1) - \frac{1}{\binom{t}{2}} E(f_2) + \frac{1}{\binom{t}{3}} E(f_3) + \dots + \frac{(-1)^{t+1}}{\binom{t}{t}} E(f_t) + (-1)^{t+2} \sum_{i=1}^N p_i^{t+1}. \end{aligned}$$

That is

$$\begin{aligned} \sum_{i=1}^N p_i (1-p_i)^t &= \sum_{j=1}^t \frac{(-1)^{j+1}}{\binom{t}{j}} E(f_j) + (-1)^{t+2} \sum_{i=1}^N p_i^{t+1} \\ \Rightarrow \sum_{i=1}^N p_i (1-p_i)^t &\approx \sum_{j=1}^t \frac{(-1)^{j+1}}{\binom{t}{j}} E(f_j) \\ \Rightarrow \text{an estimate of } \sum_{i=1}^N p_i (1-p_i)^t &\text{ is given by } \sum_{j=1}^t \frac{(-1)^{j+1}}{\binom{t}{j}} f_j. \end{aligned}$$

Now returning to equation (A4.1) it is seen that an estimate of sample coverage C is given by

$$\begin{aligned} \hat{C} &= 1 - \frac{\sum_{j=1}^t \frac{(-1)^{j+1}}{\binom{t}{j}} f_j}{\frac{z}{t}} \\ &= 1 - \frac{t}{z} \sum_{j=1}^t \frac{(-1)^{j+1}}{\binom{t}{j}} f_j. \end{aligned} \tag{A4.2}$$

Computation has shown that, in almost all situations, only the first few terms in the above summation are significant. For this reason we take the approach of Chao, Lee and Jeng(1992) and truncate the summation in (A4.2) in order to obtain the following three estimators of sample coverage :

$$\begin{aligned} \hat{C}_1 &= 1 - \frac{f_1}{z}, \\ \hat{C}_2 &= 1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z} \\ \text{and } \hat{C}_3 &= 1 - \frac{f_1}{z} + \frac{2}{(t-1)} \frac{f_2}{z} - \frac{6}{(t-1)(t-2)} \frac{f_3}{z}. \end{aligned}$$

Using the whole summation of equation (A4.2) is not recommended. This is mainly due to the fact that the higher frequencies tend to possess large variability : elements in the summation which incorporate these quantities can occasionally distort the resulting estimate of sample coverage.

Appendix 5 : Estimation of Sample Coverage : Model M_n : Continuous Time Sampling Procedure

This appendix describes how one may estimate the quantity referred to as 'sample coverage' within chapter 4. For notation and relevant background please refer to chapter 4. The following derivation appeared in Chao, Ma and Yang(1993).

Firstly note that
$$I(X_i > 0) = \begin{cases} 1 & \text{w.p. } 1 - \exp(-\lambda_i t) \\ 0 & \text{w.p. } \exp(-\lambda_i t) \end{cases}$$

$$\Rightarrow E[I(X_i > 0)] = 1 - \exp(-\lambda_i t). \quad (\text{A5.1})$$

The expectation of sample coverage C may now be written as

$$\begin{aligned} E(C) &= E \left[\frac{\sum_{i=1}^N \lambda_i I(X_i > 0)}{\sum_{j=1}^N \lambda_j} \right] \\ &= \frac{\sum_{i=1}^N \lambda_i E[I(X_i > 0)]}{\sum_{j=1}^N \lambda_j} \\ &= \frac{\sum_{i=1}^N \lambda_i [1 - \exp(-\lambda_i t)]}{\sum_{j=1}^N \lambda_j}, && \text{from A5.1.} \\ &= 1 - \frac{\sum_{i=1}^N \lambda_i \exp(-\lambda_i t)}{\sum_{j=1}^N \lambda_j} \\ &= 1 - \frac{\sum_{i=1}^N \lambda_i t \cdot \exp(-\lambda_i t)}{t \sum_{j=1}^N \lambda_j}. && (\text{A5.2}) \end{aligned}$$

Now $X_i \sim P(\lambda_i t)$

$$\Rightarrow z = \sum_{i=1}^N X_i \sim P\left(t \sum_{i=1}^N \lambda_i\right) \Rightarrow z \text{ is a good estimate of } t \sum_{i=1}^N \lambda_i.$$

Consider the quantity $f_1 = \sum_{i=1}^N I(X_i = 1)$,

$$\text{where } I(X_i = 1) = \begin{cases} 1 & \text{w.p. } \lambda_i t \cdot \exp(-\lambda_i t) \\ 0 & \text{w.p. } 1 - \lambda_i t \cdot \exp(-\lambda_i t) \end{cases}$$

It follows that

$$\begin{aligned} E[f_1] &= \sum_{i=1}^N E[I(X_i = 1)] \\ &= \sum_{i=1}^N \lambda_i t \cdot \exp(-\lambda_i t). \end{aligned}$$

$\Rightarrow f_1$ may be used as an estimator of $\sum_{i=1}^N \lambda_i t \cdot \exp(-\lambda_i t)$.

Now using equation A5.2 it follows that the value of C may be estimated by

$$\hat{C} = 1 - \frac{f_1}{z}.$$

References

- Ahuja, J. C. (1970). On the distribution of the sum of independent positive binomial variables. *Canadian Math. Bull.*, 13, 151-153.
- Bai, Z. & Chow, M. (1991). Inadmissibility of the maximum likelihood estimator in the sequential estimation of the size of a population, *Biometrika*, 78, 817-823.
- Berg, S. (1974). Factorial series distributions with applications to capture-recapture problems. *Scand. J. Statist* 1, 145-52.
- Berg, S. (1975). Some properties and applications of a ratio of Stirling numbers of the second kind. *Scand. J. Statist.* 2, 91-94.
- Berg, S. (1976). A note on the UMVU estimate in a multiple-recapture census. *Scand. J. Statist.* 3, 86-88.
- Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, 65, 625-633.
- Burnham, K. P. and Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60, 927-936.
- Carothers, A. D. (1973). Capture-recapture methods applied to a population with known parameters. *Journal of Animal Ecology*, 42, 125-146.
- Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics* 45, 427-438.
- Chao, A., Lee, S-M. and Jeng, S-L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* 48, 201-216.
- Chao, A., Ma M-C. and Yang. M. C. K. (1993). Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika*, 80, 193-201.

- Charalambides, Ch. A. (1981). On a restricted occupancy model and its applications. *Biom. J.* , 23, 601-610.
- Charalambides, Ch.A. and Singh, J.(1988). A review of the Stirling Numbers, their generalizations and statistical applications. *Commun. Statist.-Theory Meth.*, 17(8), 2533-2595.
- Cormack, R. M. (1966). A test for equal catchability. *Biometrics* 22, 330-342.
- Craig, C. C. (1953). Use of marked specimens in estimating populations. *Biometrika* 40, 170-176.
- Darroch, J. N. (1958). The multiple-recapture census. 1. Estimation of a closed population. *Biometrika*, 45, 343-359.
- Darroch, J. N. and Ratcliff, D. (1980). A note on capture-recapture estimation. *Biometrics* 36, 149-153.
- Duran, J. W. and Wiorowski, J. J. (1981). Capture-recapture sampling for estimating software error content. *IEEE Trans. Software Engineering*, SE-7, 147-148.
- Edwards, W. R. and Eberhardt, L. L. (1967). Estimating cottontail abundance from livetrapping data. *Journal of Wildlife Management*, 31, 87-96.
- Goodman, L. A. (1953). Sequential sample tagging for population size problems, *Annals of Mathematical Statistics*, 24, 56-69.
- Goudie, I. B. J. (1995). A plant-capture approach for achieving complete coverage of a population. *Comm. Statist. Theory Methods* 24, 1293-1305.
- Gould. H. W. and Hopper. A. T. (1962). Operational formulas connected with two generalizations of Hermite polynomials. *Duke Math. J.*, 29, 51-63.
- Harris, B. (1968). Statistical inference in the classical occupancy problem : Unbiased estimation of the number of classes. *J. Amer. Statist. Assoc.*, 63, 837-847.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

- Jelinski, Z. and Moranda, P. B. (1972). "Software reliability research" in statistical computer performance evaluation, ed. W.Freiberger, London : Academic Press, 465-484.
- Joe, H. and Reid, N. (1985). Estimating the number of faults in a system, *Journal of the American Statistical Association*, 80, 222-226.
- Johnson, N. L. and Kotz, S. (1977). *Urn models and their application*. Wiley. New York.
- Kelker, G. H. (1940). Estimating deer populations by a differential hunting loss in the sexes. *Proc, Utah Acad. Sci., Arts and Letters*. 17, 6-69.
- Langberg, N. and Singpurwalla, N. D. (1985). Unification of some software reliability models via the Bayesian approach. *SIAM Journal of scientific and statistical computing*, 6, 781-790.
- Laska, E. M. and Meisner, M. (1993). A plant-capture method for estimating the size of a population from a single sample. *Biometrics*, 49, 209-220.
- Lee, S-M. and Chao, A. (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, 50, 88-97.
- Lehmann, E. L. and Scheffe, H. (1950). Completeness, similar regions and unbiased estimation. *Sankhya* 10, 305-329.
- Mares, M. A., Streilein, K. E. and Willig, M. R. (1981). Experimental assessment of several population techniques on an introduced population of eastern chipmunks. *Journal of Mammalogy* 62, 315-328.
- Martin, E., Laska, E. M., Hopper, K., Meisner, M. and Wanderling, J. (1995). Issues in the use of a plant-capture method for estimating the size of the street dwelling population. (Unpublished manuscript - research conducted by staff of the U.S. Census Bureau and the Nathan S. Kline Institute.)
- Mills, H. D. (1972). On the statistical validation of computer programs. IBM Federal Syst. Div., Gothenburg, MD. Rep. 72-6015.

- Nayak, T. K. (1988). Estimating population size by recapture sampling, *Biometrika*, 75, 113-120.
- Nayak, T. K. (1991). Estimating the number of component processes of a superimposed process. *Biometrika* 78, 75-81.
- Norris, J. L. and Pollock, K. H. (1996a). Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* 52, 639-649.
- Norris, J. L. and Pollock, K. H. (1996b). Nonparametric MLE incorporation of heterogeneity and model testing into premarked cohort studies. (submitted for publication).
- Otis, D. L., Burnham, K. P., White G. C. and Anderson D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62. The Wildlife Society, Washington D.C., USA.
- Overton, W. S. (1969). Estimating the numbers of animals in wildlife populations. *Wildlife management techniques*, R.H. Giles (ed.), 403-456. Washington D.C. : The Wildlife Society.
- Pathak, P. K. (1964). On estimating the size of a population and its inverse by capture mark method. *Sankhya A* 26, 75-80.
- Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size biased sampling with applications to wildlife populations and human families. *Biometrics*, 34, 179-189.
- Patil, G. P., Boswell, M. T., Joshi, S. W. and Ratnaparkhi, M. V.(1984). Dictionary and classical bibliography of statistical distributions in scientific work. Vol. 1 Discrete Models, International Co-operative Publishing House, Fairland, Maryland, USA.
- Pollock, K. H. (1974). The assumption of equal catchability of animals in tag-recapture experiments. Ph.D. Thesis, Cornell Univ., Ithaca, N.Y.
- Pollock, K. H. (1976). Building models of capture-recapture experiments. *The Statistician* 25, 253-260.

- Pollock, K. H., Nichols, J. D., Brownie, C. and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. Wildlife Monographs 107. The Wildlife Society, Washington D.C., USA.
- 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.
- Romanovsky, V. (1934). Su due problemi di distribuzione casuale. *Giorn. Ist. Ital. Attuari*, 5, 196-218.
- Rosenberg, D. K., Overton, W. S. and Anthony, R. G. (1995). Estimation of animal abundance when capture probabilities are low and heterogeneous. *Journal of Wildlife Management* 59, 252-261.
- Rupp, R. S. (1966). Generalised equation for the ratio method of estimating population abundance. *J. Wildl. Manag.* 30, 523-526.
- Samuel-Cahn, E. (1974). Asymptotic distributions for occupancy and waiting time problems with positive probability of falling through the cells. *Annals of Probability*, 2, 515-521.
- Schnabel, Z. E. (1938). The estimation of the total fish population of a lake. *Am. Math. Mon.* 45, 348-352.
- Seber, G. A. F. (1982). The estimation of animal abundance and related parameters, 2nd edition. London : Griffin.
- Smith, E. P. and van Belle, G. (1984). Nonparametric estimation of species richness. *Biometrics* 40, 119-129.
- Skalski, J. R. and Robson, D. S. (1982). A mark and removal procedure for estimating animal abundance. *J. Wildl. Manag.* 46, 741-751.
- White, G. C., Anderson, D. R., Burnham, K. P. and Otis, D. L. (1982). Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, LA 8787-NERP, Los Alamos, N.M.

- Yip, P. (1995). Estimating the number of errors in a system using a martingale approach. *IEEE Trans. Reliability R-44*, 322-326.
- Yip, P. (1996). Effect of plant-capture in a capture-recapture experiment. *Comm. Statist. Theory Methods*, 25, 2025-2038.
- Yip, P. and Chao, A. (1996). Estimating population size from capture-recapture studies via sample coverage and estimating functions. *Comm. Statist. Stoch. Models*, 12(1), 17-35.