



Statistical Inference from Capture Data on Closed Animal Populations

Author(s): David L. Otis, Kenneth P. Burnham, Gary C. White, David R. Anderson

Source: *Wildlife Monographs*, No. 62, Statistical Inference from Capture Data on Closed Animal Populations (Oct., 1978), pp. 3-135

Published by: [Allen Press](#)

Stable URL: <http://www.jstor.org/stable/3830650>

Accessed: 24/10/2011 16:18

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

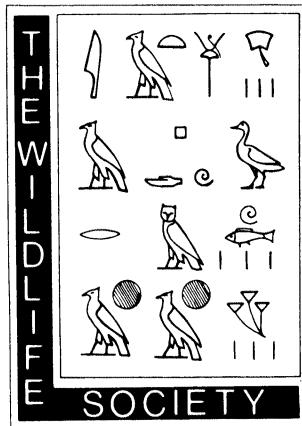


Allen Press is collaborating with JSTOR to digitize, preserve and extend access to *Wildlife Monographs*.

<http://www.jstor.org>

WILDLIFE MONOGRAPHS

A Publication of The Wildlife Society



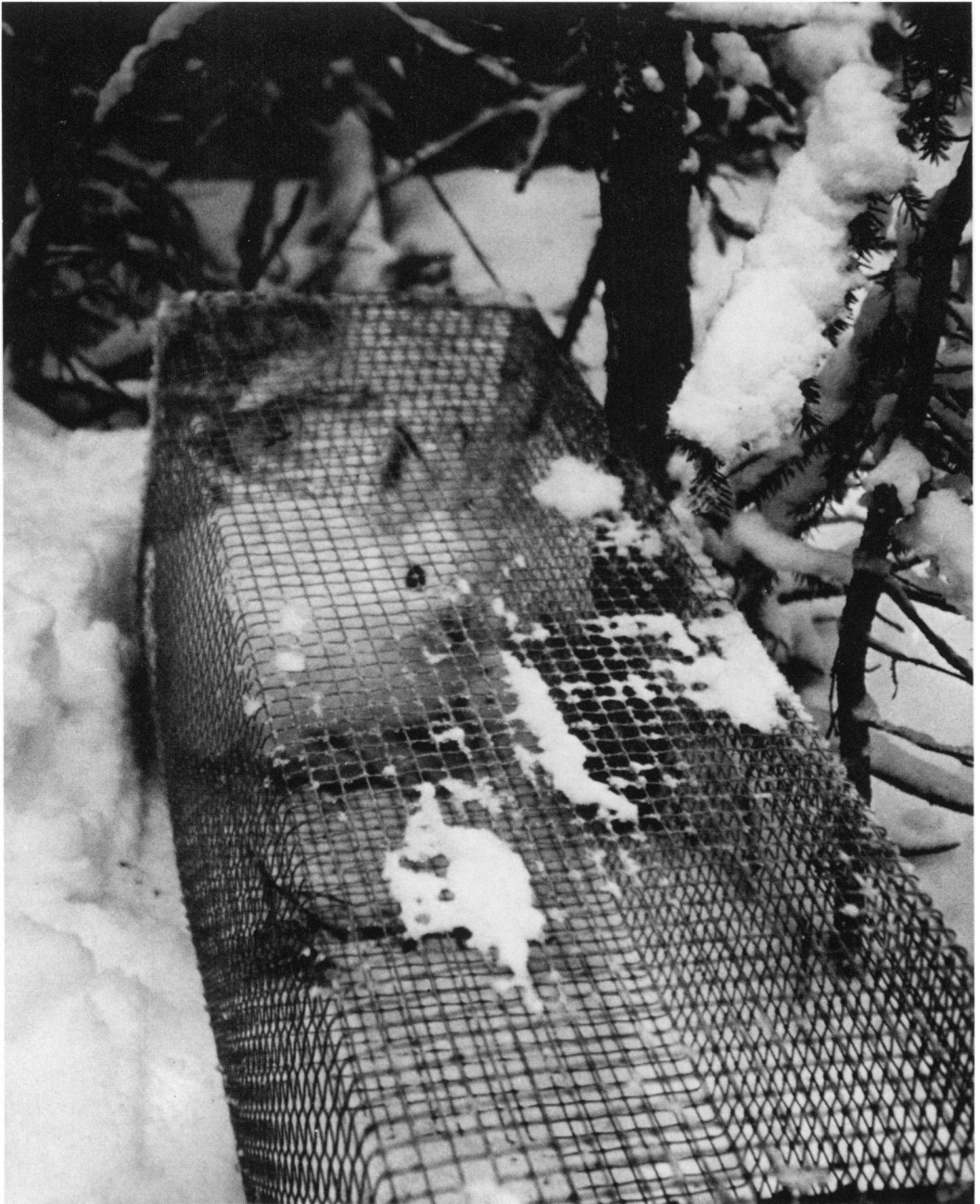
STATISTICAL INFERENCE FROM CAPTURE DATA ON CLOSED ANIMAL POPULATIONS

by

DAVID L. OTIS, KENNETH P. BURNHAM,
GARY C. WHITE, AND DAVID R. ANDERSON

OCTOBER 1978

No. 62



FRONTISPIECE. Capture-recapture studies are frequently conducted on small mammal populations such as snowshoe hares *Lepus americanus*. (Photograph by Leta Burnham.)

STATISTICAL INFERENCE FROM CAPTURE DATA ON CLOSED ANIMAL POPULATIONS

David L. Otis,¹ Kenneth P. Burnham,² Gary C. White,³
and David R. Anderson

Utah Cooperative Wildlife Research Unit, Utah State University, Logan, Utah 84322

CONTENTS

INTRODUCTION	7	<i>Example</i>	35
<i>Objectives</i>	8	<i>Example</i>	36
<i>Assumptions</i>	9	<i>Discussion</i>	37
<i>Unequal Capture Probabilities</i>	11	MODEL M _{1b} : CAPTURE PROBABILITIES VARY	
<i>Perspectives</i>	12	BY TIME AND BEHAVIORAL RESPONSE	
<i>Comments on the Use of this Monograph</i>	12	TO CAPTURE	37
ACKNOWLEDGMENTS	13	<i>Structure and Use of the Model</i>	37
FUNDAMENTAL CONCEPTS	14	<i>Discussion</i>	38
<i>Data and Parameters</i>	14	MODEL M _{1h} : CAPTURE PROBABILITIES VARY	
<i>Statistics and Notation</i>	15	BY TIME AND INDIVIDUAL ANIMAL	38
<i>Parameter Estimation</i>	16	<i>Structure and Use of the Model</i>	38
<i>Interval Estimation</i>	17	<i>Simulation Results</i>	39
HISTORICAL OVERVIEW	17	<i>Discussion</i>	39
MODEL M ₀ : CAPTURE PROBABILITIES ARE		MODEL M _{2b} : CAPTURE PROBABILITIES VARY	
CONSTANT	21	BY INDIVIDUAL ANIMAL AND BY BE-	
<i>Structure and Use of the Model</i>	21	HAVIORAL RESPONSE TO CAPTURE	40
<i>Simulation Results</i>	22	<i>Structure and Use of the Model</i>	40
<i>Example</i>	23	<i>Simulation Results</i>	41
<i>Example</i>	24	<i>Example</i>	42
<i>Discussion</i>	24	<i>Discussion</i>	42
MODEL M ₁ : CAPTURE PROBABILITIES VARY		MODEL M _{2h} : CAPTURE PROBABILITIES VARY	
WITH TIME	24	BY BEHAVIORAL RESPONSE TO CAPTURE,	
<i>Structure and Use of the Model</i>	24	TIME, AND INDIVIDUAL ANIMAL	43
<i>Simulation Results</i>	25	<i>Discussion</i>	43
<i>Example</i>	28	REMOVAL MODELS	44
<i>Discussion</i>	28	<i>Introduction</i>	44
MODEL M ₂ : CAPTURE PROBABILITIES VARY		<i>Structure and Use of the Generalized Re-</i>	
BY BEHAVIORAL RESPONSE TO CAPTURE	28	<i>moval Model</i>	44
<i>Structure and Use of the Model</i>	28	<i>Simulation Results</i>	46
<i>Simulation Results</i>	30	<i>Example</i>	46
<i>Example</i>	31	<i>Example</i>	48
<i>Example</i>	32	<i>Example</i>	49
<i>Discussion</i>	32	<i>Discussion</i>	49
MODEL M ₃ : CAPTURE PROBABILITIES VARY		TESTS OF MODEL ASSUMPTIONS	50
BY INDIVIDUAL ANIMAL	33	<i>Philosophy of the Approach</i>	50
<i>Structure and Use of the Model</i>	33	<i>Summary of Models and Estimators</i>	51
<i>Simulation Results</i>	34	<i>Specific Tests to Perform</i>	53

¹ Present address: U.S. Fish and Wildlife Service, Denver Wildlife Research Center, Denver, Colorado 80225.

² Present address: U.S. Fish and Wildlife Service, Office of Biological Services, Fort Collins, Colorado 80521.

³ Present address: Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545.

<i>On the Need for an Objective Selection Procedure</i>	56	APPENDIX B	105
<i>An Objective Model Selection Procedure</i> ..	57	<i>Estimation in Model M_o</i>	105
<i>Estimation in Alternative Models</i>	61	APPENDIX C	106
<i>Additional Examples of Model Selection</i> ..	62	<i>Estimation in Model M_t</i>	106
<i>A Test for Closure</i>	66	APPENDIX D	107
DENSITY ESTIMATION	67	<i>Estimation in Model M_b</i>	107
<i>Introduction</i>	67	APPENDIX E	108
<i>Problem Formulation</i>	68	<i>Estimation in Model M_h</i>	108
<i>Statistical Treatment</i>	69	APPENDIX F	110
<i>Discussion</i>	72	<i>Discussion of Model M_{tb}</i>	110
STUDY DESIGN	74	APPENDIX G	111
<i>Livetrapping Versus Removal Methods</i> ..	75	<i>Discussion of Model M_{th}</i>	111
<i>Closure</i>	75	APPENDIX H	112
<i>Eliminating Variation due to Time, Behavior, and Heterogeneity</i>	75	<i>Estimation in Model M_{bh}</i>	112
<i>Sample Size</i>	77	APPENDIX I	113
<i>Recording Data</i>	80	<i>Discussion of Model M_{tbh}</i>	113
<i>Data Anomalies</i>	80	APPENDIX J	114
COMPREHENSIVE EXAMPLES	81	<i>Estimation in Removal Models</i>	114
<i>A Taxicab Example</i>	81	APPENDIX K	115
<i>A Penned Rabbit Study</i>	84	<i>Tests of Model Assumptions</i>	115
<i>An Example of Trap Response</i>	87	APPENDIX L	121
<i>An Example Where No Model Fits</i>	93	<i>Density Estimation Based on Subgrids</i> ..	121
COMPREHENSIVE COMPUTER ALGORITHM ..	96	APPENDIX M	123
SUMMARY	96	<i>General Simulation Methods</i>	123
LITERATURE CITED	97	APPENDIX N	123
APPENDIX A	102	<i>Simulation Results</i>	123
<i>Notes on Estimation</i>	102	APPENDIX O	133
		<i>Interval Estimation</i>	133

INTRODUCTION

The estimation of animal abundance is an important problem in both the theoretical and applied biological sciences. Serious work to develop estimation methods began during the 1950s, with a few attempts before that time. The literature on estimation methods has increased tremendously during the past 25 years (Cormack 1968, Seber 1973).

However, in large part, the problem remains unsolved. Past efforts toward comprehensive and systematic estimation of density (D) or population size (N) have been inadequate, in general. While more than 200 papers have been published on the subject, one is generally left without a unified approach to the estimation of abundance of an animal population.

This situation is unfortunate because a number of pressing research problems require such information. In addition, a wide array of environmental assessment studies and biological inventory programs require the estimation of animal abundance. These needs have been further emphasized by the requirement for the preparation of Environmental Impact Statements imposed by the National Environmental Protection Act in 1970.

This publication treats inference procedures for certain types of capture data on closed animal populations. This includes multiple capture–recapture studies (variously called capture–mark–recapture, mark–recapture, or tag–recapture studies) involving livetrapping techniques and removal studies involving kill traps or at least temporary removal of captured individuals during the study. Animals do not necessarily need to be physically trapped; visual sightings of marked animals and electrofishing studies also produce data suitable for the methods described in this monograph.

To provide a frame of reference for what follows, we give an example of a capture–recapture experiment to estimate population size of small animals using live traps. The general field experiment is similar for all capture–recapture

studies (a removal study is, of course, slightly different). A typical field experiment is the following: a number of traps are positioned in the area to be studied, say 144 traps in a 12×12 grid, 7 m apart. At the beginning of the study ($j = 1$) a sample size of n_1 is taken from the population, the animals are marked or tagged for future identification, and then returned to the population, usually at the same point where they were trapped. After allowing time for the marked and unmarked animals to mix, a second sample ($j = 2$, often the following day) of n_2 animals is then taken. The second sample normally contains both marked and unmarked animals. The unmarked animals are marked and all captured animals are released back into the population. This procedure continues for t periods where $t \geq 2$. The animals should be marked in such a way that the capture–recapture history of each animal caught during the study is known. In practice, toes are often clipped to uniquely identify individual animals (Taber and Cowan 1969) or serially numbered tags are sometimes used on larger animals.

Such capture studies are classified by 2 schemes that are directly related to what class of models are appropriate and what parameters can be estimated. The first classification addresses the subject of closure. Closure usually means the size of the population is constant over the period of investigation, i.e., no recruitment (birth or immigration) or losses (death or emigration). This is a strong assumption and, of course, never completely true in a natural biological population. For greater generality, we define closure to mean there are no unknown changes to the initial population. In practice, this means known losses (trap death, or deliberate removals) do not violate our definition of closure. If the study is properly designed, closure can be met at least approximately. Open or nonclosed populations explicitly allow for one or more types of recruitment or losses to operate during the course of the experiment (Jolly 1965, Seber 1965, Robson 1969, Pollock 1975).

Only closed populations will be considered in this monograph.

The second classification depends on the type of data collected with 2 possibilities occurring (Pollock 1974, unpublished doctoral dissertation, Cornell University, Ithaca, New York):

- (1) only information on the recovery of marked animals is available for each sampling occasion, j , $j = 1, 2, \dots, t$.
- (2) information on both marked and unmarked animals is available for each sampling occasion, j , $j = 1, 2, \dots, t$.

In case (1), population size (N) is not identifiable, however, other parameters can be estimated (Brownie et al. 1978). In case (2), N can be estimated using a wide variety of approaches depending upon what we wish to assume. Only case (2) will be dealt with here.

Objectives

The objectives of this publication are twofold:

- (1) to give a thorough treatment of the estimation of population size given multiple capture occasions ($t > 2$) assuming
 - a. population closure,
 - b. there may exist 3 major types of variation in capture probabilities;
- (2) to extend and make available a procedure for estimating density (number of animals per unit area) from grid trapping studies.

This monograph is specifically oriented to the commonly done grid trapping and removal studies where closure can reasonably be assumed. Specifically, we do not treat the case of 2 livetrapping occasions ($t = 2$). This subject (i.e., the Petersen or Lincoln estimators and variations thereof) is adequately covered in the literature (see Seber 1973). In fact, to use the methods presented here for analysis of grid trapping data we suggest the study have 5 or more trapping occasions.

There are some types of study designs

and analysis methods we do not cover here. We do not treat sequential sampling studies (e.g., Samuel 1968), stratified populations (e.g., Darroch 1961, Arnason 1973), Bayesian schemes (e.g., Gaskell and George 1972), or change in ratio estimation (e.g., Paulik and Robson 1969). The subject of stratifying the data after the fact on such variables as sex, age, or species is not discussed primarily because there rarely are enough data for such a stratification. The contingency table approach to estimation from multiple capture studies is a promising new development (see Fienberg 1972), but currently it is relatively unexplored or developed; we do not discuss it. Finally, we do not treat studies or analysis methods for which the goal is to compute only an index to abundance (e.g., captures per 100 trap nights); standard statistical techniques are adequate for those types of studies.

Although our objective is to present comprehensive methods of analysis, the scientist must realize that no amount of sophisticated statistical analysis can compensate for poor study design or field technique (such as high trap losses). The experimenter can do far more to ensure valid estimates by having a properly planned and conducted study than he can by sophisticated analysis after the experiment. We have therefore included a section on statistical aspects of study design. That section includes comments on how to deal with anomalies such as trap losses.

This publication is intended for use by biologists. Such a goal is difficult to attain due to the generally technical and mathematically complex nature of the subject matter. We have developed a comprehensive computer program to compute estimates and test statistics for the various methods covered in subsequent sections (program CAPTURE). Biologists who wish to analyze data are urged to use the computer program rather than to try to compute the various estimates and test statistics by hand. Also, most of the mathematical and statistical details are contained in appendixes to this monograph.

We hope this publication and the associated computer program will be useful within the framework of the assumptions considered.

We undertook the theory development and the writing of this report for a variety of reasons. Several important advances have been made but are available only as unpublished dissertations (Burnham 1972, unpublished doctoral dissertation, Oregon State University, Corvallis, Oregon; Pollock). New methods have emphasized nonparametric approaches that are robust to the failure of certain assumptions. Further, the use of a sequence of statistical models seems appropriate. It is unreasonable to expect a single method to perform well on studies of various species in different habitats, or the same species at different times. Pollock (unpublished dissertation) treated 4 models, each based on specific assumptions, and suggested a statistical testing sequence. That general strategy, followed in this publication, allows models (assumptions) that are inadequate to be rejected for a particular data set. A method inappropriate for field mice *Peromyscus* spp. may work well for voles *Microtus* spp.

There exists a large body of standard statistical theory that is directly relevant and applicable to the estimation problem in capture-recapture and removal studies. Biologists need not, however, learn the theory to be able to use the results of these advanced methods. The methods employed here are often beyond the formal training of most biologists, although they should be able to make proper use of the results. We stress that we have examined the estimation and inference problems in a rigorous statistical framework as opposed to various ad hoc procedures.

Another objective of this monograph is to bring to the biologists' and statisticians' attention the computer program written to implement the complex analyses described here. Without the aid of a computer to do the calculations, development of sophisticated analyses is just an academic exercise. Our philosophy in this matter has been summed up by

Overton and Davis (1969:404): "Computers will soon prove of very great value in the routine processing of census and survey data. When they become generally available, it will be desirable to advance to even more realistic and complex solutions to the problems; there will be no premium on simplicity, so long as the users understand the principles and are able to comprehend the constraints and limitations of the models on which the computer solutions are based."

Assumptions

Every estimation method is based on a set of assumptions. The general assumptions for the capture-recapture methods we present here are listed and discussed below. The assumptions for the removal experiment are given in the section on removal studies. Four assumptions are necessary for the most restrictive experimental situations:

- (1) the population is closed,
- (2) animals do not lose their marks during the experiment,
- (3) all marks are correctly noted and recorded at each trapping occasion j , and
- (4) each animal has a constant and equal probability of capture on each trapping occasion. This also implies that capture and marking do not affect the catchability of the animal.

Before discussing the above, we must emphasize that the focal point of our work has been to relax Assumption 4. That assumption is not met in most capture-recapture studies, and a large percentage of past efforts have been directed at relaxing it. Assumptions 1-3 must be made for all models considered here. We briefly discuss the first 3 and then elaborate on the last in the following section.

(1) *Population closure*.—This assumption arises because population estimation models were initially conceptualized as extensions of urn models (Feller 1950). Such models are basically intended to provide a "snapshot" of the population size at a given point in space and time.

In that context, open and closed models become essentially noncompeting, since open models are more frequently used for purposes of monitoring populations over a longer period of time and obtaining information concerning such properties as survival and recruitment rates. If estimates of population size at a given time are also desired, however, competition between the 2 types of models does arise. In general, open models require more data than closed models due to the fact that assumptions are more rigorous and more parameters are involved. Therefore, feasibility often prohibits the use of very general stochastic models for estimating population size of open populations (Jolly 1965; Seber 1965; Robson 1969; Arnason 1972a, 1972b, 1973; Pollock 1975). If, for example, a 10-day experiment is considered, 17 basic parameters would have to be estimated using Jolly's (1965) model. Hence, data from many population estimation experiments are inadequate for obtaining estimates with acceptable precision and small bias using models for open populations. Moreover, unlike the models treated here, none of those open population models allows for unequal capture probabilities of individual animals. Let it be clear, we believe that well-developed, general models for capture data from open populations are essential in some studies. However, we also believe that for many populations of interest, the closure assumption can be met approximately and the models discussed in this monograph will be useful. For example, closure might be assumed for an 8-day study of cottontails *Sylvilagus* spp. during a nonbreeding period in a well-defined (sampled) area.

A number of tests for closure have been derived (Robson and Flick 1965, Robson and Regier 1968, Pollock et al. 1974), but they generally have little chance of rejecting closure unless the sample is large and there is a marked departure from closure. In addition, closure tests are often confounded with behavioral response to capture, e.g., an animal that becomes un-

catchable, or nearly so, is indistinguishable from one that dies or emigrates. Pollock (1972, unpublished master's thesis, Cornell University, Ithaca, New York) discussed a test for mortality in some detail. The tests for recruitment are more difficult. Thus, the biologist is forced to consider carefully the design of such studies in an effort to assure that the closure assumption is met. Finally, we note that the tests for closure implicitly assume equal capture probabilities; therefore, such tests can reject closure when in fact closure is true but equal capture probability is false. This greatly lessens the value and power of such tests. We believe closure will have to be assessed largely from a biological basis rather than from any definitive statistical tests.

The closure assumption can be relaxed in some cases. Seber (1973:70-71) showed that natural mortality will not bias some estimators if it acts equally on marked and unmarked segments of the population. In such cases, the population estimate then relates to the size of the population at the beginning of the study. However, if recruitment and mortality occur during the experiment, the estimate of N will be too high, on the average, for both initial and final population size (Robson and Regier 1968).

(2) *Permanency of marks.*—Loss of marks (tags) violates the closure assumption and will result in an overestimate of N . If the study is of short duration (to help assure the closure assumption), it seems that loss of marks will generally be a minor problem. Some exceptions, such as radioactive isotopes with a very short half-life, undoubtedly occur (cf. Seber 1973:93-100).

(3) *Reporting and recording marks (tags).*—This assumption can be easily assured by working carefully. Field reports and keypunched cards should be edited and verified. Often, a pilot study may be beneficial to train personnel and identify any problems with the marking method.

Unequal Capture Probabilities

The fourth assumption is particularly important and, for this reason, we focus on it here. It is now widely recognized that this assumption is commonly not met (e.g., Young *et al.* 1952, Geis 1955a, Huber 1962, Swinebroad 1964). Edwards and Eberhardt (1967), Nixon *et al.* (1967), and Carothers (1973a) provided clear evidence that accurate population estimation usually will require models that provide for unequal probabilities of capture. The effects of unequal capture probabilities on estimates derived from models that assume equal catchabilities have been studied by computer simulation by Burnham and Overton (1969), Manly (1970), Gilbert (1973), and Carothers (1973b). Estimators studied were generally found to be significantly biased when this assumption was violated.

This monograph presents a number of models and estimators developed to relax the critical assumption of equal catchability. We have drawn heavily from the work of Pollock (unpublished dissertation, pers. comm.) and Burnham (unpublished dissertation). Following Pollock (unpublished dissertation), we consider a sequence of models each allowing for different combinations of up to 3 types of unequal capture probabilities:

- (1) capture probabilities vary with time or trapping occasion—Model M_t ,
- (2) capture probabilities vary due to behavioral responses—Model M_b ,
- (3) capture probabilities vary by individual animal—Model M_h ($h = \text{heterogeneity among animals}$).

The *assumptions* regarding unequal capture probabilities are to be explicitly embodied in probability models that describe capture studies.

We agree with Carothers (1973b:146) that equal catchability is an unattainable ideal in natural populations (cf. Seber 1973:81–84). We discuss the 3 simplest ways to relax this assumption.

Model M_t allows capture probabilities to vary by time (e.g., each trapping oc-

casion). This situation may be common even though the number of traps might be fixed during the course of the study. For example, a cold rainy day during the study might reduce activity of the animals and reduce the probability of capture. Also, if different capture methods are used on each occasion, this model could be appropriate.

Model M_b allows capture probabilities to vary by behavioral response or “capture history,” and deals with situations in which animals become trap happy or trap shy. Carothers (1973a) referred to this as a contagion of catchability. This implies that an animal’s behavior tends to be altered after its initial capture (e.g., perhaps the animal was frightened or hurt during initial capture and marking and thereafter it will not likely enter another trap).

Model M_h allows capture probabilities to vary by individual animal. This situation has been modeled only with great difficulty and requires that additional distributional assumptions be made. Individual heterogeneity of capture may arise in many ways. Perhaps accessibility to traps (as influenced by individual home ranges), social dominance, or differences in age or sex can cause such an unequal probability structure. This is an important type of variation and has been rigorously treated by Burnham (unpublished dissertation), whose nonparametric approach is presented in a later section.

In addition to these 3 simple models, we consider all possible combinations of the 3 types of unequal capture probabilities (i.e., Models M_{tb} , M_{th} , M_{bh} , and M_{tbh}). We also treat the “null” case in which capture probability is constant with respect to all factors (Model M_0). Model M_0 corresponds to the 4 assumptions listed earlier. For simplicity, we denote estimators of population size for a specific model using the same subscript notation. For example \hat{N}_0 denotes the estimator derived from Model M_0 ; \hat{N}_t denotes the estimator derived from Model M_t ; \hat{N}_{bh} denotes the estimator derived from Model M_{bh} , and so on.

Perspectives

We wish to emphasize that a specific set of assumptions is the basis for a specific model. The assumptions and model then represent a *tentative* hypothesis when analyzing the results of a particular capture experiment conducted to estimate population size or density. Cormack (1968:456) stated, "In all cases every iota of information, both biological and statistical, must be gathered to check and countercheck the unavoidable assumptions." Statistical testing within and between models (assumptions) is emphasized here. In spite of this, more work in this direction is clearly indicated. Our approach is to derive models for an array of types of unequal probabilities of capture. We conducted statistical tests to enable selection of an appropriate model for the analysis of a particular data set (cf. Pollock unpublished dissertation). Some models are very sensitive to small departures from the underlying assumptions; therefore, testing between models and investigating the robustness of each estimator are essential.

The importance of such testing is reflected in the fact that use of an inadequate model will often lead to a highly biased estimate of population size. This is perhaps to be expected, if not obvious. More subtle is that estimates of the sampling variance (a measure of precision) are quite dependent on the correct model. Bias of the estimator may be small, but the estimate of variance may be very poor, even with large samples. This can cause, for instance, associated confidence intervals to have very poor properties. The importance of assumptions and their testing cannot be overemphasized. Paulik (1963) noted that an approximately correct estimate with low precision is always better than a highly precise incorrect estimate. Tests of assumptions concerning equal capture probabilities are especially important because estimators based on given sets of assumptions are usually not robust to departures from those assumptions (Seber 1970, Gilbert 1973).

We believe rigorous probability models explicitly incorporating various tentative assumptions represent the best approach toward estimating population size N , or density D . The tentative nature of the assumptions and the general uncertainty about biological processes make testing a key concern. As Seber (1973) pointed out, statistical models should be used with caution, due to lack of control over natural populations. All models depend on the validity of various underlying assumptions—that are often difficult to evaluate rigorously.

Finally, we believe that theory and application must be integrated. Either in the absence of the other will stifle progress. For this reason we have tried to integrate the statistical theory with the biological application. We have, however, tried to separate the more complex subjects and include them as a series of technical appendixes. We urge biologists to try to consider and understand the appendixes, and we ask statisticians to continue to be concerned with the biological complications and realities before attempting additional theory development. Through an integrated team approach we can expect further progress on this series of estimation problems.

Comments on the Use of This Monograph

We cover several topics here, and present mathematical as well as applied results. Topics covered include data analysis of short-term livetrapping and constant effort removal studies, design of such live trapping studies, and simulation results on inference procedures. Numerous examples are also given. A variety of uses of this monograph are anticipated by: (1) biologists who must analyze actual data, (2) biologists (and statisticians) faced with designing capture studies, (3) persons interested in performance of estimators presented here, (4) statisticians interested in developing more advanced models, and (5) educators who seek to teach courses on the subject of population size estimation.

Biologists who have data from closed population livetrapping studies will have to read quite a bit of this monograph before they can understand the methods. They do not need to read the appendixes. They would have to understand all sections through TESTS OF MODEL ASSUMPTIONS, except REMOVAL MODELS. We believe this can be done by anyone having had a solid course in college level algebra and beginning statistics. In order to understand the essence of what we present, the reader does not have to follow all the mathematical descriptions of models nor discussions of model properties. We have included numerous examples. In particular, the reader should benefit greatly from the section on COMPREHENSIVE EXAMPLES.

If the reader intends to do, say, data analysis according to these methods, it is virtually necessary to use program CAPTURE (see COMPREHENSIVE COMPUTER ALGORITHM). This program is available and there is a user's manual for it. Persons with many data sets to be analyzed should get the program. Conversely, we do not recommend trying to implement this computer program if one has only a few (or one) data sets to analyze. In this latter case, it is better to have the data run for you. The authors are willing to assist in running such data provided the user arranges his own keypunching of the data in the necessary format (we can supply this format).

If one's goal is to analyze some removal data, the relevant sections are those on Models M_b , M_{bh} , and the removal models (plus the introductory sections). Again, the authors would try to help users analyze removal data; within reasonable limits we may be able to run the data analyses or assist in setting them up.

Many readers will sometimes be faced with designing a capture study. The section on STUDY DESIGN covers some fundamental design aspects of livetrapping studies for closed populations. If your goal is to design a study, read that section at a minimum; to get full advantage of this monograph in terms of design, you

will need to read most sections, exceptions being HISTORICAL OVERVIEW, REMOVAL MODELS, and material following the STUDY DESIGN section.

If you are interested in obtaining insights into the performance of various estimators, you should put special effort into studying the numerous simulation results presented here. This would require reading almost all the text and careful study of Appendix N.

Persons interested in doing further research along the lines of the models and approach of this monograph will have to carefully study almost everything here, especially the appendixes.

Finally, this monograph and program CAPTURE have value for teaching and learning about population size estimation. The simulation feature of CAPTURE can be especially valuable in teaching the concepts of sampling variation and properties of estimators. Persons interested in performing such simulation of the methods presented here (either for design of studies, evaluation of estimators, or teaching purposes) will need to implement the program for their own use.

ACKNOWLEDGMENTS

Dr. K. H. Pollock, University of Reading, provided several ideas and criticisms through correspondence; in addition, we have drawn on the results of his doctoral program. Drs. M. H. Smith and J. B. Gentry, Savannah River Ecology Laboratory, contributed to this work through discussion and provided samples of their research to be used for program testing. Data to be used as examples were provided by A. D. Carothers, H. N. Coulombe, C. T. Cushwa, W. R. Edwards, S. W. Hoffman, E. C. Larsen, R. F. Raleigh, and V. H. Reid.

We appreciate extensive comments made by A. N. Arnason, A. D. Carothers, W. R. Clark, S. W. Hoffman, and R. P. Davison on an earlier draft of the manuscript. The comments by A. N. Arnason and G. A. F. Seber on the final version of

the manuscript are also much appreciated.

Dr. R. G. Streeter, U.S. Fish and Wildlife Service, provided encouragement during the investigations.

This work was performed under Contract 14-16-0008-1224 of the Coal Program, Office of Biological Services, U.S. Fish and Wildlife Service, to the Utah Cooperative Wildlife Research Unit. Funds for this work were made available to the Fish and Wildlife Service as part of the Federal Interagency Energy/Environment Research and Development Program, Office of Research and Development, U.S. Environmental Protection Agency (IAG-EPA-D5-E385). Publication costs were paid by the Oil Shale Program, Office of Biological Services, U.S. Fish and Wildlife Service, contract 14-16-0008-1197. Final debugging of the computer program, and preparation of examples were performed under the auspices of the U.S. Energy Research and Development Administration.

FUNDAMENTAL CONCEPTS

This section presents notation and discusses the statistical techniques used in this monograph. The subject matter is basically technical in nature, but we have tried to keep the presentation simple and refer the reader to appendixes for more complex details. We believe it is important for users of the methods described in this publication to understand the material presented in this section.

Data and Parameters

All the models discussed here assume population closure (except for known removals). Therefore, the parameter we wish to estimate is population size N which is constant. Moreover, because the same individual animals compose the population on each trapping occasion, $j = 1, 2, \dots, t$, we can conceive of the individuals as being numbered $i = 1, 2, \dots, N$.

The basic capture data are conveniently expressed in matrix form as

$$[X_{ij}] = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1t} \\ X_{21} & X_{22} & \dots & X_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \dots & X_{Nt} \end{bmatrix}$$

where

$$X_{ij} = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ animal is caught on the } \\ & j^{\text{th}} \text{ occasion} \\ 0 & \text{otherwise.} \end{cases}$$

The X matrix is a simple way to record the capture or noncapture of each animal in the population on each trapping occasion. Row i gives the trapping results for individual i , while column j gives the results of the j^{th} trapping occasion. Note that the matrix X may not be observed in its entirety because some animals may never be captured; therefore, those rows of X are all zeros.

A series of specific models for capture data can be derived if we define the following general structural model:

$$p_{ij} = \text{the capture probability of the } i^{\text{th}} \text{ individual in the population on the } j^{\text{th}} \text{ trapping occasion, where } i = 1, 2, \dots, N, \text{ and } j = 1, 2, \dots, t.$$

For example, if we assume the restrictions $p_{ij} = p$ for all i and j we get Model M_0 , the simplest possible model. All other models we introduce may be thought of as generalizations of Model M_0 . Hence, in the following sections, models are developed based upon capture probabilities being time specific, behaviorally related, or differing among individual animals. Therefore, capture probabilities are the crucial element of the series of models we discuss.

The above structure and assumptions suffice to specify the marginal distribution of each individual X_{ij} (i.e., they are Bernoulli random variables); however, they do not specify the joint distribution of all X_{ij} . Therefore, we have assumed joint independence of the variables in order to have a completely specified general model structure. Specifically, we assume that given the correct model (i.e., the correct specification of capture probabilities p_{ij}), then the elements of X_{ij} are

mutually independent random variables. This assumption is not testable unless one first knows what the correct model is (which we never will for real data). It is our opinion this is not a restrictive assumption and it need not be a source of concern.

Statistics and Notation

Probability models from which estimators of population size N may be developed are discussed in following sections. A few simple statistics are needed for these models. They are defined and discussed below.

n_j = the number of animals captured in the j^{th} sample, $j = 1, 2, \dots, t$,

$$= \sum_{i=1}^N X_{ij},$$

n = the total number of captures during the study $= \sum_{j=1}^t n_j$,

u_j = the number of new (unmarked) animals captured in the j^{th} sample, $j = 1, 2, \dots, t$,

f_j = the captures frequencies = the number of individuals captured exactly j times in t days of trapping, $j = 1, 2, \dots, t$. f_0 will be used for the number of individuals never captured (obviously, f_0 is not observable).

M_{t+1} = the number of distinct individuals caught during the experiment (recall that t is fixed for a given experiment),

$$= \sum_{j=1}^t f_j = \sum_{j=1}^t u_j,$$

M_j = the number of marked animals in the population at the time of the j^{th} sample, $j = 2, 3, \dots, t$. (Note that $M_1 \equiv 0$),

M = sum of the M_j [does not include M_{t+1}] $= \sum_{j=1}^t M_j$,

m_j = the number of marked animals captured in the j^{th} sample, $j =$

$2, \dots, t$. Note that $u_j = n_j - m_j$ and that $m_1 = 0$,

$$m = \text{sum of the } m_j = \sum_{j=1}^t m_j.$$

The statistics u_j , f_j , M_j , and m_j may also be computed directly from the X matrix. However, the computation is not as straightforward as that of n_j and is not given here. We denote X_ω as the number of animals with a specific capture history ω . For example X_{10011} represents those individuals caught on trapping occasions 1, 4, and 5. The set of all the possible capture histories will be symbolized as $\{X_\omega\}$. In general, with t capture occasions there are 2^t possible capture histories.

Two other terms used frequently in this monograph are:

Robustness (of an estimator).—A robust estimator is one that is not sensitive to the breakdown of a particular assumption. A specific measure of robustness is difficult to define. Therefore, a somewhat subjective determination regarding the robustness of an estimator is made relative to the general performance of the estimator. Performance is evaluated with respect to the essential criteria of bias, precision, and confidence interval coverage. For example, the estimator developed under Model M_t performs very poorly with respect to all criteria if individual heterogeneity to capture is present in the population. We say that this estimator is not robust to a particular assumption. In contrast, the jackknife estimator for Model M_h appears to be fairly robust for a number of specific assumptions.

Bias (of an estimator).—Bias is the difference between the expected value of an estimator and the true parameter being estimated, e.g., $B = E(\hat{N}) - N$. Percent relative bias, $100\{[E(\hat{N}) - N]/N\}$ is denoted as RB. Overton and Davis (1969) gave a good discussion of these and other related terms.

A final note concerns the differences between parameters (true values) and estimates. We are concerned chiefly with

making estimates of the parameters N and D , population size and density, respectively. We denote our sample estimators of these parameters as \hat{N} and \hat{D} . Biologists are referred to Kendall and Buckland (1971) for definitions of standard statistical terms.

Parameter Estimation

The data from capture-recapture or removal studies are *samples*. This imposes the need for a probabilistic treatment of the data to derive correct estimation and inference procedures. The models we consider here are termed stochastic models. Unlike the models for open populations, the only stochastic component for models under population closure relates to the sampling process: i.e., the capture probabilities. Model formulation in this context begins with a set of explicit assumptions. A probability model for the sampling distribution of the X matrix (the basic data) is derived to quantitatively express the assumptions. A probability function is a form of mathematical representation of the observed data under a specific set of assumptions. It provides a basis for quantitatively and explicitly incorporating the specific assumptions about capture probabilities and for developing the point and interval estimators by rigorous statistical estimation techniques.

Most parameter estimators in this publication were derived using the method of maximum likelihood (ML). Several models and their corresponding estimators were taken from existing literature (e.g., Zippin 1956, Darroch 1958), often with some modification. Other models and estimators were derived during the course of this study.

Estimators found by the ML method are optimal, at least for large samples. (For a discussion of optimality, refer to Appendix A). This is a generally accepted tenet of statistical estimation theory (Mood et al. 1974). In general, ML estimators of unknown parameters (e.g., N) are found by application of results from simple cal-

culus, using the likelihood function derived from the probability model. In some cases, the estimator may take a simple, easy to use form. For example, the ML estimator of N for Model M_t for 2 sample occasions ($t = 2$) is the Petersen estimator

$$\hat{N}_t = \frac{n_1 n_2}{m_2}$$

where n_1 , n_2 , and m_2 have already been defined.

However, in capture-recapture models we rarely find that the exact ML estimators exist as a simple formula as above. To illustrate this, consider the model developed by Darroch (1958) when 4 sampling occasions are considered ($t = 4$) and the capture probabilities are assumed to vary only by time (i.e., p_1 , p_2 , p_3 , and p_4). The approximate ML estimator of N for this model (see Darroch 1958) is the unique solution of the equation

$$\left(1 - \frac{M_5}{N}\right) = \left(1 - \frac{n_1}{N}\right)\left(1 - \frac{n_2}{N}\right) \cdot \left(1 - \frac{n_3}{N}\right)\left(1 - \frac{n_4}{N}\right).$$

In general, for Model M_t , the ML estimator is the solution of the equation

$$\left(1 - \frac{M_{t+1}}{N}\right) = \prod_{j=1}^t \left(1 - \frac{n_j}{N}\right).$$

For t greater than 2, this equation cannot be solved algebraically for N . In other words, it is not possible to arrange the symbols algebraically in such a way that only N appears on one side of the equation and all other terms appear on the other side. The equation can be solved, but only on a case by case basis using a numerical procedure. We say the equation does not have a simple, "closed form" solution. Complex probability models often do not have simple estimators and tests of assumptions; nonetheless, complex models appear necessary to describe many capture-recapture studies adequately.

Our work has shown that several of the

approximations of \hat{N} suggested for this model in the past are fairly poor. Furthermore, some of the iterative solutions given (e.g., Darroch 1958) produce only approximate ML estimates. We have obtained exact ML estimators for all the models in this publication (except Models M_h , M_{th} , M_{tb} , and M_{bth}) by employing numerical procedures on a digital computer (in fact no estimators can be derived for the latter 3 models). We have found the maximum of the likelihood function in such a way as to obtain exact integer valued ML estimators of N . The disadvantage here, of course, is that we cannot show simple closed form estimators. This subject is discussed further in Appendix A.

We find that the estimator of N under each model involves only simple statistics computed from the X matrix. Individual captures are not employed—only various sums (linear combinations) derived from the X matrix. Those sums are statistics such as n_j , n , u_j , and M_{t+1} . For any model we consider, there exists a set of simple statistics, called minimal sufficient statistics (MSS). Estimators should be based on only MSS. The use of the ML method results in estimators that are always functions of the MSS. This is a desired property because it can be shown that the MSS contains all the information available from the experiment for estimating the parameter(s) of interest (in our case N). An estimator based on statistics other than MSS is not using all available information and is, therefore, not optimal. Some statistical tests of assumptions will depend on information other than MSS.

The number of parameters that can be identified (estimated) is less than or equal to the number of elements in the MSS (regardless of the estimation method used). The subject of “identifiability” of parameters is important in the material that follows and, therefore, we place some emphasis on MSS. For example, under Model M_0 , the MSS is n , and M_{t+1} , whereas under Model M_h , the MSS is f_j , $j = 1, 2, \dots, t$. We make frequent use of the MSS in the following sections.

Interval Estimation

One of the several advantages of the probability model/ML approach is that estimates of sampling variance and covariances can be computed as part of the ML method. These measures of precision are essential in making inferences from the sample results of the experiment. The variance and covariance estimators are derived from “large sample” theory and usually are of unknown value as measures of precision in “smaller” samples. We have performed a large number of Monte Carlo simulation experiments (Appendixes M and N) to examine the small sample properties of such variance estimators and the confidence intervals that depend on them. Interval estimation is an old subject in the statistical literature, and we refer the interested reader to the text by Mood *et al.* (1974) for details. Seber (1973) also gave numerous examples.

Typically, the ML estimator of N is not normally distributed unless large samples are taken. Because confidence intervals commonly used depend upon an assumption of normality, we explored alternative interval estimation techniques (Appendix O). The alternative procedures were not totally satisfactory and we will use the standard procedure as follows to construct an approximate 95 percent confidence interval on N :

$$\hat{N} \pm 1.96\sqrt{\hat{\text{Var}}(\hat{N})}$$

This procedure has its limitations but, all things considered, appears to be best at present.

HISTORICAL OVERVIEW

Although the basic concept of obtaining information about an animal population by marking some of its members may be traced as far back as the 17th century (Chapman 1948) and to Petersen’s (1896) expression of the fundamental principle, one may argue that the practical beginnings of the literature concerning the marking method can be associated with

Lincoln's (1930) use of band returns to estimate the size of the North American waterfowl population. In the nearly 50 years since Lincoln's (1930) initial work, a voluminous literature has resulted from efforts directed toward deriving and refining techniques based on the capture-recapture method. In the past decade, 2 notable attempts to summarize the existing literature have been made. The first is by Cormack (1968) who provided a survey of mathematical models proposed for use in capture-recapture experiments. The second is an extensive text by Seber (1973) that attempted to bring together all the proposed techniques for estimating population abundance and related parameters. Included as a subset of those techniques are those concerned with the capture-recapture method. In the presentation of those techniques, a substantial amount of mathematical detail is provided, as are numerical examples. Furthermore, assumptions that must be met to ensure validity of a particular technique are presented; methods for testing the validity of some of the assumptions are given.

Because of the existence and quality of the cited works, no attempt is made here to present specific methods associated with the theory of capture-recapture and related experiments. Rather, we present a review that follows the chronological development of conceptual approaches in the literature.

The initial state of the art is well characterized by Lincoln's (1930:2) statement of the solution to his population estimation problem: "Given a fairly accurate statement showing the number of wild ducks killed in North America in any one season, then the total number of ducks present on the continent for that season may be estimated by a percentage computation, based upon the relation that the total number of banded ducks killed during their first season as band carriers bears to the total number banded." As one might expect, no mention is made of the statistical properties such an estimator might possess or of the underlying

assumptions, such as random sampling, that influence the validity of the method. Such considerations were at least hinted at, however, in Schnabel's (1938) paper that extended the method to the situation in which members of the population were marked and released back into the population on more than 1 occasion. Mention is made of the percentage relative bias of the estimators in an experiment in which the population size is known, and the reader is cautioned that "none of the solutions can be expected to provide more than an estimate of the general order of magnitude of the total population" (Schnabel 1938:352). Presumably, some caution is generated by the fact that "assumptions of random sampling and constant population are only rough approximations to the actual situation" (Schnabel 1938:352). More consideration was given to the uses of the capture-recapture technique in a sequence of papers by Jackson (1933, 1937, 1939, 1940), who was concerned not only with estimating population size but also with birth-immigration and death-emigration parameters. Contained in the consideration of those parameters is the concept that the population is not "closed," i.e., population size is not constant throughout the sampling period. Those methods were then being applied mainly to fish and insect populations and not to terrestrial wildlife populations, although some exceptions did exist (e.g., Green and Evans 1940). Scepticism as to the worth of the method with respect to wildlife populations was expressed by Dice (1941:402), who stated that "the application of the proportional method of calculating mammalian populations may often require as much effort as the complete trapping or counting of the whole sample population." Nonetheless, effort continued in development of the theory. Schumacher and Eschmeyer (1943) provided an alternative solution to that of Schnabel (1938) by the use of regression techniques. Evidently, their work was spurred by the desire to develop an estimator that would be more robust to de-

partures from the underlying assumptions of the Schnabel method. Moreover, unlike Schnabel, they provided an estimator for the standard error of the estimate. Similar regression techniques were also investigated by Hayne (1949a) and DeLury (1958).

A significant change in both the quality and quantity of work in the field of capture-recapture theory occurred with the appearance of several important papers of the early 1950s. Those papers signaled the beginning of a more rigorous mathematical treatment of the theory in terms of both estimation and testing of assumptions. Bailey (1951), for example, proposed a binomial model for the single mark-release situation. He used the ML theory to develop an estimator of the precision of the population size estimator, the latter estimator being the same as Lincoln's (1930). Chapman (1952) considered a hypergeometric model for the multiple capture-recapture experiment and derived an approximate expression for the resulting ML estimator. He also gave a test for determining whether the probability of capture is independent of tagging. Alternative sampling schemes were proposed by such authors as Chapman (1952, 1954) and Goodman (1953). Such schemes were designed to avoid undesirable statistical properties associated with the direct sampling method that considers the total number of animals caught on each occasion as a fixed parameter. For instance, Chapman (1952) pointed out that an estimator of population size obtained via inverse sampling (i.e., considering the number of *marked* animals caught on each occasion as fixed) is unbiased, whereas the estimator associated with direct sampling is biased. In addition, removal data, similar to the type of data used in marking experiments, was used in alternative methods proposed by Moran (1951) and DeLury (1951). Also at that time, progress was made in the theory of estimation in open populations through a sequence of papers by Leslie and Chitty (1951), Leslie (1952), and Leslie et al. (1953). Those authors used ML

theory for estimating such parameters as death rate and population size, and devoted much effort to the examination of assumptions.

The appearance of such mathematical treatments generated most of the important immediately succeeding work on the development of the theory. As an example, one can consider the work of Zippin (1956), who provided a more complete statistical treatment of the removal method first suggested by Moran (1951). An important example is the work of Darroch (1958), who was responsible for the correct derivation of the probability model for the multiple capture-recapture experiment first treated by Schnabel (1938). Moreover, Darroch presented expressions for the asymptotic bias and approximate variance for his approximate ML estimator and a method for constructing confidence intervals. Darroch's (1958, 1959) work on the closed model, the birth only and death only models stands as a cornerstone in the development of the theory.

Given the methods available, it was now possible for researchers to direct effort toward the development of statistical tests of assumptions underlying the methods of estimation. One of the most generally invoked assumptions of proposed estimation techniques was (and still is) that all animals in the population, regardless of capture history and other individual characteristics, are equally at risk to capture on each trapping occasion. Leslie (1958) devised a test directed toward that hypothesis, which was later extended by Carothers (1971). Cormack (1966) made the important point that failure of the above assumption may be caused either by each animal in the population possessing an "innate catchability" which varies among individuals over the population, or by an individual's probability of capture being affected by its capture history, or both. Cormack (1966) provided a test for the former assuming the latter is false. Seber (1962, 1965) and Robson and Youngs (1971) considered the problem of testing whether

marking an animal affects its probability of capture on subsequent trapping occasions, and Manly (1971) provided a method for estimating the effect of marking on survival of the animal. During that period, Seber (1965) and Jolly (1965) independently developed what is now known as the Jolly-Seber method of estimating open population parameters from multiple capture-recapture experiments. That model, aspects of which were later generalized by Robson (1969) and Pollock (1975), allows the population to be experiencing death, recruitment, immigration, and permanent emigration. Arnason and Baniuk (1977) provided a comprehensive computer algorithm to compute estimates for various models for open populations. Existence of such open population models points out the need for tests for closure of the population under study. Unfortunately, good tests of that assumption are still not available.

The importance of developing and using valid tests of model assumptions was further emphasized by results appearing simultaneously in the literature concerning the operating characteristics of existing estimation techniques. Edwards and Eberhardt's (1967) study on a confined rabbit population of known size revealed large biases in both the Schnabel (1938) and Schumacher-Eschmeyer (1943) methods of estimation. The authors conjectured that those biases were due to "individual animals having different or changing probabilities of capture." A simulation study by Braaten (1969) indicated serious bias in the estimators derived from DeLury's (1947) catch-effort model if the assumption of "constant catchability" is violated. A similar lack of robustness to unequal capture probabilities among animals was exhibited by estimators examined in a computer simulation study by Burnham and Overton (1969), who generated "populations" using the family of beta distributions. More recently, Carothers (1973b) sampled a population of known size, the members of which were the taxicabs of the city of Edinburgh. The assumption of popula-

tion closure was reasonable, and non-homogeneous individual capture probabilities were caused by the sampling schemes used. Various "Schnabel type" estimators, that assume equal capture probabilities, were reported as having substantial bias. In addition, 2 regression type estimators proposed by Tanaka and Kanamori (1967) and Marten (1970), each of which assumed a certain form of unequal capture probabilities, failed to reduce significantly the magnitude of the bias of the "Schnabel" estimators.

The appearance in 1965 of the Jolly-Seber method of estimating parameters of open populations did not preclude the development of additional estimation techniques in the literature, in spite of the fact that Cormack (1968:487) believed the method to be "an extremely powerful general formulation" of the capture-recapture experiment. Although the method is general in the sense that it allows for such processes as recruitment and mortality, it is restricted by the assumption that all animals have the same probability of capture on a given trapping occasion. In many experimental situations, the assumptions of population closure and unequal capture probabilities constitute a more realistic set of assumptions than the set required for the Jolly-Seber model. Hence, parameter estimators derived from models based on different sets of assumptions than the Jolly-Seber models continued to be developed by researchers such as Tanaka and Kanamori (1967), Eberhardt (1969a), and Marten (1970). The jackknife technique for bias reduction proposed by Quenouille (1949, 1956) was used by Burnham (unpublished dissertation) to derive an estimator for the situation in which each member of the population has an "innate" probability of capture that varies among individuals. That effort represents a unique attempt to develop a robust estimator of population size that is non-parametric, i.e., one that does not need to assume how capture probabilities are distributed over the population. Such non-parametric approaches are appealing

because they are robust to specific assumptions regarding the experiment and tend not to suffer from breakdown of specific assumptions used to parameterize the model. Pollock (unpublished dissertation) also considered estimation of population size under the assumption of heterogeneity of capture probabilities, but with the added complication that an animal's probability of capture may be altered by its capture history. However, no specific estimation procedure for that model had been proposed in the literature until the appearance of the generalized removal method described in this monograph.

This overview would not be complete without making some observations concerning methods of density estimation in capture-recapture experiments. The notion that the effective area of trapping is greater than the actual area of the trapping grid (i.e., the so-called edge effect) has long been recognized. Dice (1938, 1941) corrected for the effect by adding to the grid area a strip of one-half the home range of the animal, and that remains the most common practice at present. Other authors (Stickel 1954, Mohr and Stumpf 1966, Smith et al. 1975) have used recapture radii to correct for edge effect bias. Assessment lines have also been used to estimate density (Kaufman et al. 1971, Smith et al. 1971). More recently, Burnham and Cushwa (pers. comm.) have formalized MacLulich's (1951) technique for estimating density that involves using concentric trap grids to allow simultaneous estimation of density and edge width.

An underlying theme of this historical overview is that any capture-recapture experiment requires that the researcher make specific assumptions concerning the many factors that affect the results of the experiment. The assumptions that are chosen determine which statistical estimation procedures should produce the best results available from the data. Many estimation procedures have been proposed because several different assumptions can often be made for a given factor.

Unfortunately, it has been shown that misinformation results if, for a given experiment, assumptions are not valid or statistical estimators are not appropriate or both. Thus, it should be obvious that a rigorous approach to parameter estimation in capture-recapture experiments will include a statistical testing algorithm that allows the data to aid in selection of the "best" set of assumptions for the experiment. Although some tests of specific assumptions have been introduced, unified approaches to the problem have not, for the most part, received attention in the literature (an exception is the work of Pollock, unpublished dissertation). The concept of a unified approach is the basis for the development of this monograph. We believe an approach based on such a concept is a step in the direction of improved analyses of data from capture-recapture experiments. Furthermore, we hope that future research will be directed to that same objective.

MODEL M_0 : CAPTURE PROBABILITIES ARE CONSTANT

Structure and Use of the Model

Assumptions and Parameters

The simplest of all models under consideration results from the assumption that all members of the population are equally at risk to capture on every trapping occasion. Moreover, the occasions themselves do not affect capture probabilities. We thus have a model in which there is no heterogeneity of capture probability, no behavioral response to capture, and no variation in the experimental situation over time. This model is designated Model M_0 , and involves only 2 parameters: N , the population size, and p , the probability that an animal is captured on any given trapping occasion.

Statistical Treatment

The probability distribution of the set of possible capture histories $\{X_{\omega}\}$ is given by (cf. Darroch 1958):

$$P[\{X_{\omega}\}] = \frac{N!}{\left[\prod_{\omega} X_{\omega}! \right] (N - M_{t+1})!} \cdot p^n \cdot (1 - p)^{tN - n},$$

where $n = \sum_{j=1}^t n_j$ = total number of captures in the experiment, and

M_{t+1} = number of different animals captured in the experiment.

An algorithm for producing ML estimators of N and p is derived in Appendix B. (When $t = 2$, a closed form ML estimator of N exists and is given by $\hat{N}_0 = (n_1 + n_2)^2 / 4m_2$, where m_2 is the number of recaptures in the second sample.) These estimators are necessarily functions of the minimal sufficient statistic $\{n., M_{t+1}\}$. Thus, all the information relevant for estimation purposes is contained in the number of different animals captured and the total number of animals captured during the course of the experiment. Appendix B also gives an estimator for the asymptotic variance of \hat{N}_0 that we used in the construction of confidence intervals for N .

Simulation Results

Bias

A computer was used to simulate experiments from populations satisfying the assumptions of Model M_0 . By varying the population parameters N and p , some insight into the small sample bias of \hat{N} was obtained. Results indicate that the bias of \hat{N} is negligible for values of p at least as large as 0.10 and $t \geq 5$. For smaller probabilities of capture, however, positive relative biases of 15–20 percent are realized. For example, from Appendix N, Table N.1.b, for a population of size $N = 400$, one simulation consisting of 200 replications with $p = 0.10$ and $t = 5$ produced an average value of \hat{N}_0 of 406.0 (Trial 2), while another, based on 500 replications, produced an average value of \hat{N}_0 of 456.9 with $p = 0.05$ (Trial 3). Complete results of the simulation of \hat{N}_0

for Model M_0 are given in Table N.1.b of Appendix N.

Confidence Intervals

Achieved confidence coefficients of the confidence interval procedure simulated were consistently at or above the 0.90 level and hence were close to the claimed 0.95 coverage. However, the width of an average interval is so large for small values of p that not much information concerning true population size is provided. For instance, for $N = 400$, $t = 5$, and $p = 0.05$ (Trial 3) expected width $[= 2 \cdot 1.96 \cdot \text{Ave} \sqrt{\hat{\text{Var}}(\hat{N})}]$ is 628.2; and for $N = 400$, $t = 5$, $p = 0.10$ (Trial 2) this value is 217.7. However, with $p = 0.30$ (Trial 1) expected width drops to an average of 46.9, indicating that the model provides useful information concerning N when p is reasonably large. The number of replications for these 3 examples were 500, 200, and 200, respectively. One should keep in mind that extremely wide confidence intervals tend to reveal poor experimental conditions, i.e., low values of p , and thus can be of use in providing the experimenter with information concerning the success or failure of the experiment. See Table N.1.b of Appendix N for further details of the simulation results.

Robustness

Because Model M_0 is built from the assumption that no factors that affect capture probabilities are present in the experiment, it is not surprising that simulation results reveal that the estimator derived from this model is not robust to any type of variability in the capture probabilities. In particular, if capture probabilities vary by animal \hat{N}_0 exhibits significant negative bias. This property has been documented in the literature (Robson and Regier 1964, Gilbert 1973, Carothers 1973b). Common sense and some reflection on the nature of the experiment should tell us not only that behavioral response will cause bias in the

estimator, but will also indicate the direction of that bias. That is, animals becoming trap shy will cause \hat{N}_0 to overestimate N , and vice versa when animals become trap addicted. These assertions are supported by the simulation results given in Tables N.3.b and N.4.b of Appendix N. Results also reveal that the estimator is somewhat robust to changes in capture probabilities over time. However, Seber (1973) recommended, on the basis of Darroch's (1959) work, that the estimator associated with Model M_0 not be used even if the capture probabilities are suspected of not varying with time. This is good advice if large numbers of animals are being caught but such a rule could result in some loss of efficiency for small sample sizes.

Example

A capture-recapture experiment that satisfies the conditions of Model M_0 can be analogous to an urn experiment, a sampling experiment conducted in order to estimate the number of marbles in a cylinder when all marbles are the same size. We may visualize a cylinder containing N white marbles (individuals), each of which has an equal probability (p) of being picked from the cylinder on any given occasion. On each of t occasions, the following sampling scheme is carried out. A "sampling cylinder," with a diameter that is 100 percent of the diameter of the cylinder containing the marbles, is inserted into the container and a random sample of marbles removed. The numbers of white (individuals not previously "captured") and black ("recaptures") marbles in the sample are recorded. All white marbles are painted black and returned to the container along with the black marbles, and all the marbles are randomly mixed. The number of white and black marbles in the sample is recorded. Using the data from these t samples, the estimation procedure associated with Model M_0 provides the appropriate estimator of N , the number of marbles in the cylinder. Notice that all

NUMBER OF TRAPPING OCCASIONS WAS	5
NUMBER OF ANIMALS CAPTURED, $M_{(t+1)}$, WAS	98
TOTAL NUMBER OF CAPTURES, n , WAS	238
ESTIMATED PROBABILITY OF CAPTURE, \hat{p} = .4672	
POPULATION ESTIMATE IS	102 WITH STANDARD ERROR 2.3856
APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL	97 TO 107

FIG. 1. Example of population estimation with constant probability of capture under Model M_0 with simulated data based on $N = 100$, $t = 5$, and $p = 0.5$.

the assumptions of Model M_0 are fulfilled for this cylinder model. That is, the population is closed because marbles may not enter or leave the container, and every individual has the same probability of capture on every trapping occasion because (1) all marbles are the same size and thus are not "heterogeneous," (2) white and black marbles have the same capture probability and thus there is no "behavioral response to capture," and (3) the same "sampling cylinder" is used in the same manner on all t occasions and thus there is no "time variation."

The fact that an analogy can be drawn between a capture-recapture experiment modeled by Model M_0 and the simple urn experiment illustrates the point that it is not reasonable to expect that many capture-recapture studies can be adequately represented by Model M_0 . Therefore, to present an example of the estimation procedure of Model M_0 , we simulated capture-recapture sampling for 5 occasions on a population of 100 individuals, each of which had a 0.5 probability of capture. As Fig. 1 shows, the value of the minimal sufficient statistic $\{n, M_{t+1}\}$ is $\{238, 98\}$. These values, and the value of t , are used to produce the population estimate of 102. Because $N = 100$, this estimate is only 2 percent greater than the true value of N . Note also that the lower limit of the large sample 95 percent confidence interval extends below the number of different marbles seen. This undesirable operating characteristic is revealed throughout the results of this study, and is discussed in Appendix O. When this happens, the

NUMBER OF TRAPPING OCCASIONS WAS	5
NUMBER OF ANIMALS CAPTURED, $M(t+1)$, WAS	106
TOTAL NUMBER OF CAPTURES, N , WAS	149
ESTIMATED PROBABILITY OF CAPTURE, P -HAT =	.1726
POPULATION ESTIMATE IS	173 WITH STANDARD ERROR 17.9418
APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL	137 TO 209

FIG. 2. Example of population estimation with constant probability of capture under Model M_0 with meadow vole data from E. Larsen (pers. comm.).

lower limit should be increased to the number of distinct individuals seen. Finally, we mention that one would not expect the estimator \hat{N}_0 to be robust to departures from the assumptions of the urn experiment. For instance, if black (marked) marbles were larger than white (unmarked) ones, and thus had a higher probability of selection, we could expect \hat{N}_0 to exhibit significant negative bias.

Example

E. Larsen (pers. comm.) used live-trapping to estimate the population size of meadow voles *Microtus ochrogaster* on a grid near the Flint Hills of Kansas in June 1974. A 10×10 grid of live traps, spaced 40 feet (12.2 m) apart, was laid out in a tall-grass prairie that had been unburned and ungrazed for 3 years. On the first 2 nights of trapping, traps were placed on top of the deep, dense litter that uniformly covered the substrate, and as a result almost no animals were captured. On the third night, holes were dug in the litter and the traps were placed in the holes. That trapping occasion yielded only 12 animals captured, perhaps due to the adverse effect that disturbance of the environment may have had on the animals. On the last 5 nights of trapping, however, relatively large numbers of animals were captured consistently. Thus, we have chosen to analyze the data from only those occasions. When applied to those data, the discrimination procedure described in the section entitled TESTS OF MODEL ASSUMPTIONS chose Model M_0 as the appropriate model for the data.

The estimation procedure associated with this model produced the results presented in Fig. 2. Notice that the calculated 95 percent confidence interval is relatively narrow, probably due to the fact that the estimate of capture probability p is nearly 0.2.

Discussion

Model M_0 represents what might be called the "best" of all possible experimental situations considered here in that a minimum number of "nuisance" parameters is involved (one) if one is concerned only with estimation of population size N . This lack of nuisance parameters results of course from the restrictive assumptions on which the model is based. We believe that those assumptions are in most cases unrealistic, and, therefore, the estimator based on the model is, in general, of limited use. The case against the model is strengthened by the fact that its associated estimator \hat{N}_0 appears extremely nonrobust to variation in capture probabilities caused by behavioral response or heterogeneity. Moreover, it appears true in general that little is gained by using Model M_0 instead of Model M_t when only time specific changes in probabilities are present. Therefore, the greatest utility of Model M_0 lies in providing a "null" model useful in testing for sources of variation, and in providing a basic model that can be generalized in a number of different ways. Such generalizations are the subject of concern in the following 7 sections.

MODEL M_t : CAPTURE PROBABILITIES VARY WITH TIME

Structure and Use of the Model

Assumptions and Parameters

The set of assumptions used as a basis for Model M_t is the same set associated with the classical multiple capture-recapture experiment. It is assumed that all members of the population are equally at

risk to capture on the j^{th} trapping occasion. Thus, all animals have the same probability of capture on any particular trapping occasion, but that probability can change from one occasion to the next. The set of parameters involved in this model contains N , the population size, and the p_j , $j = 1, \dots, t$, where p_j is the probability of capture on the j^{th} occasion.

Statistical Treatment

Model M_t has received more statistical attention than any other in the literature (see Cormack 1968). Schnabel (1938) first used the above set of assumptions to develop a model from which the well-known Schnabel estimator was derived. Her model, however, assumed that the values of the M_j , the number of marked animals in the population at time j , are known a priori, for $j = 1, \dots, t$. It remained for Darroch (1958) to derive the correct model for the situation. Using his results, we may write the probability distribution of the set of possible capture histories $\{X_\omega\}$ as:

$$P\{\{X_\omega\}\} = \frac{N!}{\left[\prod_{\omega} X_\omega! \right] (N - M_{t+1})!} \cdot \prod_{j=1}^t p_j^{n_j} (1 - p_j)^{N - n_j},$$

where

n_j = number of animals caught on the j^{th} occasion, and

M_{t+1} = number of different animals captured in the experiment.

When $t = 2$, a closed form expression for the maximum likelihood estimator of N exists and is given by $\hat{N}_t = n_1 n_2 / m_2$, where m_2 is the number of recaptures in the second sample. This is the familiar Lincoln Index. Darroch (1958) derived an expression that may be solved iteratively to give an estimator of population size for $t > 2$. One is led to believe that this estimator produces estimates within unity of the true ML estimate of N , but this is not in fact the case. Details of the

algorithm necessary to produce the true ML estimate for a given set of data are given in Appendix C. The estimate will depend only on the value of the minimal sufficient statistic for the model, namely $\{n_1, n_2, \dots, n_t, M_{t+1}\}$. Note that the model involves $t + 1$ parameters and that the dimension of the MSS is also $t + 1$. This assures identifiability of all the parameters of the model. Darroch (1958) also developed an estimator of the asymptotic variance of the ML estimator that can be used in the construction of a confidence interval for N (see Appendix C).

Simulation Results

Bias

If the experimental situation is well represented by Model M_t , it is important to know what biases may be expected from the estimator of N discussed in the preceding section. Computer simulations of experiments on populations satisfying the assumptions of Model M_t produce the same general conclusions concerning the bias of \hat{N}_t as those produced in Model M_0 . That is, if the probabilities of capture p_j are, on the average, close to 0.1 or larger, the bias of \hat{N}_t is not significant. Again, however, if the p_j 's become smaller than 0.1, significant bias results. Some examples are given in Table 1 for experiments conducted for 5 and 7 trapping occasions (additional results are presented in Table N.2.b of Appendix N). In Table 1, as in all tables in this publication $\text{Ave}[\cdot]$ represents the average value of the quantity in brackets over all simulated replications. R represents the number of replications performed and RB represents percent relative bias.

Confidence Intervals

Confidence intervals were constructed from simulated experiments to compare achieved confidence coefficients to the stated value of 0.95. Those achieved levels depend on a number of factors of which the most important are the accuracy of the variance estimator of \hat{N}_t , the

TABLE 1.—RESULTS OF COMPUTER SIMULATION STUDIES OF THE BIAS OF \hat{N}_t (ALSO SEE APPENDIX TABLE N.2.b)

Percent relative bias RB	Average of estimates Ave[\hat{N}_t]	True population N	Model M_t Probabilities of capture $p_j, j = 1, 2, \dots, t$	Number of reps R	Number of occasions t	Trial
22.0	1,015.6	800	0.01, 0.01, 0.02, 0.03, 0.03	200	5	11
10.7	442.7	400	0.03, 0.04, 0.05, 0.06, 0.07	200	5	5
2.2	408.6	400	0.10, 0.10, 0.10, 0.10, 0.01	200	5	3
-0.7	198.7	200	0.30, 0.40, 0.10, 0.40, 0.30	100	5	11
-0.4	398.6	400	0.50, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10	100	7	7
-0.2	399.4	400	0.20, 0.40, 0.30, 0.10, 0.20, 0.30, 0.20	100	7	9

degree to which \hat{N}_t is normally distributed, and the accuracy of \hat{N}_t itself. Results indicated that the achieved confidence level was in most cases at least 0.90. Those results are encouraging because in many of the experimental situations investigated the assumptions necessary for the strict validity of the confidence interval used are not met (cf. Appendix O for a discussion of the assumptions). However, the amount and the kind of information contained in such an interval varies from one situation to the next depending on the values of the capture probabilities $p_j, j = 1, 2, \dots, t$. If those probabilities are on the average as large as 0.20 or 0.30, confidence interval coverage is good and interval width is small. Thus, very useful and reliable information concerning N can be obtained under such experimental conditions. As the capture probabilities p_j decrease, however, so does the information about population size provided by the confidence interval. That is, interval width tends to be so large as to not provide any useful information with respect to N. This result agrees with Chapman's (1951) argument, for the case $t = 2$, that variance

estimates tend to be prohibitively large when probabilities are small. In cases where capture probabilities are extremely low, lower confidence interval limits are negative, thus providing no information about N. As pointed out in the discussion of Model M_0 , however, such experimental results do inform the researcher that his data have been able to tell him essentially nothing about population size. In that respect, therefore, the calculated confidence interval is providing relevant information, not about population size but about the failure of the experiment itself.

To illustrate these points, Table 2 has been constructed, giving selected simulation results taken from Table N.2.b of Appendix N regarding confidence interval coverage (the proportion of replications R in which the confidence interval contained N), and average confidence interval width (Ave [C.I. width]).

Robustness

Because the maximum likelihood estimator (or approximations thereof) of population size N has been so frequently

TABLE 2.—SIMULATED CONFIDENCE INTERVAL WIDTHS AND COVERAGE USING \hat{N}_t (ALSO SEE APPENDIX TABLE N.2.b)

Average [C.I. width]	Coverage	Population size N	Model M_t Probabilities of capture					Number of reps R	Number of occasions t	Trial
			p_1	p_2	p_3	p_4	p_5			
1,348.3	0.89	400	0.01	0.01	0.02	0.03	0.03	1,000	5	12
140.2	0.94	100	0.05	0.05	0.10	0.15	0.15	1,000	5	
31.3	0.88	200	0.40	0.30	0.10	0.30	0.40	100	5	10
15.1	0.96	400	0.45	0.45	0.50	0.55	0.55	200	5	1

used in practice over the last 40 years, questions of robustness of those estimators with respect to departures from the assumptions of Model M_t are particularly relevant. Simulation results indicate that \hat{N}_t under Model M_t is nonrobust to failure of the assumption that all animals, regardless of capture history, are equally catchable on the j^{th} trapping occasion. As previously emphasized, departures from that assumption may be due to capture probabilities varying with the animal, or the animal's probability of capture being altered after first capture, or both. Although both departures may create significant bias in \hat{N}_t , the nature of the bias differs. That is, departure from equal catchability caused by heterogeneity results in \hat{N}_t being negatively biased, with the magnitude of the bias proportional to the amount of heterogeneity present in the population. This observation corresponds with the results of the study by Edwards and Eberhardt (1967) on a rabbit population of known size. Bias of \hat{N}_t that is a manifestation of animals exhibiting a behavioral response to capture, however, may be either positive or negative, according to whether the animals become trap shy or trap addicted, respectively. This is the same result noted for the estimator of N associated with Model M_o . The magnitude of the bias depends on the degree to which the animal's behavior is changed. Table 3 illustrates the bias of \hat{N}_t when it is used in simulated experiments conducted on populations satisfying the assumptions of Model M_b , in which animals exhibit be-

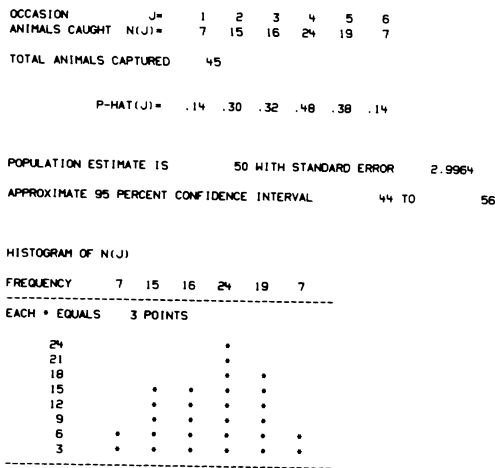


FIG. 3. Example of population estimation with time specific changes in probability of capture under Model M_t with least chipmunk data from V. Reid (pers. comm.).

havioral response to first capture, and Model M_h , in which heterogeneity of capture probabilities occurs. In Table 3, 2 examples of Model M_b are considered: Trial 8, where animals that have not been captured previously have a probability of capture of 0.2 on every trapping occasion and a probability of capture 0.05 after first capture, and Trial 9, where the probability of capture for previously captured animals changes from 0.2 to 0.5. Three examples of Model M_h are considered; for example Trial 1, in which 200 animals have a 0.05 probability of capture, 100 animals have a 0.15 probability of capture, and 100 animals have a 0.50 probability of capture. When the estimator \hat{N}_t was computed from the data generated

TABLE 3.—COMPUTER SIMULATION RESULTS ILLUSTRATING THE NONROBUSTNESS OF THE DARROCH ESTIMATOR \hat{N}_t USING DATA GENERATED UNDER OTHER MODELS. DATA MODELS ARE DESCRIBED IN TABLES N.3.a AND N.4.a OF APPENDIX N (ALSO SEE APPENDIX TABLES N.3.b AND N.4.b)

Percent relative bias RB	Ave[\hat{N}_j]	Population size N	Number of reps R	Number of occasions t	Data model
199.3	299.3	100	100	5	M_b , Trial 8 ¹
-28.3	71.8	100	100	5	M_b , Trial 9
-43.0	228.0	400	100	5	M_b , Trial 1
-13.2	173.6	200	100	10	M_b , Trial 6
-12.1	87.9	100	100	5	M_h , Trial 12

¹ For example, data generated under Model M_b were used to estimate N using estimator \hat{N}_t .

under this trial of Model M_b , an average percent bias of -43.0 resulted (Table 3).

Example

In the summer of 1975, V. Reid (pers. comm.) laid out a 9×11 livetrapping grid with traps spaced 50 feet (15.2 m) apart at a Colorado location in a bottom area dominated by sagebrush and snowberry and peripherally by gambel oak, serviceberry, and juniper. Least chipmunk *Eutamias minimus* were trapped for 6 consecutive days ($t = 6$). The discrimination procedure described in the TESTS OF MODEL ASSUMPTIONS section chose Model M_t as the most appropriate model for the data. Although that choice is subject to some suspicion (the goodness of fit test of Model M_t could not be performed because of insufficient data), the data were analyzed using the estimation procedure associated with Model M_t for purposes of illustration. Results (Fig. 3) indicate that estimates of the p_j are large enough for one to expect valid and useful confidence intervals, and such an interval does appear to result. The point estimate of 50 animals also seems to be in line with the observed data, i.e., the fact that u_6 was 0 (no new animals caught on day 6) leads us to believe that the value of \hat{N}_t should not be much larger than the number of different animals captured. A histogram of the n_j values is included to help the researcher to visually examine the data.

Discussion

Given the simulation results of this section, one might conjecture that frequent use of the "Schnabel method," i.e., Model M_b , in practice has been unfortunate unless tests of the assumptions of Model M_t have indicated the model may be appropriate. Moreover, if Model M_t appears to be an adequate representation of the experimental situation, relatively large values (at least 0.2 on the average) of the parameters p_j usually are necessary to produce useful information with re-

spect to population size. As Cormack (1968) noted, "the higher the proportion of the population marked the more statistically precise will be the estimate of population size." If the capture probabilities are small, the variance estimate of \hat{N}_t , and hence the confidence interval for N , tends to be quite large, telling the experimenter that the estimate of population size N is unreliable. Because of the apparent positive correlation between average capture probability and the amount of useful information about N contained in the experiment, it is wise for the experimenter to calculate the maximum likelihood estimates of the p_j , $j = 1, 2, \dots, t$, from the data at hand. (Formulas for these ML estimators are given in Appendix C.) Such calculations should provide more feel for both the quantity and quality of information contained in the data.

MODEL M_b : CAPTURE PROBABILITIES VARY BY BEHAVIORAL RESPONSE TO CAPTURE

Structure and Use of the Model

Assumptions and Parameters

This model deals with the failure of the assumption that initial capture does not affect the probability of capture on subsequent occasions. That is, the model allows an animal to exhibit a behavioral response to capture and become either "trap addicted" or "trap shy." Overton and Davis (1969) pointed out that "it is well known that so-called trap-happy animals are often encountered." Moreover, the existence of trap response has been well documented (Geis 1955a, 1955 unpublished doctoral dissertation, Michigan State University, East Lansing, Michigan; Tanaka 1956, 1963; Flyger 1959; Bailey 1968; Pucek 1969). Formally, Model M_b assumes that on any given trapping occasion, all unmarked animals have one probability of capture, and all marked animals have another probability of capture. It is assumed that there is no

difference between trapping occasions, i.e., that capture probabilities do not vary with time. An implication of those assumptions is that all members of the population have the same probability of capture at the beginning of the experiment. Note also that the assumption is made that an animal's capture probability is altered only once, after first capture. Although one might think it more realistic to allow the capture probability to be changed more than once (e.g., after both first and second capture) this more general assumption turns out to have no effect on the estimation of population size N . Therefore, for simplicity of presentation the assumption is made that all marked animals, regardless of the number of times they might have been captured, have the same capture probability.

The assumptions of Model M_b result in 3 model parameters: N , population size; p , the probability of capture of an unmarked animal on any trapping occasion; and c , the probability that an animal is captured on any trapping occasion subsequent to the occasion on which it was first captured.

Statistical Treatment

The probability distribution of the set of possible capture histories $\{X_\omega\}$ from a multiple capture-recapture experiment on a population satisfying the assumptions of Model M_b may be written as (Pollock unpublished dissertation):

$$P\{\{X_\omega\}\} = \frac{N!}{\prod_{\omega} X_\omega!(N - M_{t+1})!} \cdot p^{M_{t+1}}(1 - p)^{N - M_{t+1} - M} \cdot c^m(1 - c)^{M - m},$$

where $M = \sum_{j=1}^t M_j =$ sum (over j) of the

number of marked animals in the population at the time of the j^{th} trapping occasion,

$m = \sum_{j=1}^t m_j =$ total number of marked animals, caught, and

$M_{t+1} =$ number of different animals caught during the entire experiment.

A MSS for this probability distribution is given by $\{M_{t+1}, m., M.\}$. Notice that this is a 3-dimensional vector of statistics, and because the model involves 3 parameters, each of those parameters is identifiable. What is most noteworthy concerning the above distribution is the fact that the estimation of c is independent of the estimation of the parameters N and p . That is, in Model M_b , once an animal has been captured, subsequent recaptures of the animal provide no information with respect to the estimation of N and p . The recapture information of the experiment is used only in the estimation of the nuisance parameter c , the probability of recapture. The reader is referred to Appendix D for details of the above argument and for the derivation of the maximum likelihood estimators of N and p . Those estimators are essentially equivalent to those given by Zippin (1956, 1958), who provided a basic statistical analysis of the removal experiment first proposed by Moran (1951). In Zippin's discussion, however, animals usually are physically removed by killtrapping or electrofishing, whereas in Model M_b , animals are "removed" from the population by being marked; in both, estimators of N and p depend only upon first captures.

Since estimation of population size under the conditions of Model M_b is equivalent to estimation in Zippin's removal model, the "failure criterion" associated with the removal method applies. Seber and Whale (1970) showed that valid ML estimators for N and p are obtained from the data when the criterion

$$\sum_{j=1}^t (t + 1 - 2j)(n_j - m_j) > 0$$

is satisfied. If that condition, which tests whether the population is being sufficiently "depleted" by the "removal" of new animals, is satisfied, a confidence interval for N may be constructed using

TABLE 4.—RESULTS OF COMPUTER SIMULATION STUDIES OF THE BIAS OF \hat{N}_b (ALSO SEE APPENDIX N, TABLE N.3.b)

Percent relative bias RB	Average of estimates Ave[\hat{N}_b]	True population N	Probability of capture p	Number of reps R	Number of occasions t	Trial
-18.0	82.0	100	0.10	161	5	10
15.4	461.5	400	0.10	196	5	4
9.8	109.8	100	0.20	199	5	8
-0.7	198.6	200	0.25	100	7	6
-1.1	197.9	200	0.30	200	5	5

the asymptotic variance estimator of \hat{N}_b given by Zippin (Appendix D).

Simulation Results

Bias

Because behavioral response to capture is common in practice, an estimator of population size that exhibits relatively small biases in populations that satisfy the assumptions of Model M_b would prove useful in practice. Simulation results of experiments conducted on such populations indicate that \hat{N}_b , the ML estimator of N, is a good estimator if the probability of capture for an unmarked animal (p) is at least 0.2. Frequency of "failure" of the experiment is substantial (20%) with p values ≤ 0.1 , and when the experiment does succeed in these cases, \hat{N}_b tends to be significantly biased. A few examples taken from Table N.3.b of Appendix N of the simulated bias of \hat{N}_b are given in Table 4 which illustrates that biases on the order of 15–20 percent occur when p = 0.10, but that such bias gradually decreases as probability of first capture increases to reasonable levels greater than 0.20. Although the first 2 simulations included in Table 4 have the same probability of first capture, the case for N = 100 is negatively biased, whereas the case for N = 400 is positively biased. These seemingly contradictory results are caused by the estimator failing more often for the N = 100 case. When the estimator is close to failing (i.e., the failure criterion is close to zero), the estimates tend to be biased high. Because actual failure of the method occurs much more

frequently in cases where N is small, a relatively large number of potentially large estimates of N are "lost." This phenomenon therefore effects a significant reduction in the value of Ave[\hat{N}].

Confidence Intervals

Confidence intervals were simulated for various populations satisfying the assumptions of Model M_b (Table 5). A high correlation between the values of p and adequate performance of confidence intervals is indicated by Table 5. Evidence indicates that high confidence levels and small (hence informative) interval widths can be expected from experiments in which the probability of first capture is at least 0.30. On the other hand, in an experiment with N = 100 and p = 0.10 lower limits of confidence intervals were negative in more than 40 percent of those simulated experiments that succeeded. Such results illustrate the point that, in general, confidence intervals constructed from experiments with an insufficiently large value of p serve only to inform the researcher that his data cannot provide any real information with respect to population size.

Robustness

If one examines the failure criterion for estimation in Model M_b , it becomes clear that the number of new animals captured (removed) should decrease for each successive trapping occasion. Simulation results seem to indicate that steep declines in new captures over time produce good estimates of and informative confidence

TABLE 5.—SIMULATED CONFIDENCE INTERVAL WIDTHS AND COVERAGE USING \hat{N}_b (ALSO SEE APPENDIX N, TABLE N.3.b)

Average [C.I. width]	Coverage	Population size N	Probability of capture p	Number of reps R	Number of occasions t	Trial
253.9	0.71	100	0.10	161	5	10
145.0	0.95	200	0.20	98	5	
51.5	0.90	200	0.30	100	5	5
78.0	0.92	400	0.30	100	5	1

intervals for the parameter N. (This relates directly to the previous observation that larger values of the parameter p produce estimators with better properties.) One can see that if capture probabilities vary from one trapping occasion to the next, the vector of “removals” represented by (u_1, u_2, \dots, u_t) may be perturbed in such a way as to prevent the desired monotone decrease in the removals over time. This results in \hat{N}_b exhibiting a large bias and confidence intervals for N having large expected width. These observations support the conjecture that estimation of N based on Model M_b will be sensitive to significant changes in capture probabilities over time. Moreover, if capture probability varies among individuals, independently of the animal’s capture history, \hat{N}_b tends to underestimate N. The magnitude of the bias depends directly on the number of animals in the population that are essentially untrappable, i.e., those that have small (<0.1) “innate” probabilities of capture. This non-robustness to heterogeneity of capture probability is of the same nature as that exhibited by \hat{N}_o and \hat{N}_t . These remarks are based on results obtained from simulation experiments on populations that satisfy the assumptions of Model M_h (see following section for a complete description of this model). A few examples follow concerning the bias of \hat{N}_b in experiments on Model M_h populations. In a population of size 100, with 40 animals that have a 0.05 probability of capture, 40 animals with a 0.10 probability, and 20 animals with a 0.30 probability (Trial 8), an experiment with $t = 10$ trapping occasions was simulated resulting in an average \hat{N}_b estimate of 78 for N. In a pop-

ulation of size 400, with 200 animals that have a 0.05 probability of capture, 100 with a 0.15 probability, and 100 with a 0.50 probability (Trial 1), an experiment with $t = 5$ trapping occasions produced an average estimate of 258 animals.

Example

E. Larsen (pers. comm.) reported the results of a livetrapping experiment conducted in the summer of 1976 as part of a study on community succession. A total of 55 live traps, spaced 15 m apart and covering 0.81 ha, were set out in a field in which vegetation was dominated by a mixture of sagebrush *Artemisia tridentata* and rabbitbrush *Chrysothamnus* spp. Trapping was conducted for 10 consecutive nights. Data collected on the deer mouse *Peromyscus maniculatus* from those 10 occasions (Fig. 4) were analyzed by the discrimination procedure described in the section entitled TESTS OF MODEL ASSUMPTIONS and it was determined that Model M_b would be an appropriate model for the data (the significance level of a goodness of fit test of Model M_b was approximately 0.47). Therefore, the estimation procedure associated with Model M_b was used to produce point and interval estimates for N. Two aspects of those estimates (Fig. 4) are worth noting: animals tend to become trap happy since recapture probability (\hat{c}) is more than twice the value of initial capture probability (\hat{p}) and the small value of $\hat{p} = 0.09$ is the major reason why the 95 percent confidence interval for N is wide and the lower limit of the computed interval extends far below the number of animals actually seen (as mentioned be-

OCCASION	J=	1	2	3	4	5	6	7	8	9	10	
TOTAL CAUGHT	M(J)=	0	14	19	30	37	42	48	51	57	61	69
NEWLY CAUGHT	U(J)=	14	5	11	7	5	6	3	6	4	8	

ESTIMATED PROBABILITY OF CAPTURE, $P\text{-HAT} = .091026$
 ESTIMATED PROBABILITY OF RECAPTURE, $C\text{-HAT} = .233983$

POPULATION ESTIMATE IS 112 WITH STANDARD ERROR 31.1966
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVALS 50 TO 174

HISTOGRAM OF U(J)

FREQUENCY	14	5	11	7	5	6	3	6	4	8

EACH * EQUALS	2 POINTS									
14	*									
12	*	*								
10	*	*	*							
8	*	*	*	*						*
6	*	*	*	*	*				*	*
4	*	*	*	*	*	*		*	*	*
2	*	*	*	*	*	*	*	*	*	*

FIG. 4. Example of population estimation with constant probability removal estimator under Model M_b with deer mouse data from E. Larsen (pers. comm.).

fore, the lower limit should be taken as 69, not 50, in this situation).

Example

V. Reid (pers. comm.) reported the results of livetrapping deer mice *Peromyscus maniculatus* in a drainage bottom of sagebrush, gambel oak, and serviceberry with pinyon pine and juniper on the uplands. The area, in Rio Blanco County, Colorado, was trapped for 6 consecutive nights in the summer of 1975. Traps were arranged in a 9×11 grid and spaced 50 feet (15.2 m) apart. Analysis of the resulting data by the model selection procedure indicated that Model M_b would be appropriate for use in estimating N (e.g., the significance level of a goodness of fit test of Model M_b was approximately 0.43). The data and the results of using \hat{N}_b to produce point and interval estimates for N are given in Fig. 5. Note that the mice tend to become trap happy, as in the previous example. In that study, however, the estimate of first capture probability is $\hat{p} = 0.34$, a fact that largely accounts for the narrow width of the 95 percent confidence interval. Again, the lower limit of the confidence interval could be taken as 38, the number of different animals captured.

OCCASION	J=	1	2	3	4	5	6	
TOTAL CAUGHT	M(J)=	0	15	23	29	32	35	38
NEWLY CAUGHT	U(J)=	15	8	6	3	3	3	

ESTIMATED PROBABILITY OF CAPTURE, $P\text{-HAT} = .342413$
 ESTIMATED PROBABILITY OF RECAPTURE, $C\text{-HAT} = .611940$

POPULATION ESTIMATE IS 41 WITH STANDARD ERROR 3.0518
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVALS 35 TO 47

HISTOGRAM OF U(J)

FREQUENCY	15	8	6	3	3	3

EACH * EQUALS	2 POINTS					
16	*					
14	*					
12	*	*				
10	*	*	*			
8	*	*	*	*		
6	*	*	*	*	*	
4	*	*	*	*	*	*
2	*	*	*	*	*	*

FIG. 5. Example of population estimation with constant probability removal estimator under Model M_b with deer mouse data from V. Reid (pers. comm.).

Discussion

Simulation results concerning the performance of \hat{N}_b in populations that satisfy the assumptions of Model M_b seem encouraging in that the estimator and its associated confidence intervals provide useful information in the presence of behavioral response to first capture, if the probability of first capture is sufficiently large (>0.1). (Hence, calculation of the ML estimate of p should assist in assessing the amount of information concerning N that is contained in the data). However, the estimator of N appears nonrobust to other factors that may have an effect on an animal's probability of capture. Moreover, the estimation procedure associated with Model M_b is a special case of a more general estimation procedure which involves fewer assumptions than those associated with Model M_b . This procedure is described in the section on removal models. For these reasons, the estimator of population size N associated with Model M_b is useful only in those instances where all unmarked animals have the same capture probability on all trapping occasions.

MODEL M_h : CAPTURE PROBABILITIES
VARY BY INDIVIDUAL ANIMAL

Structure and Use of the Model

Assumptions and Parameters

The assertion that each member of the population has its own probability of capture independent of all other members of the population is the basis of Model M_h . The assumptions are made that there is no difference between trapping occasions and no behavioral response to capture, but that there is heterogeneity among the capture probabilities of individuals. Cormack (1968) stated that a test for the assumption of "heterogeneity of individuals" is impossible unless an independent experiment is conducted on a population of known size that is "representative" of the population of interest. Because of that apparent difficulty (and others), models similar to Model M_h have been largely ignored, although the assumptions behind the model fit more experimental situations than do most of those commonly used. As Eberhardt (1969a) pointed out, "various sets of data indicate . . . that the equal-probability-of-capture assumption is not fulfilled."

Conceptually, Model M_h involves $N + 1$ parameters: the population size N and the set of capture probabilities $\{p_i\}$, $i = 1, 2, \dots, N$, where p_i is the probability of capture of the i^{th} animal on any trapping occasion. This more general formulation of Model M_h does not allow estimation of population size N because of the presence of too large a number of nuisance parameters. Therefore, for most of the development in this chapter, it is more useful to think of $\{p_i\}$ as a random sample of size N from some probability distribution $F(p)$ defined on the interval $[0, 1]$.

Unfortunately, there is no completely satisfactory estimator when Model M_h is true. Therefore, one should design the study to minimize heterogeneity. In fact, it is worth repeating that it is very important to design any capture study carefully with the goal of having as few factors as possible affecting capture probabilities (see STUDY DESIGN).

Statistical Treatment

Recall that the random variable X_{ij} takes on the value 1 when the i^{th} animal is caught on the j^{th} trapping occasion and is zero otherwise. If it is assumed that $\{p_i\}$ results from a random sample from $F(p)$ then the probability distribution for the set of variables $\{X_{ij}\}$ reveals that a sufficient statistic for Model M_h is given by the capture frequencies $\{f_1, f_2, \dots, f_j\}$, where f_j = the number of animals caught exactly j times in the experiment. Hence, all the information for estimating N is contained in the frequency of capture statistics. Estimators based on frequency of capture statistics have been proposed (Craig 1953, Tanton 1965, Eberhardt 1969a), but are the result of essentially ad hoc approaches. If the particular family of distributions of which $F(p)$ is a member is specified, a maximum likelihood estimator for N may be derived. For example, Burnham (unpublished dissertation) assumed that $\{p_i\}$ results from a random sample from a 2-parameter beta distribution and investigated the properties of the resulting ML estimator of N . Theoretical and simulation results indicated that this estimator has quite unsatisfactory operating characteristics, and this inspired Burnham to develop a nonparametric estimation procedure that would not require specification of $F(p)$. Such an estimator was developed using an extension of the jackknife method of bias reduction first proposed by Quenouille (1949, 1956). The resulting form of the estimator \hat{N}_h can be written as

$$\hat{N}_h = \sum_{j=1}^t a_j f_j.$$

Each set of constants a_1, a_2, \dots, a_t , generated by using a different "order" of the jackknife procedure, corresponds to a different "jackknife" estimator. Burnham (unpublished dissertation) suggested a sequence of statistical tests designed to choose the best one of the estimators for any given data set. For a more detailed presentation of that estimation procedure, consult Appendix E which also pre-

TABLE 6.—RESULTS OF COMPUTER SIMULATION STUDIES OF THE BIAS OF \hat{N}_h (ALSO SEE APPENDIX, TABLE N.4.b)

Percent relative bias RB	Average of estimates Ave $\{N_h\}$	True population N	Probabilities of capture $p_i, i = 1, 2, \dots, N$	Number of reps R	Number of occasions t	Trial
-17.2	331.1	400	$p_i = 0.05, i = 1, 200; p_i = 0.15, i = 201, 300; p_i = 0.50, i = 301, 400.$	200	5	1
0.2	100.2	100	$p_i = 0.05, i = 1, 40; p_i = 0.10, i = 41, 80; p_i = 0.30, i = 81, 100.$	200	10	8
15.0	460.1	400	$p_i = 0.10, i = 1, 100; p_i = 0.20, i = 101, 200; p_i = 0.25, i = 201, 300; p_i = 0.30, i = 301, 400.$	200	5	3
4.3	417.1	400	$p_i = 0.01, i = 1, 50; p_i = 0.15, i = 51, 200; p_i = 0.25, i = 201, 300; p_i = 0.30, i = 301, 400.$	100	5	4
3.7	207.0	200	$p_i = 0.05, i = 1, 50; p_i = 0.15, i = 51, 150; p_i = 0.25, i = 151, 200.$	100	10	6
11.0	443.9	400	$p_i = 0.20, i = 1, 100; p_i = 0.30, i = 101, 200; p_i = 0.40, i = 201, 300; p_i = 0.50, i = 301, 400.$	100	5	5
13.4	226.8	200	$p_i = 0.15, i = 1, 50; p_i = 0.20, i = 51, 100; p_i = 0.25, i = 101, 150; p_i = 0.30, i = 151, 200.$	100	7	7

sents a formula for a variance estimator of \hat{N}_h .

Simulation Results

Bias

Cormack (1968:497) noted that when animals exhibit heterogeneous capture probabilities "any available estimate of population size [is] markedly biased." We performed simulation experiments with Model M_h populations to determine whether, in a given population, "marked bias" is associated with \hat{N}_h . Table 6 lists some examples of such results. The complete details of the simulation are given in Tables N.4.a and N.4.b of Appendix N. Results of those simulations and of those conducted by Burnham (unpublished dissertation) seem to indicate 2 major points concerning the bias of the jackknife estimator: (1) general statements concerning the magnitude and direction of the bias are not easily made since the estimator exhibits both small and large and positive and negative bias, depending on the values of N and the set $\{p_i\}$. One may conjecture, however, that if the number of trapping occasions is sufficiently large (say greater than 5) and if a negligible number of animals is for all practical purposes untrappable, then the bias of \hat{N}_h will be within a tolerable

range; and (2) even though significant bias may sometimes be present in the jackknife estimator, the magnitude of such bias will, in general, be substantially less than the bias that would be associated with any other estimator thus far discussed. Evidence of this assertion can be seen in Tables N.1.b, N.2.b, and N.3.b of Appendix N that present the results of using estimators other than \hat{N}_h on data simulated from Model M_h .

Confidence Intervals

Construction of confidence intervals from simulated experimental data generated from populations of Model M_h show a large variance in the achieved confidence coefficients (Table 7). Coverage is poor and ranges from an estimated 0 percent (Trials 2, 11) to 87 percent (Trial 6) (Appendix N, Table N.4.b). Because results indicate that the distribution of the jackknife estimator is approximately normal, poor coverage is caused either by significant negative bias of the variance estimator, a significant bias of the estimator \hat{N}_h , or both. Unless an experimenter has data from an adequate number of trapping occasions and feels that very few, if any, members of the population under study are untrappable, the confidence interval constructed by the

TABLE 7.—SIMULATED CONFIDENCE INTERVAL WIDTHS AND COVERAGE USING THE JACKKNIFE ESTIMATOR \hat{N}_h (ALSO SEE APPENDIX N, TABLE N.4.b)

Average [C.I. width]	Coverage	Population size N	Probabilities of capture $p_i, i = 1, 2, \dots, N$	Number of reps R	Number of occasions t	Trial
89.9	0.18	400	$p_i = 0.05, i = 1, 200; p_i = 0.15, i = 201, 300; p_i = 0.50, i = 301, 400.$	200	5	1
104.9	0.40	400	$p_i = 0.10, i = 1, 100; p_i = 0.20, i = 101, 200; p_i = 0.25, i = 201, 300; p_i = 0.30, i = 301, 400.$	200	5	3
45.4	0.87	200	$p_i = 0.05, i = 1, 50; p_i = 0.15, i = 51, 150; p_i = 0.25, i = 151, 200.$	100	10	6
45.3	0.68	100	$p_i = 0.05, i = 1, 40; p_i = 0.10, i = 41, 80; p_i = 0.30, i = 81, 100.$	200	5	8

methods associated with Model M_h should be considered unreliable. The reader may refer to Table N.4.b of Appendix N for complete results concerning simulated coverage of these confidence intervals.

Robustness

The jackknife estimator was constructed with the objective of having an estimator that is robust to heterogeneity of capture probabilities. The degree to which the objective is satisfied has been discussed previously. The question remains: is the jackknife estimator robust to other forms of departure from the assumption of equal catchability? Results indicate that of the estimators discussed in this monograph the jackknife is the most robust. This does not mean that it is a good estimator no matter what assumptions apply to the population under study. In general, it may be said that the

jackknife estimator will provide an adequate estimate of population size in an experiment in which many animals are caught a relatively large number of times. Some examples of the bias of the jackknife estimator when data are simulated from models other than M_h are given in Table 8. More information may be obtained from Tables N.2.b, N.3.b, and N.5.b of Appendix N. In the table, the information under the heading Data model refers to the population on which the simulated experiments were performed. For details of these populations, see Appendix N, Tables N.2.a, N.3.a, and N.5.a.

Example

Carothers (1973a) conducted a capture-recapture experiment on the "population" of taxicabs in Edinburgh, Scotland. The population, known to be of size 420 and assumed to be closed, was sam-

TABLE 8.—COMPUTER SIMULATION RESULTS ILLUSTRATING THE ROBUSTNESS OF THE JACKKNIFE ESTIMATOR \hat{N}_h TO DATA GENERATED UNDER OTHER MODELS (ALSO SEE APPENDIX N, TABLES N.2.b, N.3.b AND N.5.b)

Percent relative bias RB	Ave[\hat{N}_h]	Population size N	Number of reps R	Number of occasions t	Data model
-12.5	349.2	400	200	5	M_t , Trial 3 ¹
- 1.3	98.7	100	1,000	5	M_t , Trial 12
-17.6	82.3	100	100	5	M_b , Trial 9
-13.1	347.7	400	200	5	M_b , Trial 4
0.7	402.9	400	100	5	M_{th} , Trial 1
19.7	478.7	400	100	5	M_{th} , Trial 2

¹ For example, data generated under Model M_t were used to estimate N using the estimator \hat{N}_h .

NUMBER OF TRAPPING OCCASIONS HAS 10
 NUMBER OF ANIMALS CAPTURED, $M_{(T+1)}$, HAS 283
 TOTAL NUMBER OF CAPTURES, $N_{..}$, HAS 500

FREQUENCIES OF CAPTURE, $F_{(i)}$
 1 * 1 2 3 4 5 6 7 8 9 10
 $F_{(i)}$ * 142 81 49 7 3 1 0 0 0 0

COMPUTED JACKKNIFE COEFFICIENTS

	N(1)	N(2)	N(3)	N(4)	N(5)
1	1.900	2.700	3.400	4.000	4.500
2	1.000	.289	-.878	-2.278	-3.722
3	1.000	1.000	1.476	2.535	4.042
4	1.000	1.000	1.000	.743	.077
5	1.000	1.000	1.000	1.000	1.103

THE RESULTS OF THE JACKKNIFE COMPUTATIONS

	N(1)	SE(1)	.95 CONF. LIMITS	TEST OF $N_{(1+1)}$ VS. $N_{(1)}$
1	283			25.877
0	410.8	15.58	380.3	441.3
2	466.8	25.20	417.4	516.2
3	495.0	36.42	423.7	566.4
4	516.9	49.98	418.9	614.9
5	540.4	65.29	412.4	668.4

AVERAGE P-HAT = .1062

INTERPOLATED POPULATION ESTIMATE IS 471 WITH STANDARD ERROR 26.8266
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 418 TO 524

HISTOGRAM OF $F_{(i)}$
 FREQUENCY 142 81 49 7 3 1 0 0 0 0
 EACH * EQUALS 15 POINTS

135	*
120	*
105	*
90	*
75	*
60	*
45	*
30	*
15	*

FIG. 6. Example of population estimation with variable probability of capture by animal under Model M_h with the complete set of Scheme A taxicab data from Carothers (1973a).

pled for 10 consecutive days. The sampling was done by observing taxicabs that passed preselected points and recording "capture" or "recapture" according to whether or not the particular cab had been observed previously. In the author's sampling scheme "A," a different set of sampling points was selected each day, and the time of sampling (i.e., morning, afternoon, night) was also varied. We applied the model selection procedure (described in TESTS OF MODEL ASSUMPTIONS) to the data collected from this scheme. The procedure indicated that an appropriate model for the data would be Model M_h . (For example, a goodness of fit test of Model M_h has a significance level greater than 0.99). Thus, the estimation procedure associated with Model M_h was used to produce point and interval estimates for N (Fig. 6). Although \hat{N}_h overestimates the true value of N by approximately 12 percent, the confidence

NUMBER OF TRAPPING OCCASIONS HAS 6
 NUMBER OF ANIMALS CAPTURED, $M_{(T+1)}$, HAS 68
 TOTAL NUMBER OF CAPTURES, $N_{..}$, HAS 145

FREQUENCIES OF CAPTURE, $F_{(i)}$
 1 * 1 2 3 4 5 6
 $F_{(i)}$ * 25 22 13 5 1 2

COMPUTED JACKKNIFE COEFFICIENTS

	N(1)	N(2)	N(3)	N(4)	N(5)
1	1.833	2.500	3.000	3.333	3.500
2	1.000	.467	-.233	-.833	-1.167
3	1.000	1.000	1.225	1.542	1.750
4	1.000	1.000	1.000	.956	.914
5	1.000	1.000	1.000	1.000	1.001

THE RESULTS OF THE JACKKNIFE COMPUTATIONS

	N(1)	SE(1)	.95 CONF. LIMITS	TEST OF $N_{(1+1)}$ VS. $N_{(1)}$
1	68			1.410
0	88.8	6.18	76.7	100.9
1	93.8	9.40	75.4	112.2
2	93.8	12.95	69.0	118.6
3	92.0	15.45	62.5	123.1
4	92.2	17.06	58.7	125.6

AVERAGE P-HAT = .2778

INTERPOLATED POPULATION ESTIMATE IS 87 WITH STANDARD ERROR 5.8220
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 75 TO 99

HISTOGRAM OF $F_{(i)}$
 FREQUENCY 25 22 13 5 1 2
 EACH * EQUALS 3 POINTS

24	*
21	*
18	*
15	*
12	*
9	*
6	*
3	*

FIG. 7. Example of population estimation with variable probability of capture by animal under Model M_h with snowshoe hare data from Burnham and Cushwa (pers. comm.).

interval constructed for N does in fact contain the true population value of 420. These results are as satisfactory as any produced by various estimation techniques used by Carothers.

Example

In 1972, Burnham and Cushwa (pers. comm.) laid out a livetrapping grid in a black spruce forest 30 miles (48.3 km) north of Fairbanks, Alaska. The basic grid was 10×10 , with traps spaced 200 feet (61 m) apart. Trapping for snowshoe hares *Lepus americanus* was carried out for 9 consecutive days in early winter. Traps were not baited for the first 3 days, and therefore we have chosen to analyze the data from the last 6 days of trapping. The model selection procedure designed to identify a proper model for estimating N chose Model M_h as the most appropriate model for the data. A goodness of fit

test of Model M_h had a significance level of 0.06. The data, and the estimates produced by the estimation procedure of Model M_h are given in Fig. 7. The results of the entire 9 days of trapping show 74 different animals caught while the lower 95 percent confidence limit for N given in Fig. 7 is 75. Although this result is satisfying, we reemphasize that the true confidence coefficient of the confidence intervals associated with Model M_h is, for most populations, much less than the stated level of 0.95.

Discussion

Theoretically, 2 main complicating factors are associated with the existence of heterogeneity of individual capture probabilities. First, a parameter must be introduced into the model for every individual in the population, and this makes straightforward statistical estimation of population size impossible. Second, the assumption of heterogeneity can render some individuals nearly “invisible” with respect to any estimation procedure based on marking methods because such individuals have nearly zero catchability. The jackknife estimation procedure is the most robust procedure thus far proposed for dealing with problems associated with heterogeneity. Moreover, the estimator seems robust to other factors that may have an effect on capture probabilities. It should be pointed out, however, that in some cases the bias of the estimator is not negligible. In general, if the jackknife procedure is to be considered a good nonparametric and robust technique, the trapping must be performed on a large number of occasions, and the number of recaptures on each occasion must be substantial.

We remark that the observations contained in this section reflect Gilbert’s (1973:524) opinion that “the presence of heterogeneity per se is not as important as the particular pattern of heterogeneity and whether the heterogeneity is near zero or one.”

MODEL M_{tb} : CAPTURE PROBABILITIES VARY BY TIME AND BEHAVIORAL RESPONSE TO CAPTURE

Structure and Use of the Model

Assumptions and Parameters

If it is assumed that an animal’s probability of capture changes after initial capture and that temporal changes also have an effect on capture probabilities, Model M_{tb} results. This model has been conceptualized in the literature (see Eberhardt et al. 1963, Seber 1973) but it has not received mathematical treatment with respect to estimation of population parameters. Although many variations of the basic model exist (see Appendix F), for purposes of this chapter we shall define the parameters of Model M_{tb} as:

- N = population size,
- p_j = the probability an unmarked animal is captured on the j^{th} trapping occasion, $j = 1, 2, \dots, t$,
- c_j = probability that a marked (previously captured) animal is captured on the j^{th} trapping occasion, $j = 2, 3, \dots, t$.

Note that again, as in Model M_b , an animal exhibits only 1 behavioral response to capture after its initial capture.

Statistical Treatment

The joint probability distribution of the set of possible capture histories $\{X_\omega\}$ can be written as

$$P[\{X_\omega\}] = \frac{N!}{\left[\prod_{\omega} X_\omega! \right] (N - M_{t+1})!} \prod_{j=1}^t p_j^{u_j} \cdot (1 - p_j)^{N - M_{j+1}} (c_j)^{m_j} (1 - c_j)^{M_j - m_j},$$

where

- u_j = number of unmarked animals caught on the j^{th} occasion, $j = 1, 2, \dots, t$,
- M_j = number of marked animals in the population the the time of the j^{th} sample, $j = 1, 2, \dots, t$, and
- m_j = number of marked animals captured on the j^{th} occasion, $j = 1, 2, \dots, t$.

A minimal sufficient statistic for this distribution is given by $\{u_1, u_2, \dots, u_t, m_2, m_3, \dots, m_t\}$ which has dimension $2t - 1$. Since the model involves $2t$ parameters, not all parameters can be estimated and maximum likelihood estimation of N proves to be impossible. The structure of $P\{X_{\omega}\}$ reveals that estimation of the parameters N, p_1, p_2, \dots, p_t depends only upon the vector of removals $\{u_1, u_2, \dots, u_t\}$. Thus, estimation of N in Model M_{th} would be equivalent to the estimation in Zippin's (1956) removal model generalized to allow the probability of removing an animal, p_j , to be different for every trapping occasion. This involves estimating $t + 1$ parameters with t statistics. Hence, in order to make N an identifiable parameter, one has to make the assumption that at least 2 of the p_j are equal. Because there appears to be no biological justification for making such an assumption, only ad hoc estimation procedures are available. For example, Tanaka (1951, 1952) has proposed a regression technique that involves plotting $y_j = m_j/n_j$ versus M_j on a logarithmic scale; however, Seber (1973) pointed out there are serious problems of interpretation involved in using this technique. It is true that graphical techniques such as this may possess some utility in that they encourage the researcher to examine the data carefully. However, regression methods are not contained in the class of estimation methods considered in this monograph. Therefore, Tanaka's (1951, 1952) method and those similar to it will not be considered here.

Discussion

In view of the previous discussion, we believe that Model M_{th} is not useful for estimation purposes. However, the model does have utility with respect to the problem of choosing a "best" estimation procedure given the data at hand. For details of how the model is used in such a testing procedure see the section on model selection.

MODEL M_{th} : CAPTURE PROBABILITIES VARY BY TIME AND INDIVIDUAL ANIMAL

Structure and Use of the Model

Assumptions and Parameters

If, on the j^{th} trapping occasion, the i^{th} animal has a capture probability of $p_{ij} = p_i p_j$ that is independent of its capture history (i.e., there is no behavioral response to capture), then Model M_{th} is the appropriate probability model for a capture-recapture experiment on such a population. Notice that the structure of p_{ij} implies that variation in capture probabilities due to time is independent of the variation caused by individual heterogeneity. In effect, this means that the factors responsible for time variation, e.g., environmental conditions, affect all members of the population similarly. For purposes of this section it shall be assumed that the $p_i, i = 1, 2, \dots, N$, are a random sample of size N from some probability distribution-function $F(p; \theta)$ that is parameterized by the vector θ and defined for p in the interval $[0, 1]$. Note that it is necessary that each of the parameters, $p_j, j = 1, 2, \dots, t$, be subject to the constraint that $0 \leq p_i p_j \leq 1$ for $i = 1, 2, \dots, N$. Conceptually, the form of the distribution function $F(p; \theta)$ and the parameters $\theta, p_1, p_2, \dots, p_t$ are unknown elements of the model, as is the parameter N .

Statistical Treatment

Estimation of population size under the assumptions of Model M_{th} has not been considered in the literature. If one is willing to completely specify the distribution $F(p; \theta)$, then maximum likelihood estimation of N becomes possible (cf. Appendix G). In general, such an assumption is not realistic. Furthermore, simulation results of Burnham (unpublished dissertation) indicate that an ML estimator resulting from such an assumption may have poor operating characteristics.

If no assumptions concerning $F(p; \theta)$ are made, ML estimation of N is not pos-

TABLE 9.—COMPUTER SIMULATION OF THE BIAS ASSOCIATED WITH THE ESTIMATORS \hat{N}_h AND \hat{N}_t FOR DATA GENERATED UNDER MODEL M_{th} (ALSO SEE APPENDIX N, TABLE N.5.b)

Percent relative bias RB	Ave[\hat{N}_t]	Percent relative bias RB	Ave[\hat{N}_h]	Population size N	Number of reps R	Number of occasions t	Trial ¹
0.7	402.9	-24.1	303.6	400	100	5	1
19.7	478.7	- 7.5	369.9	400	100	5	2
-11.5	353.9	-22.8	309.0	400	100	5	3
-11.8	352.8	-31.8	272.8	400	100	5	4
8.9	217.7	- 6.1	187.9	200	100	7	5
-10.9	178.2	-17.9	164.2	200	100	5	6

¹ The probability of capture for data generated under Model M_{th} is complex; therefore, the specific values are given in Appendix N, Table N.5.a.

sible. At present, we are not aware of any rigorous estimation technique that is appropriate for estimation in Model M_{th} .

Simulation Results

Bias

To gain some insight into the performance of estimators associated with models closely related to Model M_{th} , we simulated capture-recapture experiments on populations that satisfy the assumptions of Model M_{th} and calculated the estimators associated with Model M_t and Model M_h from the resulting data. Examples are given in Table 9. See Tables N.5.a and N.5.b of Appendix N for a description of the populations used and for results of other simulations. A choice between the 2 estimators with respect to performance is obviously not clearcut. The magnitude of the bias of \hat{N}_h seems in general to be less than that of \hat{N}_t . However, direction of bias of \hat{N}_h is not consistent, whereas the estimator \hat{N}_t appears to consistently exhibit negative bias. Thus, if the data seem to fit Model M_{th} , computing both estimates of N should give the experimenter some sense of the size of the population. However, it should be emphasized that the model does not seem particularly useful in providing a single satisfactory estimate of the value of N.

Confidence Intervals

Coverage of confidence intervals constructed both on the basis of the esti-

mators of Model M_h and Model M_t was poor (cf. Table N.5.b, Appendix N). The lack of coverage is due, in general, to the significant bias present in the estimates. Therefore, we recommend that confidence intervals constructed from data that apparently fit Model M_{th} be considered only as a *very crude indication* of the possible values of N and not as a valid statement of inference.

Robustness

A discussion of the robustness of the estimator associated with Model M_{th} is somewhat inappropriate because no single estimator is exclusively associated with the model. The robustness of \hat{N}_h and \hat{N}_t has been discussed previously. Those discussions indicated that the former estimator is relatively robust and that the latter is not. Such assertions seem to be supported by the results of this section concerning bias.

Discussion

Model M_{th} is one of the most difficult models dealt with in this monograph. That difficulty is reflected in the fact that no estimation procedure can be proposed that is specifically suited to all assumptions of the model. Estimators associated with 2 previous models have been investigated for possible use but neither seems completely adequate. Thus, at present, satisfactory techniques for constructing point and interval estimates of N are not

available. Hopefully, further research involving Model M_{th} will be forthcoming so that the model can be dealt with in a more adequate statistical manner. Until then, one must be satisfied with obtaining only very rough estimates of population size when Model M_{th} is the appropriate representation of the experiment.

MODEL M_{bh} : CAPTURE PROBABILITIES VARY BY INDIVIDUAL ANIMAL AND BY BEHAVIORAL RESPONSE TO CAPTURE

Structure and Use of the Model

Assumptions and Parameters

Model M_{bh} assumes that every member of the population has a specific pair of capture probabilities: p_i , the probability that the i^{th} animal is caught on any trapping occasion given that it has not been previously captured; and c_i , the probability that the i^{th} animal is recaptured given that it has been captured at least once previously. Thus, the model allows both behavioral response to first capture and individual heterogeneity of capture probabilities. An important and appealing characteristic of the model is that it allows the behavioral response to capture to vary with the animal, i.e., all members of the population do not exhibit an identical response to first capture.

The most general formulation of the model involves $2N + 1$ total parameters consisting of N (population size), and 2 capture probabilities for each of the N members of the population. Obviously, no real experiment will be able to provide enough information for the explicit estimation of this number of parameters. Further assumptions are necessary before estimation of population size is possible.

Statistical Treatment

Pollock (unpublished dissertation) introduced Model M_{bh} and assumed that the pairs (p_i, c_i) are a random sample of size N from a bivariate probability distri-

bution function $G(p, c; \theta)$ that is parameterized by the vector of unknown constants θ . If one assumes that the form of the distribution $G(p, c; \theta)$ is known and if the dimension of θ is no larger than $[t(t + 1)/2] - 1$, where t is the number of trapping occasions, then theoretically maximum likelihood estimation could be used to provide an estimator of N . However, a "nonparametric" approach is also possible. We discuss this approach in the context of the assumption that the bivariate density function $G(p, c; \theta)$ can be factored as $G_1(p; \theta_1) G_2(c; \theta_2)$. That is, an individual's initial capture probability and its probability of recapture are independent. With that assumption, Pollock (unpublished dissertation) showed that the probability distribution of the set of possible capture histories $\{X_\omega\}$ can be written as

$$P[\{X_\omega\}] = \frac{N!}{u_1!u_2! \cdots u_t!(N - M_{t+1})!} \cdot \pi_1^{u_1}\pi_2^{u_2} \cdots \pi_t^{u_t} \left(1 - \sum_{j=1}^t \pi_j\right)^{N - M_{t+1}} \cdot P^*[\{X_\omega\} | u_1, u_2, \dots, u_t],$$

where

u_j = number of unmarked animals caught at time j , $j = 1, 2, \dots, t$,

$$\pi_j = E[(1 - p)^{j-1}p] = \int_0^1 (1 - p)^{j-1}p dG_1$$

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

$P^*[\{X_\omega\} | u_1, u_2, \dots, u_t]$ = a conditional probability distribution that does not depend upon the parameter N or the distribution $G_1(p; \theta_1)$.

(Refer to Appendix H for further details). Thus, the "removals" $\{u_1, u_2, \dots, u_t\}$ are the relevant statistics for the purpose of estimating N and the parameters θ_1 in $G_1(p; \theta_1)$, and these statistics have a multinomial distribution with parameters N and $\pi_1, \pi_2, \dots, \pi_t$ (Seber 1973:316 gave this model for the special case $t = 3$). We

can transform the π_j by writing $\pi_j = (1 - \bar{p}_1)(1 - \bar{p}_2) \cdots (1 - \bar{p}_{j-1})\bar{p}_j$, $j = 1, \dots, t$, where \bar{p}_j is the average conditional probability of capture on the j^{th} occasion for those animals not previously captured. The distribution of the t -dimensional vector of removals now depends upon the $t + 1$ parameters $N, \bar{p}_1, \bar{p}_2, \dots, \bar{p}_t$, and thus all the parameters of the model are not identifiable. This is similar to the situation that resulted in Model M_{tb} where it was concluded that ML estimation of N is not possible. If individual probabilities of first capture vary over the population, however, it is logical to assume that $\bar{p}_1 > \bar{p}_2 > \bar{p}_3 > \cdots > \bar{p}_t$, because the individuals with high first capture probability will tend to be removed first, the animals with slightly lower first capture probabilities removed second, and so on. Moreover, the assumption that $(\bar{p}_1 - \bar{p}_2) > (\bar{p}_2 - \bar{p}_3) > \cdots > (\bar{p}_{t-1} - \bar{p}_t)$, i.e., the bigger differences in the conditional probabilities of removal occur in the initial stages of the experiment, does not seem illogical. These assumptions are the basis for the development of an estimation procedure appropriate for Model M_{bh} . The estimator \hat{N}_{bh} derived under Model M_{bh} is obtained by sequentially testing (through goodness of fit tests) for differences among the \bar{p}_j . The process begins with testing whether or not all the \bar{p}_j are equal. If not, we let \bar{p}_1 be different and test whether or not $\bar{p}_2 = \bar{p}_3 = \cdots = \bar{p}_t$. This testing continues until it is concluded that the last $t - k + 1$ capture probabilities are not significantly different, whereas the first k capture probabilities do differ. In the case $k = 1$, we are saying the simple Model M_b (all \bar{p}_j equal) adequately fits the removals. For $k > 1$ we are concluding Model M_b does not fit the data, due to the presence of heterogeneity (or time variation in capture probabilities). The estimator \hat{N}_{bh} is the ML estimator under the selected model.

The above technique, called the generalized removal method, is further described in Appendix H. The appendix also points out that it is not necessary that the factorization $G(p, c; \theta) = G_1(p; \theta)$

$G_2(c; \theta_2)$ be possible in order to use the generalized removal method to estimate population size. It is important to realize that this method is a generalization of Zippin's (1956, 1968) removal method which assumes no variation in first capture probabilities. Therefore, the fact that the generalized removal method has utility in removal experiments as well as capture-recapture experiments is not surprising (cf. REMOVAL MODELS).

Finally, we mention that the method can "fail" if a mathematical criterion involving the removals u_1, u_2, \dots, u_t is not satisfied. This failure criterion, similar to the one involved in Zippin's removal method, ensures that a sufficient decline in the number of newly captured animals is being effected by successive trapping occasions. A formula expressing this criterion is given in Appendix H.

Simulation Results

Bias

Experiments were simulated on populations behaving according to the assumptions of Model M_{bh} . The estimator \hat{N}_{bh} was calculated to develop some idea of the bias involved. Results indicate that relative bias ranges from 3 to 38 percent for populations used in the simulations. That range is somewhat misleading however, because for all populations except one (in which half the population was essentially untrappable), the range of bias was 3 to 15 percent. Some examples are given in Table 10. The reader is referred to Tables N.6.a and N.6.b of Appendix N for further results and for descriptions of the populations. In general, it seems there is no serious bias in \hat{N}_{bh} if relatively few members of the population are essentially uncatchable (i.e., probability of first capture less than 0.05) and the number of trapping occasions is adequate. (Recall that since the estimation technique depends on removals only, probabilities of recapture have no effect on the performance of \hat{N}_{bh} .) Considering the complicated model structure and the assumptions required to produce the

TABLE 10.—COMPUTER SIMULATION OF THE BIAS ASSOCIATED WITH THE ESTIMATOR \hat{N}_{bh} FOR DATA GENERATED UNDER MODEL M_{bh} (ALSO SEE APPENDIX N, TABLE N.6.b)

Percent relative bias RB	Ave[\hat{N}_{bh}]	Population size N	Number of reps R	Number of occasions t	Trials ¹
-38.4	246.6	400	100	5	1
-14.8	340.8	400	100	5	2
-5.9	94.1	100	100	5	7
-12.3	175.5	200	100	10	5
-3.2	193.7	200	100	7	6
-4.3	383.0	400	100	5	4

¹ The probability of capture for data generated under Model M_{bh} is complex; therefore, the specific values are given in Appendix N, Table N.6.a.

generalized removal estimator, these results are encouraging with respect to practical use of the estimator.

Confidence Intervals

Simulation results are not so encouraging, however, when it comes to placing a confidence interval on N . The variance estimator associated with \hat{N}_{bh} appears ill behaved and the distribution of \hat{N}_{bh} is nonnormal. Those factors resulted in simulated confidence coefficients averaging approximately 0.50. Further work is necessary to develop useful confidence intervals for N when Model M_{bh} seems appropriate because normal theory confidence intervals apparently are inappropriate for practical use.

Robustness

The generalized removal estimator can be expected to be robust to Models M_h and M_b , since both are special cases of Model M_{bh} . If capture probabilities exhibit significant variation over time, the utility of the method is undoubtedly decreased. That is particularly true if probabilities tend to increase with time, although the failure criterion should help to provide a safeguard against use of the method when it is inappropriate. In general, it is expected that the generalized removal method used in the context of a capture-recapture experiment will pro-

vide reasonable estimates of population size if the number of unmarked animals captured on each trapping occasion exhibits a definite decrease over time. Such a trend indicates that conditional probabilities of capture on the j^{th} occasion are reasonably large, which is the primary requirement for adequate estimation of population size.

Example

E. Larsen (pers. comm.) provided the results of livetrapping Great Basin pocket mice *Perognathus parvus* in a desert community, Curlew Valley, Utah. The area was trapped for 7 consecutive nights during June 1977. Traps were arranged in a 12×12 grid and spaced 15 m apart. Analysis of the resulting data by the model selection procedure (described in TESTS OF MODEL ASSUMPTIONS) indicated that either Model M_{tbb} or Model M_{bh} would be appropriate for use in estimating N . The data on first capture and the results of using \hat{N}_{bh} to produce point and interval estimates for N are given in Fig. 8. Note that the estimator rejects the model for $k = 1$, but accepts $k = 2$. Also of interest is that the estimation method fails for $k = 3$, but estimates for $k = 4$ and $k = 5$ are produced that appear reasonable. The estimated variance of \hat{N}_{bh} is quite large for $k = 2$. Possibly time variation is quite important and causes this imprecision. The probabilities of capture decline for $k = 2$, as they should under Model M_{bh} . A histogram of the u_j values is included to help the researcher visually examine his data.

Discussion

Model M_{bh} is one of the most realistic and useful models for a capture-recapture experiment considered in this monograph. Realism results from allowing every animal to possess its own probability of first capture and from not constraining the members of the population so that every individual's behavioral response to first capture is identical. The

OCCASION	J=	1	2	3	4	5	6	7	
TOTAL CAUGHT	M(J)=	0	23	32	35	40	46	52	55
NEWLY CAUGHT	U(J)=	23	9	3	5	6	6	3	

K	N-HAT	SE(N)	CHI-SQ.	PROB.	ESTIMATED P-BAR(J), J=1, . . . , 7						
1	60.61	4.456394	9.965	.0762	.2802	.2802	.2802	.2802	.2802	.2802	.2802
2	76.73	24.66909	3.486	.4801	.2998	.1377	.1377	.1377	.1377	.1377	.1377
3	FAILURE CRITERION = -2 NO ESTIMATES FOR THIS STEP.										
4	67.44	18.18989	1.308	.5201	.3410	.2025	.0846	.2067	.2067	.2067	.2067
5	58.65	5.896785	.623	.3643	.3922	.2525	.1126	.2114	.3953	.3953	.3953

POPULATION ESTIMATE IS 77 WITH STANDARD ERROR 24.6691
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 28 TO 126

HISTOGRAM OF U(J)

FREQUENCY	23	9	3	5	6	6	3

EACH * EQUALS	3 POINTS						
24	*						
21	*						
18	*						
15	*						
12	*						
9	*	*					
6	*	*	*	*	*	*	
3	*	*	*	*	*	*	*

FIG. 8. Example of population estimation with variable probability removal estimator under Model M_{bh} with pocket mouse data from E. Larsen (pers. comm.).

model is useful because an estimation procedure based on the assumptions of the model is available for estimating population size. Although the operating characteristics of the estimator are not completely satisfactory, the estimator would seem superior to any other that might be used if, in fact the assumptions of Model M_{bh} are satisfied. Moreover, the method is nonparametric in the sense that one does not have to specify a probability distribution for the pairs (p_i, c_i) .

MODEL M_{tbh} : CAPTURE PROBABILITIES VARY BY BEHAVIORAL RESPONSE TO CAPTURE, TIME, AND INDIVIDUAL ANIMAL

Discussion

For purposes of this monograph, Model M_{tbh} is useful only conceptually. Because all 3 factors that may affect capture probabilities are assumed to be operating

in Model M_{tbh} , every other model considered here is of course a special case of this model. Thus it may be useful to think of Model M_{tbh} as the most realistic of all models that assume population closure. Unfortunately, this extreme realism precludes the model from having any practical use in the estimation of population size. If the experimenter is led to believe that all the assumptions of Model M_{tbh} are required for the capture-recapture experiment at hand, none of the models considered in this publication are capable of producing valid statistical information concerning the population parameters of interest. If such a failure occurs, the researcher is forced to reevaluate the experimental design to discern how the experiment might be conducted to obtain useful information. This reevaluation may cause the capture-recapture method to be discarded in favor of some other completely different technique, e.g., line transect methods. Whatever the conclu-

sions, such an approach is much preferred over one that ignores necessary assumptions associated with the appropriate capture–recapture model and proceeds to use some simpler but inappropriate model. This latter approach requires the researcher to give credence to statistically invalid information about the population, and therefore violates basic tenets of scientific research.

REMOVAL MODELS

Introduction

In a capture–recapture experiment, all members of the population captured on a given trapping occasion are marked and released back into the population. In a removal experiment, however, as the name implies, members of the population are permanently removed and are not reintroduced into the population after first capture. The removal may be accomplished by killtrapping, electrofishing, trawling, or merely livetrapping the animals and physically displacing them to another area.

Removal methods have traditionally been associated with catch–effort experiments, which have been used often in practice (Omand 1951, Ketchen 1953, Fischler 1965). Such experiments are based on the idea that the size of the population will be decreased gradually as will the catch per unit effort by applying a known amount of removal effort on a number of occasions. The basic assumption behind catch–effort techniques is that the number of animals removed from the population directly depends upon the amount of effort expended in removing them. By using the size of the successive decreases and the known efforts that effected those decreases, initial population size N can be estimated. The experiment usually is modeled by assuming that all animals have probability of capture $p_j = 1 - \exp(-ke_j)$ on the j^{th} trapping occasion, where k is often called the “Poisson catchability coefficient” and e_j represents the known effort expended on the j^{th} occasion. Under such conditions and the

condition that units of effort act independently, a number of estimation techniques have been proposed. Although maximum likelihood estimation of N is possible (cf. Seber 1973:297), the best known estimation techniques are the regression techniques proposed by Leslie and Davis (1939), DeLury (1947), and Ricker (1975). All those methods involve regressing catch per unit effort against some function of the “catchability coefficient” and the efforts expended, and then using least squares analysis to provide an estimator of N and its variance.

Such catch–effort techniques will not be considered in our approach to removal experimentation for a number of reasons. First, and most important, we restrict our consideration of removal studies to those in which effort is deliberately kept constant on all occasions. In such studies, the concept of effort has no utility. Second, because the general approach in this monograph has been to void parametric assumptions concerning capture probabilities, the assumption that $p_j = 1 - \exp(-ke_j)$ is inappropriate for our purposes. Third, it is necessary for purposes of least squares analysis to make certain assumptions that involve the variance structure of the observations. Because to each postulated variance structure there corresponds a formula by which the estimate of N is calculated, such procedures are to some extent arbitrary in the absence of valid tests for determining proper variance structure. Finally, Braaten (1969) investigated the robustness of the DeLury (1958) estimator to departures from some of the assumptions and found that the estimator may exhibit significant negative bias.

Structure and Use of the Generalized Removal Model

If it is assumed that sampling effort is constant over trapping occasions, and that all animals in the population have the same probability of removal, then the removal model and estimation method considered by Moran (1951) results. That

method was discussed in conjunction with estimation in Model M_b . The context here, however, is different since animals are not reintroduced into the population after initial capture. Zippin (1956) showed that the joint distribution of $\{u_1, u_2, \dots, u_t\}$, where u_j represents the number of animals removed on the j^{th} occasion, can be written as

$$P\{\{u_1, u_2, \dots, u_t\}\} = \frac{N!}{\left[\prod_{j=1}^t u_j!\right] (N - M_{t+1})!} p^{M_{t+1}} \cdot (1-p)^{tN - \sum_{j=1}^t (t-j+1)u_j},$$

where p is the probability of removal of any animal on any trapping occasion. Because the parameter space is 2 dimensional (N and p are the only parameters) and the minimal sufficient statistic

$$\left\{M_{t+1}, \sum_{j=1}^t ju_j\right\}$$

is 2 dimensional, both parameters are identifiable and may be estimated by maximum likelihood. Such estimators are derived in Appendix D, and are the same as those used to estimate N and p in Model M_b .

An estimator of population size N based on the above model will clearly be unsatisfactory if all animals present do not have an equal probability of removal on a given trapping occasion. In removal experiments, unequal capture probabilities can be caused by time variation or heterogeneity or both. Behavioral variation is nonexistent since members of the population are removed after first capture. We maintain that proper planning and design of the removal experiment can be used to control or reduce time variation (see STUDY DESIGN). Thus, the main problem lies in dealing with heterogeneity of capture (removal) probabilities. In the following development, it is argued that the estimation procedure previously proposed for Model M_{bh} is also appropriate for estimation in removal

models. Such a procedure represents a generalization of Moran's (1951) and Zippin's (1956, 1958) constant probability removal model to the case in which heterogeneous probability of removal exists. If the set of removal probabilities p_i , $i = 1, 2, \dots, N$, is assumed to be a random sample of size N from some probability distribution $G(p; \theta)$ parameterized by θ and defined on $[0, 1]$, then the distribution of the vector of removals $\{u_1, u_2, \dots, u_t\}$ can be written as

$$P\{\{u_1, u_2, \dots, u_t\}\} = \frac{N!}{\left[\prod_{j=1}^t u_j!\right] (N - M_{t+1})!} \{E[p]\}^{u_1} \{E[(1-p)p]\}^{u_2} \dots \{E[(1-p)^{t-1}p]\}^{u_t} \{E[(1-p)^t]\}^{N-M_{t+1}},$$

where

$$E[(1-p)^{j-1}p] = \int_0^1 (1-p)^{j-1}p \, dG(p; \theta), \quad j = 1, \dots, t,$$

$$E[(1-p)^t] = 1 - \sum_{j=1}^t E[(1-p)^{j-1}p].$$

This distribution is identical to the distribution of the removals $\{u_1, u_2, \dots, u_t\}$ under the conditions discussed in the section on Model M_{bh} . Thus, it is easily argued (cf. Appendix J) that the generalized removal method developed for Model M_{bh} is also appropriate for estimation in removal models in which it is assumed that heterogeneity of first capture probability exists. Details of this estimation technique are given in Appendix H, along with an estimator of the asymptotic variance of the population estimator \hat{N}_R (the R stands for "removal," but we note that \hat{N}_R is mathematically the same as \hat{N}_{bh}) and a "failure" criterion that must not hold if parameters are to be validly estimated. This criterion ensures that a sufficient decline in the population is being effected by the successive removals.

As explained in the material on estimation of N in the section on Model M_{bh} ,

our procedure is to look at a sequence of removal models. These correspond to a sequence of assumptions as follows: all \bar{p}_j are equal, or $\bar{p}_1 \neq \bar{p}_2$ but $\bar{p}_2 = \bar{p}_3 = \dots = \bar{p}_t$, or $\bar{p}_1 \neq \bar{p}_2 \neq \bar{p}_3$, but $\bar{p}_3 = \bar{p}_4 = \dots = \bar{p}_t$, and so forth. The most general model allows all \bar{p}_j to be different, but this most general model does not allow estimation of N . We define the specific removal model M_{Rk} as the model in which the last $t - k + 1$ values of \bar{p}_j are the same, and for $k > 1$, the first \bar{p}_1 to \bar{p}_k are different.

The estimation procedure is to test the goodness of fit of the removal models sequentially from M_{R1} to $M_{R,t-2}$. Thus, we are first testing whether the constant probability model fits. If it does, we use the corresponding maximum likelihood estimator of N . If this simple model is rejected by the chi-square test (at the 20% significance level in program CAPTURE), we then examine the case (model M_{R2}) of $\bar{p}_1 \neq \bar{p}_2$, but $\bar{p}_2 = \bar{p}_3 = \dots = \bar{p}_t$. The estimator of N , \hat{N}_R , used in this generalized removal approach is taken as the ML estimator of N for the selected removal model.

Simulation Results

Bias and Confidence Intervals

A discussion of the bias of the estimator of N associated with the generalized removal method and of the usefulness of its associated confidence intervals was presented in the material on Model M_{bh} . Briefly, the simulation study revealed that the bias of the estimator of N was in most cases not serious, but that confidence intervals achieved only 50 percent coverage on the average. Although the results and discussion of that section were presented in the context of a capture-recapture experiment, all the material is directly applicable here. Because the 2 estimation procedures are identical, the relevant data in both cases involve only the removal statistics from populations with heterogeneous probabilities of first capture. Denoting the estimator of N as \hat{N}_R is intended to remind the reader

that this section deals with strict removal data.

Robustness

In the context of removal experiments, the generalized removal method represents a significant step forward with respect to robust estimation of population size. It should be emphasized, however, that the procedure is designed to be robust to failure of the assumption that all animals have the same probability of first capture, and not to failure of the assumption that sampling effort is uniform over trapping occasions. There is some indication in the simulation results that the method performs adequately when there is no heterogeneity of capture probability but there is nonuniformity in sampling rates over time. The method performs very poorly, however, when both heterogeneity and nonuniform sampling rates are present. First, the percentage of experiments which "fail," as determined by the failure criterion, can be very high. For instance, in the population defined by Trial 1 of Model M_{th} , the experiment failed in every one of 100 simulations. Secondly, when the experiment does succeed, bias is usually significant. Some simulated examples are given in Table 11 which help to substantiate these remarks. Complete simulation results are given in Table N.6.b of Appendix N.

Example

Andrzejewski and Jezierski (1966) reported the results of a study designed to estimate population density of European hare *Lepus europaeus* on experimental hunting grounds in Poland. Hares were captured and removed by driving them into nets surrounding the area. Results of the application of the generalized removal method to the data resulting from the 4 drives (removal occasions) are given in Fig. 9. When $k = 1$ (i.e., all animals are assumed to have the same probability of removal), a poor fit to the data results (chi-square goodness of fit value is 13.5, 2 df, and $p = 0.0014$). For $k = 2$, how-

OCCASION	J=	1	2	3	4
TOTAL CAUGHT	M(J)=	0	722	913	982 1018
NEWLY CAUGHT	U(J)=	722	191	69	36

K	N-HAT	SE(N)	CHI-SQ.	PROB.	ESTIMATED P-BAR(J), J=1, . . . , 4			
1	1028.21	3.789703	13.150	.0014	.6806	.6806	.6806	.6806
2	1039.10	7.658751	1.528	.2164	.6948	.5916	.5916	.5916

POPULATION ESTIMATE IS 1039 WITH STANDARD ERROR 7.6588
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 1023 TO 1055

HISTOGRAM OF U(J)

FREQUENCY	722	191	69	36

EACH * EQUALS	73 POINTS			
730	*			
657	*			
584	*			
511	*			
438	*			
365	*			
292	*			
219	*	*		
146	*	*		
73	*	*	*	

FIG. 9. Example of population estimation under the variable probability removal model with data on European hare from Andrzejewski and Jezierski (1966).

ever, an acceptable fit results ($p = 0.2164$) and, therefore, the point and interval estimates corresponding to Model M_{R2} are chosen. In particular, the point estimate $\hat{N}_R = 1,039 \pm 7.66$ seems much preferable to the estimate of 1,010 that results from the (commonly used) regression method used by Andrzejewski and Je-

TABLE 11.—COMPUTER SIMULATION RESULTS ILLUSTRATING THE ROBUSTNESS OF THE GENERALIZED REMOVAL ESTIMATOR \hat{N}_R USING DATA GENERATED UNDER OTHER MODELS (ALSO SEE APPENDIX N, TABLES N.2.b, AND N.5.b)

Percent relative bias RB	Ave[\hat{N}_R]	Population size N	Number of reps R	Number of occasions t	Data model
- 3.2	387.1	400	100	10	M_t , Trial 6 ¹
- 9.0	364.1	400	100	7	M_t , Trial 7
2.6	410.3	400	100	7	M_t , Trial 9
- 0.2	199.7	200	100	5	M_t , Trial 10
- 1.9	272.6	400	100	5	M_{th} , Trial 2
22.3	489.1	400	100	5	M_{th} , Trial 4
-10.3	179.4	200	100	7	M_{th} , Trial 5

¹ For example, data generated under Model M_t were used to estimate N, using the generalized removal method \hat{N}_R .

OCCASION	J=	1	2	3	4	5	
TOTAL CAUGHT	M(J)=	0	181	192	196	201	204
NEWLY CAUGHT	U(J)=	181	11	4	5	3	

K	N-HAT	SE(N)	CHI-SQ.	PROB.	ESTIMATED P-BAR(J), J=1, ..., 5				
1	204.00	.2094765	70.820	0.0000	.8160	.8160	.8160	.8160	.8160
2	206.77	3.697937	1.555	.4596	.8754	.4029	.4029	.4029	.4029
3	208.43	8.365668	1.021	.3124	.8684	.4010	.3306	.3306	.3306

POPULATION ESTIMATE IS 207 WITH STANDARD ERROR 3.6979
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 199 TO 215

HISTOGRAM OF U(J)

FREQUENCY	181	11	4	5	3

EACH * EQUALS	19 POINTS				
190	*				
171	*				
152	*				
133	*				
114	*				
95	*				
76	*				
57	*				
38	*				
19	*	*			

FIG. 10. Example of population estimation under the variable probability removal model with aquatic insect data from R. F. Raleigh (pers. comm.).

zierski, because the number of hares actually removed was 1,018. The estimated capture probabilities for Model M_{R2} are $\bar{p}_1 = 0.6948$, and $\bar{p}_2 = \bar{p}_3 = \bar{p}_4 = 0.5916$, a result that supports the idea that heterogeneity is operating.

Example

R. F. Raleigh (pers. comm.) provided some results of a removal experiment involving a species of mayfly *Ephemeroptera*. Mayflies were sampled from 10 randomly placed 0.25-m² areas in a section of the Poudre River streambed near Fort Collins, Colorado, with 5 removal occasions at each site. A special benthic aquatic sampler was used in the study. Because these are true removal data, the

generalized removal estimation procedure was used to produce point and interval estimates for N (Fig. 10). As in the previous example, there is a poor fit for the simple model with constant capture probability ($k = 1$). When $k = 2$, however, a good fit to the data results and an estimate $\hat{N}_R = 207$, corresponding to Model M_{R2} is produced. Note the difference between $\bar{p}_1 = 0.8754$ and $\bar{p} = 0.4029$, indicating that a significant heterogeneity in removal probabilities may exist. The computed confidence interval is somewhat unsatisfactory because its lower limit is less than the number of animals actually seen, and reminds one that in removal and livetrapping experiments, conditions necessary for construction of normal theory confidence intervals are

OCCASION	J=	1	2	3	4	5	6	7	
TOTAL CAUGHT	M(J)=	0	25	51	66	79	91	104	109
NEWLY CAUGHT	U(J)=	25	26	15	13	12	13	5	

K	N-HAT	SE(N)	CHI-SQ.	PROB.	ESTIMATED P-BAR(J), J=1, . . . , 7							
1	138.07	14.67943	3.879	.5670	.1980	.1980	.1980	.1980	.1980	.1980	.1980	.1980
2	132.46	13.64926	3.623	.4594	.1887	.2218	.2218	.2218	.2218	.2218	.2218	.2218
3	141.63	25.85191	2.754	.4311	.1765	.2229	.1829	.1829	.1829	.1829	.1829	.1829
4	130.96	20.14541	2.688	.2608	.1909	.2454	.1876	.2339	.2339	.2339	.2339	.2339
5	118.54	10.21073	2.133	.1442	.2109	.2779	.2221	.2474	.3675	.3675	.3675	.3675

POPULATION ESTIMATE IS 138 WITH STANDARD ERROR 14.6794

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 109 TO 167

HISTOGRAM OF U(J)

FREQUENCY	25	26	15	13	12	13	5

EACH * EQUALS	3 POINTS						
27		*					
24	*	*					
21	*	*					
18	*	*					
15	*	*	*				
12	*	*	*	*	*	*	
9	*	*	*	*	*	*	
6	*	*	*	*	*	*	*
3	*	*	*	*	*	*	*

FIG. 11. Example of population estimation data under the variable probability removal model with whitefish data from Ricker (1958:150).

not satisfied. In this case, one should truncate the lower limit to the actual number removed, rather than using a lower limit less than M_{t+1} .

Example

Ricker (1958:150) removed whitefish *Coregonus clupeaformis* from Shakespeare Island Lake in Ontario, Canada, on 7 successive occasions by means of gillnetting. Members of whitefish removed in the 13- to 14-inch (3.3–3.6 cm) length class are shown in Fig. 11, with the results of the generalized removal method. Notice that a good fit to the removals is achieved for $k = 1$, so that Zippin's (1956, 1958) constant probability removal model seems appropriate for estimating N . The population estimate of $\hat{N}_R = 138$ is in close agreement with Ricker's estimate of 136 which was obtained by using DeLury's (1947) regres-

sion method assuming equal effort. (We wish to reemphasize that on theoretical grounds the ML estimation procedure is superior to the various regression techniques often used with removal studies.) The 95 percent confidence interval estimate of [109, 167] is very informative. Moreover, the true confidence level of the interval probably is close to the stated level of 0.95, since the estimated probability of removal is nearly 0.20.

Discussion

Conducting a removal experiment for purposes of estimating population size may sometimes prove more feasible than a capture-recapture approach (refer to STUDY DESIGN for further discussion). In such cases, the experimenter has available 2 classes of estimation procedures, the catch-effort techniques usually associated with Leslie and Davis (1939) and

DeLury (1947) or the "removal" techniques first introduced by Moran (1951), refined by Zippin (1956, 1958), and generalized here. It is felt that catch-effort techniques are often not appropriate, either because of the assumptions involved or because the concept of effort may be meaningless in many experimental situations. In those frequent situations, we believe the generalized removal method provides the best approach to estimating population size. The operating characteristics of this method are by no means completely satisfactory, in view of the results concerning confidence interval coverage and the failure of the experiment in some situations. However, the fact that the method fails in a given experiment does at least prevent the use of wildly inaccurate estimates in practice and helps to inform the experimenter that the assumptions of the method are not met and the quality of the experiment needs to be upgraded or the design altered, or both. Nevertheless, the method is the most general now available in the literature that is capable of providing useful results and, with further research, improvements in the method should be forthcoming.

TESTS OF MODEL ASSUMPTIONS

In preceding sections, we have recognized 3 distinct sources of variation in capture probabilities, and have given 8 models corresponding to presence or absence of specific types of variability. No estimator of population size has been derived from 3 of those models (M_{th} , M_{tb} , M_{tbh}). Consequently, we have introduced only 5 estimators that are identified below, along with the models where they should be used. The main purpose of this section is to give a strategy for selecting one of the estimation procedures based on tests of the various underlying assumptions.

Philosophy of the Approach

Our philosophy is that one should first present the most general model structure

(i.e., assumptions) possible for the given data. Then a series of specific alternative models should be developed based on specific assumptions that cover the various cases of realistic or theoretical interest. Those alternative models form the basis for both testing of assumptions and estimating parameters. Under FUNDAMENTAL CONCEPTS, we gave the general model framework considered here: population closure is assumed and captures are assumed to be independent events with capture probabilities

$$p_{ij}, j = 1, \dots, t, i = 1, \dots, N.$$

Such a model structure is useful only as a conceptual starting point because none of the parameters are estimable. Next, we recognized 3 different sources of variation acting on these capture probabilities: (1) variation over time, (2) behavioral variation as a result of first capture (trap response), and (3) variation over individuals (heterogeneity). We also recognized the various possible combinations of these sources of variation.

An almost infinite variety of very specific models could be constructed. We examined 8 models likely to be generally useful for estimation or testing purposes. Those models can all be expressed in terms of the assumptions about variation in the capture probabilities. If each of the 8 models introduced had its own unique estimation procedure, then selection of an estimator would be synonymous with selection of a model. Such is not the case with only 5 estimators and 8 models. Our strategy is to present a testing sequence designed to lead to selection of the most appropriate model. We define "appropriate model" as the simplest model that "fits" the data. When that selection gives a model such as M_{th} for which there is no theoretically appropriate estimator, we recommend (in some instances) using the next best fitting model for which an estimator exists. It will occasionally be necessary to conclude that no estimate of N can legitimately be calculated from the data.

It is important to recognize that we do

not claim this model selection procedure results in the “true” model. By true model for a capture study we mean that the stated model is an exact representation of the true capture probability structure for that study. Any model is a simplification that can realistically represent only the more predominant features of the study. Also, we must recognize that the most appropriate model will depend upon the amount of data we have. With good data (large grids, many occasions) one might be able to show, for example, that Model M_{bh} was necessary. Yet, if the same study used a much smaller grid, and fewer days, it might be impossible to demonstrate heterogeneity from the data themselves and Model M_b might be indicated as appropriate.

The conceptual goal of our model selection procedure is to achieve an acceptable trade-off between precision and bias. If one uses too simple a model, the estimated population size is likely to be severely biased yet have a deceptively small sampling variance. If the model is complex, but still the wrong one, then estimators are again of uncertain bias. If a model is chosen that is too general, the risk of bias is much reduced, or negligible, but the sampling variance is larger than it needs to be (i.e., we lose precision). Of those 2 cases, the safer one in our opinion is to choose the more general model in hopes of minimizing biases. We state this as our conceptual goal because, not knowing the true model, we can never be certain when analyzing real data that we have made the best choice of a model.

This model selection procedure is based on tests of assumptions of 2 types: (1) tests between specific models, and (2) general goodness of fit tests to specific models. Before discussing these tests we summarize the primary models and associated estimators.

Summary of Models and Estimators

Capture-Recapture Models

Because these models have been described in detail, only a summary is given

below, in terms of assumptions and corresponding capture probabilities.

MODEL M_o : $p_{ij} \equiv p$.

Comment: This is the most restrictive model because capture probabilities are assumed to be constant.

Estimation: We recommend the ML estimator (see Appendix B); it does not exist in closed form except for $t = 2$.

MODEL M_t : $p_{ij} \equiv p_j, j = 1, \dots, t$.

Comment: Capture probabilities vary only with time. This is the model most frequently assumed (often implicitly) in the literature.

Estimation: We recommend the ML estimator (see Appendix C), it does not exist in closed form, except for $t = 2$ (in which case it is the Lincoln estimator). The Schnabel estimator, or variations thereof, is often used with this model, but is only an approximation to the ML estimator.

MODEL M_b : $p_{ij} \equiv \begin{cases} p & \text{for first capture,} \\ c & \text{for recapture.} \end{cases}$

Comment: This is the simplest model of behavioral (trap) response and it has been recognized in the literature (Tanaka 1956).

Estimation: The ML estimator, which is recommended (Appendix D), is a special case of the generalized removal estimator. Zippin's removal model is equivalent to Model M_b for purposes of estimating N , but his suggested estimator is only a close approximation to the true ML estimator.

MODEL M_h : $p_{ij} \equiv p_i, i = 1, \dots, N$.

Comment: This model allows heterogeneity of capture probabilities over animals, but allows no variation in capture probabilities over time.

Estimation: The jackknife estimator (Appendix E) is recommended when M_h is the underlying model; that estimator was derived specifically for this model. Note, however, that there

is no way to derive an optimal estimator for Model M_h if nothing is known about how the capture probabilities vary.

$$\text{MODEL } M_{tb}: p_{ij} \equiv \begin{cases} p_j & \text{for any first capture, } j = 1, \dots, t, \\ c_j & \text{for any recapture, } j = 2, \dots, t. \end{cases}$$

Comment: This model allows variation in capture probabilities due to both time and behavior (trap response).
 Estimation: Population size N is not estimable under this model. Therefore, there is no satisfactory estimation procedure associated with this model.

$$\text{MODEL } M_{th}: p_{ij} \equiv p_i p_j, \quad j = 1, \dots, t, \quad i = 1, \dots, N.$$

Comment: This is useful as a conceptual model of how time and heterogeneity might operate as a simple product. No estimator can be derived from this model.

Estimation: Population size N is not estimable under this model if straightforward methods are used. Therefore, there is no satisfactory estimation procedure associated with this model

$$\text{MODEL } M_{bh}: p_{ij} \equiv \begin{cases} p_i & \text{for first capture,} \\ c_i & \text{for all recaptures,} \\ & i = 1, \dots, N. \end{cases}$$

Comment: This model assumes heterogeneity with trap response, but no time variation as such in the capture probabilities.

Estimation: The generalized removal estimator should perform well here (Appendix H). However, if none of the generalized removal models fit, no estimate can be produced using this model.

$$\text{MODEL } M_{tth}: p_{ij} \equiv \begin{cases} p_{ij} & \text{for first capture,} \\ c_{ij} & \text{for all recaptures,} \\ & j = 1, \dots, t, \\ & i = 1, \dots, N. \end{cases}$$

Comment: This is a conceptually useful model because it represents the

TABLE 12.—MODELS AND SUGGESTED ESTIMATED PROCEDURES

Model	Theoretically appropriate estimator
M_o	Null (\hat{N}_o)
M_t	Darroch (\hat{N}_t)
M_b	Zippin (\hat{N}_b)
M_h	Jackknife (\hat{N}_h)
M_{tb}	(none)
M_{th}	(none)
M_{bh}	Generalized removal (\hat{N}_{bh})
M_{tth}	(none)

case in which all 3 sources of variation operate.

Estimation: Population size N is not estimable under this model using straightforward methods. Therefore, there is no satisfactory estimation procedure associated with this model.

Relationships Between Models

Table 12 summarizes some of the above information. It is important to understand the relationship of the models to one another in order to understand testing between models. One can only truly test one model against another when one model is a special case of the other. For example, Model M_o is a special case of each of the 3 models M_t , M_b , and M_h . But those 3 models are not related to each other in any simple manner that allows one to choose or test between them with any simple statistical tests.

The relationships among these models is shown in Fig. 12 using an arrow between 2 models to indicate that one is a special case of the other. For example, $M_o \leftarrow M_t$ means Model M_o is a special case of Model M_t , because if Model M_t is restricted by the assumption $p_1 = p_2 = \dots = p_t = p$, Model M_o results.

The selection of an appropriate capture-recapture model is neither simple nor straightforward and there is, in fact, no optimal or rigorous statistical theory to guide this selection. The jumble of arrows in Fig. 12 is indicative of the difficulty.

Specific Tests to Perform

Before we can give our strategy for model, hence estimator, selection, it is necessary to discuss exactly what hypothesis can be tested concerning variations in capture probabilities. Two types of tests are made:

- (1) Specific tests of one model versus a more general model. For example, testing Model M_0 vs. Model M_t tests for time variation in average daily capture probabilities;
- (2) General goodness of fit tests of a given model to the data. For example, testing whether Model M_t fits the data amounts to testing whether time variation is the only source of variation in capture probabilities.

There is a basic difference in the nature of the hypotheses associated with those 2 different classes of tests. The objective of the first class of tests is to establish whether or not the more general model produces a significantly better fit to the data than the more restrictive model. On the other hand, the goodness of fit tests of the second class test whether or not the data might reasonably have arisen from the specified model.

Table 13 gives the 7 specific tests we have used in the model selection procedure. The reader should study Table 13 before proceeding. Program CAPTURE has been used to produce an example of the tests. Results are presented in Fig. 13, that is based on simulated data under Model M_0 with parameters $N = 100$, $p = 0.50$, and $t = 5$. The first 3 tests of the model selection procedure separately test each of the simple models M_b , M_{bh} , M_t against Model M_0 . These are all chi-square tests (see Appendix K for details). The significance level of the tests is given by the program. For example, in Fig. 13 under the test of Model M_0 vs. Model M_b (test 1) we have a chi-square test statistic of 5.007 (3 degrees of freedom) and the significance level (labeled PROBABILITY OF LARGER VALUE) is 0.17127. Thus, that test is not significant, and we have no

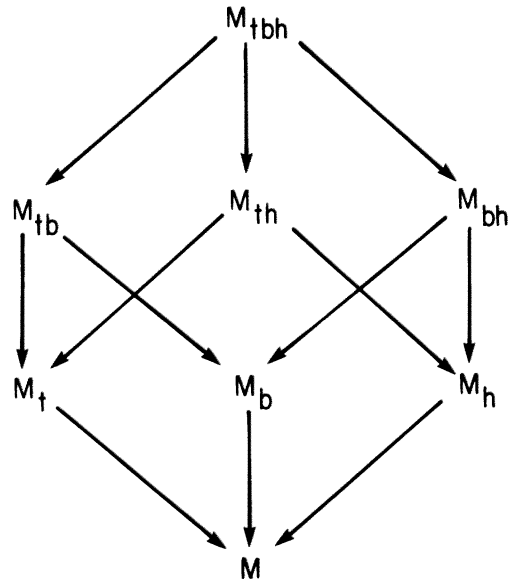


FIG. 12. Relationships among models; arrows indicate that one model is an immediate special case of another model.

basis to believe there is any heterogeneity of capture probabilities.

Similarly, the simple tests for behavioral variation (Model M_0 vs. Model M_{bh} , test 2) and for time variation (Model M_0 vs. Model M_t , test 3) are not significant in Fig. 13. Those tests had significance levels of 0.98056 and 0.31935, respectively. Based on only these results, we would (correctly, in this case) conclude the appropriate model is Model M_0 . When a model more complex than M_0 is the appropriate one, additional tests are needed.

In the second part of the model selection procedure, 4 more tests are presented. The first 3 are simple goodness of fit tests to models M_b , M_{bh} , and M_t . The seventh is a test for behavioral response in the presence of heterogeneity; hence this test helps make a choice between M_b and M_{bh} . That last test is included here because we have an estimator for Model M_{bh} .

The goodness of fit test of Model M_h (test 4) in Fig. 13 results in a chi-square value of 4.556 (4 df) and has a significance

OCCASION	J=	1	2	3	4	5
ANIMALS CAUGHT	N(J)=	55	47	54	60	48
TOTAL CAUGHT	M(J)=	0	55	75	88	96
NEWLY CAUGHT	U(J)=	55	20	13	8	2
FREQUENCIES	F(J)=	10	38	26	20	4

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)

CHI-SQUARE VALUE = 5.007 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .17127

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)

CHI-SQUARE VALUE = .001 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = .98056

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)

CHI-SQUARE VALUE = 4.701 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .31935

4. GOODNESS OF FIT TEST OF MODEL M(H)

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)

CHI-SQUARE VALUE = 4.556 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .33601

TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
(FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

NUMBER OF CAPTURES CHI-SQUARE D.F. PROBABILITY

NUMBER OF CAPTURES	CHI-SQUARE	D.F.	PROBABILITY
1	5.000	4	.28730
2	1.298	4	.86167
3	1.949	4	.74519
4	5.500	4	.23973

5. GOODNESS OF FIT TEST OF MODEL M(B)

NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)

CHI-SQUARE VALUE = 5.183 DEGREES OF FREEDOM = 6 PROBABILITY OF LARGER VALUE = .52053

5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

CHI-SQUARE VALUE = 2.459 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .48276

5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

CHI-SQUARE VALUE = 2.724 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .43610

6. GOODNESS OF FIT TEST OF MODEL M(T)

NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)

CHI-SQUARE VALUE = 59.465 DEGREES OF FREEDOM = 54 PROBABILITY OF LARGER VALUE = .28338

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)

CHI-SQUARE VALUE = 7.471 DEGREES OF FREEDOM = 10 PROBABILITY OF LARGER VALUE = .68036

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL CRITERIA	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
	1.00	.96	.26	.51	0.00	.44	.33	.64

TABLE 13.—SPECIFIC TESTS OF ASSUMPTIONS USED IN THE MODEL SELECTION PROCEDURE

Test number	Source of variation tested for	Null hypothesis	Alternative hypothesis	Comments
1.	Heterogeneity	Model M_0 fits the data	Model M_h fits the data	This test examines the capture frequencies to see if there is evidence of variability among individual capture probabilities
2.	Trap response after first capture	Model M_0 fits the data	Model M_b fits the data	This is a test for gross behavior effects on capture probabilities
3.	Time variation in capture probabilities	Model M_0 fits the data	Model M_t fits the data	This tests for variation in average daily capture probabilities. Unfortunately the actual test is also sensitive to Model M_b
4.	Trap response and/or time variation given heterogeneity	Model M_h fits the data	Model M_h fails to fit the data	If M_h is the true model we expect this test not to reject. We would also expect test 1 to reject Model M_0 in favor of M_h
4a.	As above	As above	As above	These tests are a specific form of test 4, where for each significantly large capture frequency an individual test can be made for trap response or time variation or both
5.	Heterogeneity and/or time variation given trap response	Model M_b fits the data	Model M_b fails to fit the data	If M_b is the true model we expect this test not to reject. We also would expect test 2 to reject Model M_0 in favor of M_b . This goodness of fit test can be partitioned into two specific tests, useful in their own right (5a, 5b)
5a.	Heterogeneity and/or time variation using first capture only	First capture probabilities are constant	First capture probabilities vary by time and/or animals	This test is identical to the goodness of fit test for the simple removal model
5b.	Heterogeneity and/or time variation using only recaptures	Recapture probabilities are all constant	Recapture probabilities vary by time and/or animals	If M_b is true then this null hypothesis should not be rejected
6.	Trap response and/or heterogeneity given time variation	Model M_t fits the data	Model M_t fails to fit the data	If M_t is the model, then we expect this test to fail to reject, and we expect Model M_0 to be rejected in favor of M_t
7.	Trap response given heterogeneity	Model M_h fits the data	Model M_{bh} fits the data	This test is useful because if we reject M_h in favor of M_{bh} the estimator to use is the generalized removal method

←

FIG. 13. Example of the model selection procedure tests with simulated data. True model is M_0 , with parameters $N = 100$, $t = 5$, and $p = 0.5$. Appropriate model is M_0 . Suggested estimator is Null.

level of 0.33601. Thus, in this example Model M_h fits the data. This is expected because Model M_o is a special case of Model M_h , and hence the tests should not reject.

When individual capture frequencies are large enough (we have used the criteria $f_i > 2t$) we have computed a goodness of fit test of Model M_h based on the data for animals captured exactly f_i times. Theoretically, this can be done for all frequencies f_1, \dots, f_{i-1} that are large enough.

The goodness of fit test of Model M_b (test 5) can also be partitioned into 2 components. Test 5a is actually the goodness of fit test to Zippin's removal model, i.e., it tests whether the probability of first capture is constant over time. In Fig. 13, the test statistic value is 2.459 with significance level 0.48276. Thus, in this case there is no evidence of variation in first capture probability over time. The second component of the test of Model M_b (test 5b) examines for time variation in recapture probabilities. In Fig. 13, the test statistic value is 2.724. The goodness of fit test statistic for Model M_b is the sum of those 2 chi-square values. In Fig. 13, the value is 5.183 (6 df) with significance level 0.52053. Thus, we conclude Model M_b fits the data. Again, we point out that Model M_o is a special case of M_b , so the result is not surprising.

The final goodness of fit test is for Model M_t (test 6). In Fig. 13, the test statistic value of 59.465 (54 df) is not significant ($P = 0.28338$).

An alternate way of thinking of these tests is as follows. The test of Model M_o vs. M_h tests whether there is heterogeneity of capture probabilities while the goodness of fit test of Model M_h tests whether there is any additional source of variation in capture probabilities due either to behavior or time or both. If we rejected Model M_o in favor of M_h and the goodness of fit test to M_h indicates M_h fits the data, we could conclude Model M_h was appropriate.

Similarly, the tests of Model M_o vs. M_b test for whether there is behavioral variation in capture probabilities, while the

goodness of fit tests to Model M_b examine whether any other source of variation is operating.

The 7 tests have been simulated to study their size and power, and the results are presented in Appendix N. In general, all the tests have good size properties, i.e., they tend not to reject when the null hypothesis is true. However, most of the tests lack power, i.e., they do not always reject when they should. This is particularly true for tests of heterogeneity (M_o vs. M_h) and the goodness of fit test of M_h .

On the Need for an Objective Selection Procedure

The above discussion and Fig. 13 illustrate the general model selection approach. Conceptually, we examine the results of the 7 tests of assumptions, and on the basis of the results choose the appropriate model. In the example used (Fig. 13), the selection was not difficult. However, for other models, and for much real data we have seen, the judgemental selection of an appropriate estimator can be very difficult because (1) the tests are not independent, in particular if behavioral variation in capture probabilities is present this can strongly affect the test for time and heterogeneity effects; and (2) for real data all 3 sources of variability probably are present in varying degrees, so all we can hope to do is select the most appropriate model (the one that best describes the data). With respect to the first point, consider what happens if Model M_b is true. For example, assume 7 trapping occasions with M_b as the true model, and $p = 0.40$, $c = 0.20$. That is, prior to being captured animals have a 0.40 capture probability, but due to behavioral response (trap avoidance in this case) their capture probabilities drop to only 0.20 after first capture. The expected daily capture probability is given by

$$p_j = \frac{E(n_j)}{N}, \quad j = 1, \dots, 7.$$

These are constant only if the expected

daily captures are constant. But as trapping progresses the average daily capture probability decreases as more and more animals “shift” their capture probability from 0.40 to 0.20 as a result of trap response. A formula for p_j in this case is

$$p_j = \frac{(1 - (1 - p)^{j-1})c + (1 - p)^{j-1}p^2}{1 - (1 - p)^j},$$

$j = 2, \dots, t$

($p_1 \equiv p$). For the case above, we have

j	p_j
1	0.400
2	0.275
3	0.237
4	0.220
5	0.211
6	0.207
7	0.204

When Model M_b is true, the test for time variation in capture probabilities (Model M_o vs. M_t) will tend to reject Model M_o because the behavioral response does indeed “cause” time variation to be present also.

The same sort of argument shows that behavioral response “causes” heterogeneity on any given day. Consequently, when Model M_b is true, there is a strong tendency for all of the simple tests to reject Model M_o , making selection of the correct model difficult.

In the final analysis, selection of a model involves examining a point in a 7-dimensional space and trying to classify it into 1 of 8 categories. We have not been successful ourselves in providing a set of simple rules for this choice and believe that field biologists without rigorous statistical training might have great difficulty arriving at a proper choice.

*An Objective Model
Selection Procedure*

The problem identified in the previous section falls in the realm of multivariate statistics, specifically classification of an unknown entity into one of several different known populations. In the classi-

cal version of this problem, one has a sample of measurements from each known population, and from those data a mathematical rule is constructed for classifying future cases based on their measurements. In our case, the “measurements” are the significance levels from the 7 tests. The “populations” are the 8 models. What is unknown is what model best fits any capture data at hand. Given this conceptualization, we chose to use discriminant function analyses to construct the classification function (see Cooley and Lohnes 1962). The objective of the discriminant analysis is to weight and linearly combine the significance levels of the 7 tests in some fashion so that the models are forced to be as statistically distinct as possible.

The model classification function was derived from simulated data where the true model is known. The capture probability structure of these simulated data is presented in Table 14. These parameter values were chosen because they are, in our opinion, representative of many real capture studies. For each population, 100 replications were generated, giving a total of 1,600 cases. The probability levels from each of the 7 tests were then used to construct the classification function. In addition to the 7 probabilities, 9 additional variables were constructed by taking all possible products between probability levels from tests 1–3 and tests 4–6, i.e. by taking the 9 pairwise products of the probabilities from the 3 specific tests against Model M_o and the 3 goodness of fit tests. This provides a total of 16 variables that were transformed by $X = \log(P + 0.01)$. That transformation tends to weight the smaller probability levels more heavily than the larger values close to 1. The classification rule resulting from that series of transformations was one of many rules examined; this particular rule was chosen because it had the best performance.

The transformed variables were then used to compute a classification function using the SPSS discriminant procedure (Nie et al. 1975). The classification func-

TABLE 14.—CAPTURE PROBABILITY STRUCTURE FOR THE SIMULATED DATA USED TO GENERATE THE CLASSIFICATION FUNCTION. ONE HUNDRED REPLICATIONS OF EACH TYPE WERE GENERATED, GIVING A TOTAL SAMPLE SIZE OF 1,600

Model	Probability Structure
LARGE POPULATION (N = 400, t = 7)	
M _o	p = 0.30.
M _h	p ₁ = 0.15, i = 1,100; p ₁ = 0.25, i = 101,200; p ₁ = 0.35, i = 201,300; p ₁ = 0.45, i = 301,400.
M _b	p = 0.2; c = 0.5.
M _{bh}	p ₁ = 0.15, c ₁ = 0.375, i = 1,100; p ₁ = 0.25, c ₁ = 0.625, i = 101,300; p ₁ = 0.35, c ₁ = 0.875, i = 301,400.
M _t	p ₁ = 0.15; p ₂ = 0.45; p ₃ = 0.35; p ₄ = 0.25; p ₅ = 0.30; p ₆ = 0.20; p ₇ = 0.40.
M _{th}	p _{ij} = p _i · p _j ; p ₁ = 0.45, i = 1,100; p ₁ = 0.55, i = 101,200; p ₁ = 0.65, i = 201,300; p ₁ = 0.75, i = 301,400; p ₁ = 0.65; p ₂ = 0.75; p ₃ = 0.55; p ₄ = 0.35; p ₅ = 0.60; p ₆ = 0.50; p ₇ = 0.80.
M _{tb}	p _{ij} = p _i · c; c = 2.50; p ₁ = 0.05; p ₂ = 0.35; p ₃ = 0.25; p ₄ = 0.15; p ₅ = 0.20; p ₆ = 0.10; p ₇ = 0.30.
M _{tbn}	p _{ij} = p _i · p _j · c; p ₁ = 0.15, i = 1,100; p ₁ = 0.25, i = 101,200; p ₁ = 0.35, i = 201,300; p ₁ = 0.45, i = 301,400; c = 2.50; p ₁ = 0.65; p ₂ = 0.75; p ₃ = 0.55; p ₄ = 0.45; p ₅ = 0.60; p ₆ = 0.50; p ₇ = 0.70.
SMALL POPULATION (N = 100, t = 5)	
M _o	p = 0.10.
M _h	p ₁ = 0.05, i = 1,25; p ₁ = 0.15, i = 26,50; p ₁ = 0.25, i = 51,75; p ₁ = 0.35, i = 76,100.
M _b	p = 0.40; c = 0.20.
M _{bh}	p ₁ = 0.15, c ₁ = 0.075, i = 1,33; p ₁ = 0.25, c ₁ = 0.125, i = 34,66; p ₁ = 0.35, c ₁ = 0.175, i = 67,100.
M _t	p ₁ = 0.15; p ₂ = 0.45; p ₃ = 0.35; p ₄ = 0.25; p ₅ = 0.30.
M _{th}	p _{ij} = p _i · p _j ; p ₁ = 0.35, i = 1,25; p ₁ = 0.45, i = 26,50; p ₁ = 0.55, i = 51,75; p ₁ = 0.65, i = 76,100; p ₁ = 0.65; p ₂ = 0.75; p ₃ = 0.55; p ₄ = 0.35; p ₅ = 0.60.
M _{tb}	p _{ij} = p _i · c; c = 0.5; p ₁ = 0.35; p ₂ = 0.25; p ₃ = 0.15; p ₄ = 0.20; p ₅ = 0.30.
M _{tbn}	p _{ij} = p _i · p _j · c; p ₁ = 0.15, i = 1,25; p ₁ = 0.25, i = 26,50; p ₁ = 0.35, i = 51,75; p ₁ = 0.45, i = 76,100; c = 0.75; p ₁ = 0.65; p ₂ = 0.75; p ₃ = 0.55; p ₄ = 0.45; p ₅ = 0.60.

tion consists of 8 sets of coefficients (1 set for each of the 8 models) that are used to form linear combinations of the 16 variables plus an intercept. The classification equations are derived from the pooled within-model covariance matrix and the

centroids for the transformed probability levels. The *i*th function has the form

$$C_i = c_{i0} + c_{i1} \cdot X_1 + c_{i2} \cdot X_2 + \dots + c_{i16} \cdot X_{16}.$$

The 136 classification coefficients are

TABLE 15.—CLASSIFICATION FUNCTION COEFFICIENTS USED IN THE MODEL SELECTION PROCEDURE

Coefficient	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
C0	-8.09154	-14.33446	-25.69392	-17.15981	-31.90722	-23.20432	-26.11839	-15.07368
C1	2.44865	1.55835	2.41361	2.09104	4.85352	4.52842	4.58145	1.01685
C2	0.01513	0.76894	0.34328	-0.30063	4.19868	3.36253	-1.18146	-0.78482
C3	5.60627	7.61750	4.45155	2.65005	-3.83028	-4.33794	3.81010	1.97090
C4	2.31654	2.43157	-1.90418	1.85727	4.11214	3.84041	0.07484	5.00658
C5	0.39009	0.24251	2.57294	2.95937	-0.37644	-0.30277	-0.73170	-1.36695
C6	1.59847	1.50362	5.40943	0.82379	3.29315	2.76447	4.77294	0.91642
C7	-1.07536	-0.75833	-1.39342	-0.92010	-1.87794	0.47319	-1.46921	-0.87191
C8	-0.95870	-2.99247	-0.29576	0.83512	-1.39938	-1.28509	-1.44532	-2.06608
C9	-4.23372	-5.76230	-1.30645	-3.97856	-4.93318	-4.85310	-3.75607	-7.44977
C10	0.73672	2.16608	1.71422	-0.11956	1.32575	2.28866	1.74703	2.68257
C11	-0.90843	-0.51321	0.06944	-2.67855	-2.69316	-2.01668	-1.73071	1.08225
C12	1.54069	2.72345	-4.56388	-2.53049	-5.55330	-2.64194	0.53878	2.03048
C13	-5.49376	-6.30792	-2.38615	-2.14175	4.38634	-0.03381	-2.06993	-5.17029
C14	-3.30107	-2.40404	-5.13204	-2.38473	-3.79996	-4.33330	-4.76823	-2.41632
C15	-0.19891	-1.42895	-2.26381	-0.48135	0.35309	-2.42112	-1.92578	-0.09665
C16	-2.04687	-3.46579	-4.06512	-1.73548	-4.64956	-1.59132	-3.89432	-1.80314

TABLE 16.—CAPTURE PROBABILITY STRUCTURE USED TO SIMULATE GOOD, MEDIUM, AND POOR DATA TO STUDY THE OPERATING CHARACTERISTICS OF THE MODEL SELECTION PROCEDURE. A POPULATION OF 400 WAS TRAPPED 7 OCCASIONS FOR 200 CASES

Model	Probability Structure
GOOD DATA (Average $p \approx 0.35$)	
M_o	$p = 0.35$.
M_h	$p_i = 0.5, i = 1,100; p_i = 0.4, i = 101,200; p_i = 0.3, i = 201,300; p_i = 0.2, i = 301,400$.
M_b	$p = 0.5; c = 0.2$.
M_{bh}	$p_i = 0.65, c_i = 0.10, i = 1,100; p_i = 0.55, c_i = 0.15, i = 101,200; p_i = 0.45, c_i = 0.10, i = 201,300; p_i = 0.35, c_i = 0.50, i = 301,400$.
M_t	$p_1 = 0.35, p_2 = 0.45, p_3 = 0.25, p_4 = 0.35, p_5 = 0.20, p_6 = 0.50, p_7 = 0.35$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.9, i = 1,100; p_i = 0.8, i = 101,200; p_i = 0.7, i = 201,300; p_i = 0.6, i = 301,400; p_1 = 0.5; p_2 = 0.6; p_3 = 0.4; p_4 = 0.5; p_5 = 0.35; p_6 = 0.65; p_7 = 0.5$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 1.5; p_1 = 0.3; p_2 = 0.45; p_3 = 0.15; p_4 = 0.3; p_5 = 0.2; p_6 = 0.4; p_7 = 0.3$.
M_{tth}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.5, i = 1,100; p_i = 0.4, i = 101,200; p_i = 0.3, i = 201,300; p_i = 0.2, i = 301,400; c = 1.5; p_1 = 0.7; p_2 = 0.7; p_3 = 0.4; p_4 = 0.4; p_5 = 0.7; p_6 = 0.9; p_7 = 0.87$.
MEDIUM DATA (Average $p \approx 0.20$)	
M_o	$p = 0.20$.
M_h	$p_i = 0.3, i = 1,100; p_i = 0.25, i = 101,200; p_i = 0.15, i = 201,300; p_i = 0.10, i = 301,400$.
M_b	$p = 0.15; c = 0.03$.
M_{bh}	$p_i = 0.35, c_i = 0.05, i = 1,100; p_i = 0.4, c_i = 0.1, i = 101,200; p_i = 0.2, c_i = 0.05, i = 201,300; p_i = 0.25, c_i = 0.2, i = 301,400$.
M_t	$p_1 = 0.02; p_2 = 0.35; p_3 = 0.05; p_4 = 0.2; p_5 = 0.1; p_6 = 0.3; p_7 = 0.2$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.7, i = 1,100; p_i = 0.55, i = 101,200; p_i = 0.45, i = 201,300; p_i = 0.3, i = 301,400; p_1 = 0.4; p_2 = 0.55; p_3 = 0.25; p_4 = 0.4; p_5 = 0.3; p_6 = 0.5; p_7 = 0.4$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 0.33; p_1 = 0.30; p_2 = 0.45; p_3 = 0.15; p_4 = 0.30; p_5 = 0.20; p_6 = 0.40; p_7 = 0.30$.
M_{tth}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.3, i = 1,100; p_i = 0.25, i = 101,200; p_i = 0.15, i = 201,300; p_i = 0.10, i = 301,400; c = 2.00; p_1 = 0.50; p_2 = 0.65; p_3 = 0.35; p_4 = 0.50; p_5 = 0.40; p_6 = 0.70; p_7 = 0.50$.
POOR DATA (Average $p \approx 0.05$)	
M_o	$p = 0.05$.
M_h	$p_i = 0.09, i = 1,100; p_i = 0.05, i = 101,300; p_i = 0.10, i = 301,400$.
M_b	$p = 0.075; c = 0.01$.
M_{bh}	$p_i = 0.07, c_i = 0.03, i = 1,100; p_i = 0.03, c_i = 0.08, i = 101,300; p_i = 0.01, c_i = 0.09, i = 301,400$.
M_t	$p_1 = 0.05; p_2 = 0.01; p_3 = 0.09; p_4 = 0.05; p_5 = 0.07; p_6 = 0.03; p_7 = 0.05$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.20, i = 1,100; p_i = 0.25, i = 101,200; p_i = 0.05, i = 201,300; p_i = 0.10, i = 301,400; p_1 = 0.33; p_2 = 0.10; p_3 = 0.10; p_4 = 0.333; p_5 = 0.50; p_6 = 0.50; p_7 = 0.33$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 0.10; p_1 = 0.10; p_2 = 0.15; p_3 = 0.05; p_4 = 0.10; p_5 = 0.19; p_6 = 0.01; p_7 = 0.10$.
M_{tth}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.20, i = 1,100; p_i = 0.25, i = 101,200; p_i = 0.05, i = 201,300; p_i = 0.10, i = 301,400; c = 2.00; p_1 = 0.20; p_2 = 0.15; p_3 = 0.20; p_4 = 0.05; p_5 = 0.05; p_6 = 0.15; p_7 = 0.37$.

given in Table 15. The 8 C_i values are then standardized over the interval 0 to 1 to give the model selection criteria. These standardized "model selection criteria" are printed by program CAPTURE right after the 7 tests (cf. Fig. 13).

Given the classification functions, an evaluation of their usefulness must be made. Again, data are required where the underlying models are known; this is only possible with simulated data. Hence, we simulated 3 additional data sets: 200 replications each of good, medium, and poor data for each model. In all cases, we used

$N = 400$ and $t = 7$. The good data had an approximate overall average capture probability \bar{p} of 0.35, medium \bar{p} was approximately 0.20, and poor \bar{p} was approximately 0.05. The capture probability structure of these data is given in Table 16.

We analyzed each data set and selected a model based on the classification function described above. This allows an evaluation of the selection procedure, i.e., how often a wrong model is chosen, and which models tend to be confused. These results are given in Tables 17–19

TABLE 17.—PERFORMANCE OF THE MODEL SELECTION PROCEDURE WITH GOOD DATA. THE TRUE MODEL FROM WHICH THE DATA WERE GENERATED IS AT THE TOP, AND THE MODEL SELECTED BY THE CLASSIFICATION FUNCTION ON THE SIDE. THE ENTRIES ARE THE PERCENTAGE OF TIMES THE MODEL WAS SELECTED, THE $Ave(\hat{N})$, AND THE 95% CONFIDENCE INTERVAL COVERAGE

Model selected	Data generated from model								All data
	M_o	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbh}	
M_o	70.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0	9.1
	399.7	382.2	—	—	—	—	—	—	399.1
	0.957	0.200	—	—	—	—	—	—	0.931
M_h	21.5	93.5	0.0	0.0	0.0	0.0	0.0	0.0	14.4
	414.0	427.4	—	—	—	—	—	—	424.9
	0.349	0.059	—	—	—	—	—	—	0.113
M_b	0.0	0.0	90.0	0.0	0.0	0.0	0.0	0.0	11.3
	—	—	399.7	—	—	—	—	—	399.7
	—	—	0.922	—	—	—	—	—	0.922
M_{bh}	6.5	1.5	0.5	93.0	0.0	0.0	0.0	0.0	12.7
	394.2	396.3	399.0	396.6	—	—	—	—	396.4
	0.692	1.00	1.00	0.575	—	—	—	—	0.591
M_t	1.0	0.0	0.0	0.0	82.0	58.5	2.0	0.0	17.9
	404.0	—	—	—	399.8	396.3	372.0	—	398.0
	1.00	—	—	—	0.951	0.829	0.0	—	0.888
M_{th}	0.0	1.0	0.0	0.0	6.5	39.5	0.5	0.0	5.9
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
M_{tb}	0.0	0.5	9.5	0.0	11.0	2.0	95.0	37.0	19.4
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
M_{tbh}	1.0	1.0	0.0	7.0	0.5	0.0	2.5	63.0	9.4
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
Mean population estimate	402.5	425.8	399.7	396.6	399.8	396.3	372.0	—	
Coverage	0.808	0.077	0.923	0.575	0.951	0.829	0.0	—	

for the 3 types of data. However, our interest in these simulated data does not stop with the results of the model selection. Even more informative is how well the estimator performed when the wrong model was selected (we know that estimator performance usually is acceptable when the correct model is used). Information on the $Ave(\hat{N})$ and confidence interval coverage is also given in Tables 17–19 for all estimators. The tables also provide summaries by model selection and by data type.

Those tables again emphasize the importance of high capture probabilities. The estimates and coverage coefficients drastically decline in usefulness as the capture probabilities decline. In addition,

the correct model is much less likely to be selected for poor data. Note that in Table 19 (the poor data case), Model M_o is selected 34.7 percent of the time (whereas it is the true model only 12.5% of the time), and that $Ave(\hat{N}_o)$ is 745.5.

For good data, such a discrimination procedure will do quite well in selecting the appropriate model. We suggest that users knowledgeable in statistics and experienced in analysis of capture data may sometimes be able to render an improved judgement about the appropriate model for real data. However, the field biologist should probably accept the model recommended by the model selection algorithm. The primary reason for this is the problems with dependence among tests

TABLE 18.—PERFORMANCE OF THE MODEL SELECTION PROCEDURE WITH MEDIUM DATA. THE TRUE MODEL FROM WHICH THE DATA WERE GENERATED IS AT THE TOP, AND THE MODEL SELECTED BY THE CLASSIFICATION FUNCTION ON THE SIDE. THE ENTRIES ARE THE PERCENTAGE OF TIMES THE MODEL WAS SELECTED, THE $Ave(\hat{N})$, AND THE 95% CONFIDENCE INTERVAL COVERAGE

Model selected	Data generated from model								All data
	M_o	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbb}	
M_o	68.0	28.0	0.0	0.0	0.0	0.0	0.0	0.5	12.1
	400.8	361.1	—	—	—	—	—	234.0	388.4
	0.971	0.089	—	—	—	—	—	0.000	0.710
M_h	18.5	66.5	0.5	0.0	0.0	0.0	0.0	0.0	10.7
	484.7	449.7	890.4	—	—	—	—	—	459.9
	0.000	0.399	0.000	—	—	—	—	—	0.310
M_b	1.0	0.0	81.0	58.0	0.0	0.0	0.0	1.5	17.7
	427.0	—	397.2	388.4	—	—	—	783.0	397.9
	0.500	—	0.951	0.681	—	—	—	1.000	0.837
M_{bh}	7.5	3.0	9.0	32.0	0.0	0.0	0.0	3.5	6.9
	409.3	382.0	471.8	390.1	—	—	—	467.1	410.6
	0.667	0.333	0.949	0.688	—	—	—	1.000	0.727
M_t	1.0	0.0	0.0	0.0	57.5	67.5	0.0	1.5	15.9
	405.0	—	—	—	400.3	377.9	—	251.3	386.7
	1.000	—	—	—	0.974	0.593	—	0.000	0.761
M_{th}	3.0	2.0	0.0	0.0	10.5	25.5	0.0	11.5	6.6
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
M_{tb}	0.0	0.0	8.5	8.5	31.5	6.5	98.0	65.0	27.2
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
M_{tbb}	1.0	0.5	1.0	1.5	0.5	0.5	2.0	16.5	2.9
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
Mean population estimate	418.0	422.2	407.4	389.0	400.3	377.9	—	471.9	
Coverage	0.755	0.308	0.945	0.683	0.974	0.593	—	0.714	

and the fact that with real data it is highly unlikely that any of these 8 models will be exactly “true.”

Estimation in Alternative Models

When the model selection procedure described above has classified a model as the best one for a given set of data, 2 problems may still arise. First, the model may not have an associated estimation procedure for estimating N . Second, the model with the largest selection value (always 1.0) may not really fit the data, even though it is the “best” model available. In the first case, the investigator should scan for a large selection value (say 0.90) corresponding to a model hav-

ing an estimator. If such a model is found, and if the relevant tests of model assumptions suggest that the goodness of fit of the model is adequate, the corresponding estimator can be used with some confidence. We would caution, however, against using models with selection values less than 0.75 to produce estimates of N , especially if there is a poor fit of the model to the data. In the second case, none of the models with corresponding estimation procedures seems to fit the data well; where fit is judged by the model selection procedure and inspection of individual tests, then we can see no justification for granting statistical validity to any calculated population estimate.

TABLE 19.—PERFORMANCE OF THE MODEL SELECTION PROCEDURE WITH POOR DATA. THE TRUE MODEL FROM WHICH THE DATA WERE GENERATED IS AT THE TOP, AND THE MODEL SELECTED BY THE CLASSIFICATION FUNCTION ON THE SIDE. THE ENTRIES ARE THE PERCENTAGE OF TIMES THE MODEL WAS SELECTED, THE AVE(\hat{N}), AND THE 95% CONFIDENCE INTERVAL COVERAGE

Model selected	Data generated from model								All data
	M_o	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbh}	
M_o	79.0	74.0	39.5	83.0	0.0	1.0	0.0	1.0	34.7
	417.5	316.3	3,229.7	270.2	—	342.5	—	146.5	745.5
	0.949	0.547	0.557	0.386	—	0.500	—	0.000	0.615
M_h	13.0	18.5	4.5	11.0	1.0	9.0	0.0	5.0	7.8
	377.7	336.5	612.2	255.6	374.6	347.1	—	210.5	342.8
	0.923	0.514	0.000	0.000	1.000	0.611	—	0.000	0.452
M_b	0.5	1.0	14.0	0.0	0.0	0.0	0.0	0.0	1.9
	158.0	156.0	272.8	—	—	—	—	—	261.6
	0.000	0.000	0.250	—	—	—	—	—	0.226
M_{bh}	3.0	4.0	33.0	2.0	0.0	0.0	0.0	0.0	5.3
	174.7	172.1	359.5	123.3	—	—	—	—	317.2
	0.000	0.000	0.924	0.000	—	—	—	—	0.726
M_t	1.0	0.5	0.0	1.5	83.5	24.5	2.0	68.0	22.6
	366.5	246.0	—	216.7	406.7	325.9	2,983.0	164.5	331.0
	1.000	0.000	—	0.000	0.940	0.653	0.000	0.059	0.550
M_{th}	1.5	0.5	0.5	1.5	5.0	5.0	0.0	0.5	1.8
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
M_{tb}	0.5	0.0	4.5	0.0	1.0	0.0	92.5	0.0	12.3
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
M_{tbh}	1.5	1.5	4.0	1.0	9.5	60.5	5.5	25.5	13.6
	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—
Mean population estimate	402.7	312.2	1,604.5	264.7	406.3	331.9	2,983.0	167.4	
Coverage	0.912	0.510	0.615	0.328	0.941	0.638	0.000	0.054	

Additional Examples of Model Selection

As part of a population ecology study on salt marsh rodents, Coulombe (1965, unpublished master's thesis, University of California, Los Angeles, California), conducted a livetrapping study on an outbreak of feral house mice *Mus musculus* in a salt marsh in mid-December 1962, at Ballana Creek, Los Angeles County, California. A square 10×10 grid was used with traps spaced 3 m apart and trapping

was done twice daily, morning and evening, for 5 days. Thus there are 10 trapping occasions, but we can expect time variation in capture probabilities between morning and night occasions. The entire data set of 173 distinct individuals captured included young and adult, and male and female. Thus, we might also expect some heterogeneity of capture probabilities.

From the model selection procedure (Fig. 14), there is clear evidence of time

→

FIG. 14. Example of the model selection procedure applied to Coulombe's (unpublished thesis) full data set. Appropriate model probably is M_{th} . No estimator results from the model.

OCCASION	J=	1	2	3	4	5	6	7	8	9	10
ANIMALS CAUGHT	N(J)=	68	61	62	52	74	41	76	36	76	39
TOTAL CAUGHT	M(J)=	0	68	102	128	140	156	159	171	171	173
NEWLY CAUGHT	U(J)=	68	34	26	12	16	3	12	0	2	0
FREQUENCIES	F(J)=	2	64	40	31	16	13	5	1	0	1

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)

CHI-SQUARE VALUE = 48.576 DEGREES OF FREEDOM = 6 PROBABILITY OF LARGER VALUE = 0.00000

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)

CHI-SQUARE VALUE = 1.848 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = .17400

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)

CHI-SQUARE VALUE = 62.246 DEGREES OF FREEDOM = 9 PROBABILITY OF LARGER VALUE = 0.00000

4. GOODNESS OF FIT TEST OF MODEL M(H)

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)

CHI-SQUARE VALUE = 57.151 DEGREES OF FREEDOM = 9 PROBABILITY OF LARGER VALUE = 0.00000

TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
(FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

NUMBER OF CAPTURES	CHI-SQUARE	D.F.	PROBABILITY
2	14.027	9	.12136
3	6.857	9	.65199
4	27.387	9	.00121

5. GOODNESS OF FIT TEST OF MODEL M(B)

NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)

CHI-SQUARE VALUE = 68.087 DEGREES OF FREEDOM = 15 PROBABILITY OF LARGER VALUE = 0.00000

5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

CHI-SQUARE VALUE = 24.780 DEGREES OF FREEDOM = 7 PROBABILITY OF LARGER VALUE = .00083

5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

CHI-SQUARE VALUE = 43.307 DEGREES OF FREEDOM = 8 PROBABILITY OF LARGER VALUE = .00000

6. GOODNESS OF FIT TEST OF MODEL M(T)

NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)

CHI-SQUARE VALUE = 162.329 DEGREES OF FREEDOM = 125 PROBABILITY OF LARGER VALUE = .01386

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)

CHI-SQUARE VALUE = 31.439 DEGREES OF FREEDOM = 24 PROBABILITY OF LARGER VALUE = .14153

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL CRITERIA	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
	.15	.24	.03	.22	.26	1.00	0.00	.66

and heterogeneity variation, but little indication of behavioral variation. Note also that none of the simple models (M_h , M_b , M_t) fit the data. The model selection criteria strongly suggest Model M_{th} as appropriate, but there is no estimator associated with this model. To obtain an estimator, we can look for the next most likely model, which is M_{tbh} (selection criteria = 0.66); however, there is no estimator for that model either. In fact, to get a model allowing estimation we would have to use either M_t , M_h , or M_{bh} . Unfortunately, the value of the selection criteria corresponding to those models is too low to allow legitimate choice of one of the models for estimation purposes. Therefore, the investigator must realize that none of the models can be used to estimate population size with the data in the present form.

One alternative is to simply take the total number seen (M_{t+1}) as the best estimate. Given good data, M_{t+1} will be nearly equal to N . Also with good data one will tend to reject the simpler models. The real measure, however, of good data comes from such things as $t, n/M_{t+1}$, the pattern of the new captures (the u_j 's) and the apparent average capture probability (average \hat{p} is 0.3 here using $\hat{N} = 173 = M_{t+1}$). In good data, very few new animals will be caught by the last few occasions; in Coulombe's data, only 2 new animals were caught after the seventh occasion. Similarly, the capture frequency data should show many animals caught 2, 3, 4, or more times and there should not be a strong spike at f_1 (captured once only). By all these measures, Coulombe's data suggest that almost all the population was caught. Thus, here we would suggest taking 173 as the estimated population size.

Under those circumstances, we would also expect the point estimate of N from

all 5 possible estimators to be in close agreement with the value of 173. In fact the results for Coulombe's full data set are:

Estimator	Value	Standard error
\hat{N}_o	176	1.8
\hat{N}_t	175	1.6
\hat{N}_b	174	1.6
\hat{N}_h	175	1.8
\hat{N}_{bh}	173	0.2

These are all very similar and precise. Note that in general it is not appropriate to compute all estimates. We recommend it only when there is strong evidence all animals were captured, in which case it provides an additional check.

As a further example, Fig. 15 shows the model selection results using only the morning capture data from Coulombe's (unpublished thesis) study (i.e., pretending the evening captures never occurred). Presumably, this would eliminate most of the time variation, but not heterogeneity. The selection criteria verify this conjecture.

Model	Model selection criteria
M_o	0.99
M_h	1.00
M_b	0.58
M_{bh}	0.74
M_t	0.00
M_{th}	0.46
M_{tb}	0.53
M_{tbh}	0.80

When the criteria value for 2 or more models exceed 0.95, the program does not just suggest 1, but names the 2 models that have the highest criteria. Thus, in this case the choice between models M_o and M_h is not clear cut. As a standard operating procedure, we rec-

→

FIG. 15. Example of the model selection procedure applied to Coulombe's (unpublished thesis) morning trapping occasions (Occasions 1, 3, 5, 7, 9). Appropriate model probably is M_h or M_o . Suggested estimator is jackknife.

OCCASION	J=	1	2	3	4	5
ANIMALS CAUGHT	N(J)=	68	62	74	76	76
TOTAL CAUGHT	M(J)=	0	68	103	127	153
NEWLY CAUGHT	U(J)=	68	35	24	26	7
FREQUENCIES	F(J)=	45	59	36	15	5

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
 NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)

CHI-SQUARE VALUE = 5.051 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .16813

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
 NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)

CHI-SQUARE VALUE = 2.271 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = .13184

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
 NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)

CHI-SQUARE VALUE = 3.667 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .45295

4. GOODNESS OF FIT TEST OF MODEL M(H)
 NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)

CHI-SQUARE VALUE = 3.674 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .45191

TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
 (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

NUMBER OF CAPTURES	CHI-SQUARE	D.F.	PROBABILITY
1	3.333	4	.50367
2	4.362	4	.35928
3	3.074	4	.54551
4	2.667	4	.61506

5. GOODNESS OF FIT TEST OF MODEL M(B)
 NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)

CHI-SQUARE VALUE = 10.240 DEGREES OF FREEDOM = 6 PROBABILITY OF LARGER VALUE = .11492

5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

CHI-SQUARE VALUE = 7.735 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .05183

5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

CHI-SQUARE VALUE = 2.505 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .47440

6. GOODNESS OF FIT TEST OF MODEL M(T)
 NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)

CHI-SQUARE VALUE = 115.230 DEGREES OF FREEDOM = 101 PROBABILITY OF LARGER VALUE = .15766

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
 NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)

CHI-SQUARE VALUE = 16.497 DEGREES OF FREEDOM = 10 PROBABILITY OF LARGER VALUE = .08626

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL CRITERIA	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
	.99	1.00	.58	.74	0.00	.46	.53	.80

OVERALL TEST RESULTS --
 Z-VALUE -0.403
 PROBABILITY OF A SMALLER VALUE .34362

TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
 (FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES	Z-VALUE	PROBABILITY
2	-.660	.25477
3	-.322	.37360
4	-.066	.47356
5	.694	.75607
6	-1.085	.13906

FIG. 16. Example of the test for closure procedure with feral house mouse data from Coulombe (unpublished thesis).

commend using the jackknife estimator in such cases because it is robust to heterogeneity and will tend to do well even if a competing model is true. For the morning data, the estimate of N from the jackknife estimator was 194 ± 7.6 , while that from the null estimator was 171 ± 4.1 . Hence, both choices given by the model selection procedure provided similar estimates.

As a final check on whether time variation exists in the form of morning and evening differences, note that there were more captures in the mornings (occasions 1, 3, 5, 7, 9) than in the evening (occasions 2, 4, 6, 8, 10).

A Test for Closure

Throughout this monograph our philosophy has been that assumptions should be tested. One of the most critical assumptions behind this entire work is that of population closure. Although it is desirable to test closure, there are no truly suitable tests for this assumption. Closure is difficult to test for, because some types of variations in capture probabilities (especially behavioral) are difficult to distinguish from a failure of closure.

Burnham and Overton (pers. comm.) suggest a closure test based on Model M_h (cf. Appendix K). Fig. 16 gives an example produced by program CAPTURE of this closure test applied to Coulombe's (unpublished thesis) full data set. The

test statistic value (z-value) is -0.403 , and the significance level of the test is 0.34362 . Thus, for Coulombe's data, closure is not rejected.

If individual capture frequencies are 10 or more, program CAPTURE also computes and prints a partitioned version of the closure test for those frequencies. For example, from Fig. 16, for animals caught twice the test value is -0.660 and is not significant. These partitioned test cases are of interest, but we emphasize the overall test is the one to use for the final judgement on closure.

Failure of closure means that during the study animals are either entering or leaving the population at risk of capture, or both. This could be caused, for example, by death, emigration, or the trap grid itself attracting animals from surrounding areas (especially likely in removal trapping). In any of those cases, the animals that enter or leave have zero capture probabilities during the time they are not part of the trapped population. The corresponding probability model has similarities to our behavioral Models M_b , M_{bh} , M_{tb} , and M_{tth} . Indeed, as previously stated, behavioral responses are extremely difficult to untangle from true failures of closure. Our simulation results have shown the closure test rejects strongly when Model M_b is true, and is not a truly unbiased test whenever there is strong behavioral variation in capture probabilities. However, the test does not seem to be affected by heterogeneity or random time variations. We have not used other tests from the literature (e.g., Pollock et al. 1974) because, to our knowledge, those tests all are implicitly developed under the assumption that Model M_0 is the true model under the closure assumption. Because we feel Model M_0 probably is never acceptable, those tests will be even more untrustworthy than our current closure test.

The biologist is advised to look carefully at the closure test. If the test statistic is not significant, this tends to support the validity of the closure assumption. If that test rejects closure, before accepting that result, it is necessary to look at the other

tests and the indicated model. If strong behavioral variation is indicated, the closure test is not valid.

Additional evidence regarding closure is obtained from a test of average captures per trap as a function of trap distance from the edge of the grid. That test is discussed in the section on DENSITY ESTIMATION; it tests for whether the grid attracts animals.

Finally, we mention that the closure test used here is oriented toward detecting breakdowns in closure only during the initial and final stages of the experiment. The test is not appropriate, for instance, for identifying situations in which animals emigrate during the middle of the study period and then immigrate back to the study area in the latter stages of an experiment.

DENSITY ESTIMATION

Introduction

The models discussed to this point involve only population size N as the parameter of interest. Often, interest may lie in population density, the number of animals per unit area (e.g., squirrels/hectare). One could naively take \hat{N} divided by the area enclosed by the trapping grid as an estimate of density. That approach, however, leads to severe overestimation as a result of what has been called "edge effect," i.e., not all animals have their entire home range within the trapping grid, but may still be caught because some traps near the grid boundary are within their home range. Although biologists have recognized this problem for decades (Dice 1938, 1941; Stickel 1954), statisticians concerned with estimation of population abundance have tended to ignore or have failed to recognize the problem. This probably is due in part to the fact that abstract models for capture studies, such as ball-and-urn models, have no spatial component, hence do not include any concept of density as distinct from numbers of animals.

Three basic approaches are given in the biological literature to solve this

problem. Two of them rely on Dice's (1938) concept of a boundary strip about the grid such that the effective trapping area is the grid area plus this boundary strip area. Dice assumed the boundary strip to be one-half the average diameter of the home range of the population being trapped. The first 2 approaches attempt to estimate this strip width, W , from the capture data. These 3 approaches are given below:

- (1) Home range size is estimated from the locations of different captures for the same animal and is used to estimate the strip width W . A variety of approaches ad hoc and subject to numerous problems, e.g., results depend upon trap spacing and numbers of recaptures (Hayne 1949b, Stickel 1954, Tanaka 1972). The basic idea can be developed into an elaborate statistical estimation problem (Jennrich and Turner 1969), but as far as we know estimation of W based on movement data remains unsatisfactory.
- (2) The parameter W is directly estimated based on data drawn from selected subgrids (MacLulich 1951, Hansson 1969, Seber 1973:51, Smith et al. 1975). As suggested by Burnham and Cushwa (pers. comm.) that idea can be developed into a procedure allowing joint estimation of D and W from data on 1 sufficiently large grid. We discuss this approach in detail below.
- (3) The use of "assessment" lines is the most complex approach to density estimation. It involves designing the study to specifically estimate the effective trapping area as well as the size of the population at risk of capture. There are numerous variants on this approach (Smith et al. 1971, 1972, 1975, Swift and Steinhorst 1976, O'Farrell et al. 1977). We have not pursued this approach here because the proper data analysis depends upon the study design.

Of those 3 approaches, only the second seems to be formulated in a rigorous sta-

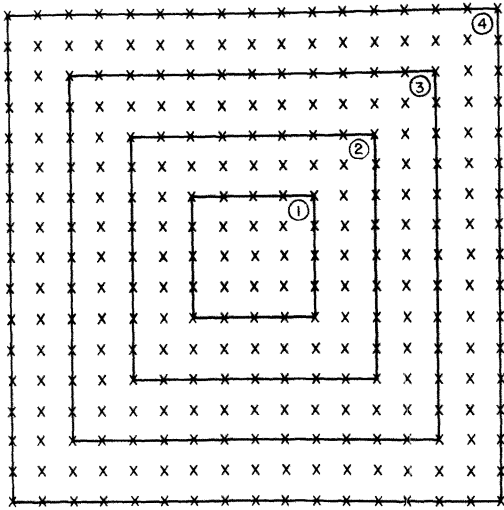


FIG. 17. A 17 × 17 trapping grid with 4 nested subgrids.

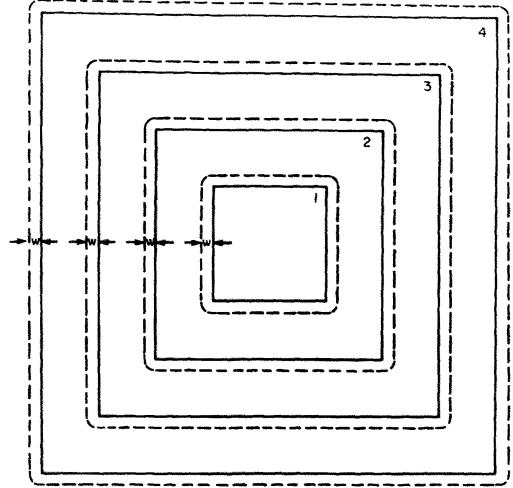


FIG. 18. A series of 4 nested subgrids with a boundary strip of width W around each subgrid. Note the corner of the boundary strip is a quarter circle of radius W .

tistical fashion without arbitrary assumptions. However, even the second approach cannot be made workable without some arbitrary ad hoc features.

Problem Formulation

We will illustrate the problem with some data from a livetrapping study of feral house mice *Mus musculus* in a southern California coastal salt marsh in December 1962 by Coulombe (unpublished thesis), that were also used to illustrate the model selection procedure. Only the basic specifications of the study will be needed here: 100 Sherman live traps were laid out in a 10 × 10 square grid. Morning and evening trapping sessions were conducted for 5 days and provided 10 trapping occasions, although only the morning sessions will be used

in the calculations. Traps were 3 m apart and the area enclosed by the grid was 0.073 ha. We can delineate subgrids of smaller size by examining smaller groups of traps; in particular, we can conceptualize a series of nested subgrids (see for example Fig. 17).

The size of the population at risk of capture for each grid must be estimated. The choice of a population estimator should be based on the model selection procedures using the data for the entire grid. For $i = 1, 2, \dots, k$, assume that a population estimate N_i has been made for the i^{th} grid by using data obtained only from that grid. The naive estimator of density of the i^{th} grid is given by

$$\hat{Y}_i = \hat{N}_i / A_i$$

$$i = 1, 2, \dots, k,$$

TABLE 20.—THE INFORMATION NEEDED TO ESTIMATE DENSITY USING THE FOUR NESTED SUBGRIDS AND THE DATA FROM COULOMBE (UNPUBLISHED THESIS)

No.	Grid Size	Area A_i (ha)	\hat{N}_i (individuals)	\hat{Y}_i ($\frac{\text{individuals}}{\text{ha}}$)	$Se(\hat{Y}_i)$
1	4 × 4 inner grid	0.0081	47	5770.9	607.8
2	6 × 6 middle inner grid	0.0225	107	4750.7	452.6
3	8 × 8 middle outer grid	0.0441	166	3768.5	243.4
4	10 × 10 entire grid	0.0729	194	2654.4	104.9

where A_i is the size (area) of the i^{th} trapping grid. The grid data are given in Table 20. Note that the standard error of \hat{Y}_i is $\hat{S}e(\hat{N}_i)/A_i$.

The four \hat{Y}_i values in Table 20 cannot reasonably be considered each an unbiased estimate of a single parameter. The bias is attributed to the phenomenon of edge effect, wherein the area used by individuals at risk of capture is not restricted to the area contained within the trapping grid; rather, there are areas outside the trapping grid that are part of the home ranges of individuals at risk of capture.

We assume that the population density is constant in the area of trapping. That is, there is not a trend in density across the grid. Under these conditions it is reasonable to think that there is an area of constant width about the trapping grid such that the actual area used by the population at risk of capture is the total area of the trapping grid plus that of the boundary strip. Let W equal the width of this strip. This is illustrated for the four grids in Fig. 18. The concept of a boundary strip goes back at least as far as Dice (1938). We need not interpret W , but from the inception of the boundary strip concept, biologists have considered that W is related to home range size (Dice 1938, 1941). In fact they arbitrarily assume W is one-half the maximum linear dimension of the average home range of the species.

Statistical Treatment

The approach to density estimation that we advocate for use with grid trapping is to formulate the problem as one of joint estimation of D and W , with the parameter N being a function of D and W . Then, by having 2 or more grids of different sizes, we can estimate those parameters with a weighted nonlinear least squares procedure.

In what follows, we assume that density (D) is expressed in individuals per unit area and strip width (W) in linear units, such as meters. Consider a layout

of traps in the shape of a square or rectangle. That is, by connecting the outermost traps, a square or rectangle results. Let P be the measured perimeter of the grid. Let A be the area within this perimeter, and let $A(W)$ equal the area obtained by adding a boundary strip of constant width W . Then the fundamental relationships is

$$A(W) = A + PW/c + \pi W^2/c,$$

where c is a conversion factor to express PW or W^2 in units of A . For example, to convert m^2 to ha , $c = 1,000 m^2/ha$.

Let there be k such grids identified. These may be subgrids of one grid, or may be physically separate study grids. Assume a constant density applies for each grid. Then, for grid i we would expect to have

$$N_i = D A_i(W) = D[A_i + P_i W/c + \pi W^2/c],$$

where N_i is the population at risk of capture with respect to the i^{th} grid of traps only. Conceptually, we should let the actual numbers of individuals at risk of capture on grid i be a random variable with expectation $DA_i(W)$. Then we would write $E(\hat{N}_i) = DA_i(W)$, and it is really $E(\hat{N}_i)$ we are estimating for the i^{th} grid. Letting $i = 1, \dots, k$ we obtain structural equations relating the parameters D , W , and the induced parameters N_i . The area A_i , and the grid perimeter P_i must be known. Next we redefine the basic structural equations as

$$Y_i = \frac{N_i}{A_i} = D[1 + a_i W + b_i W^2] \quad i = 1, 2, \dots, k,$$

where

$$a_i = \frac{P_i}{A_i c} \text{ and } b_i = \frac{\pi}{A_i c}.$$

Assume that from the trapping data of each grid we have estimates of the N_i , expressed as \hat{N}_i , and estimates of their variances $\text{Var}(\hat{N}_i)$. Assume further that the \hat{N}_i are good estimates in the sense that they have small bias, so we can write

$$\frac{\hat{N}_i}{A_i} = \hat{Y}_i = D[1 + a_i W + b_i W^2] + \epsilon_i,$$

$$i = 1, 2, \dots, k,$$

where ϵ_i is a random error with $E(\epsilon) = 0$, and covariance matrix $E(\underline{\epsilon} \underline{\epsilon}') = \Sigma$. The small bias of the \hat{N}_i is a big assumption that relies heavily on previous sections. Note that $\hat{Y}_i = \hat{N}_i/A_i$ is what we would call the naive estimator of density, D , from the i^{th} trapping grid; clearly it can be significantly biased unless A_i is large relative to $(P_i W/c + \pi W^2/c)$, the area that is added by assuming a strip width of W . Note that for a sequence of k subgrids of increasing size, the most biased estimate of D is obtained from \hat{Y}_1 because A_1 is smallest relative to $(P_1 W/c + \pi W^2/c)$. On the other hand, \hat{Y}_k is the best estimate of D because A_k is largest relative to $(P_k W/c + \pi W^2/c)$. If we had a very large grid, say the state of Wyoming, the contribution of area due to the boundary strip W would be negligible, and we could ignore the problem. Because this is not the case, we use our biased estimates of Y_i to find an unbiased estimate of D .

We assume $E(\underline{\epsilon} \underline{\epsilon}') = \Sigma$ because in general the naive density estimates, \hat{Y}_i , are not independent from one another if these subgrids derive from 1 overall grid. Hence, we must estimate the covariance matrix Σ of the k estimates of \hat{Y}_i . A simple and intuitive procedure is to assume the correlation between \hat{Y}_i and \hat{Y}_j is just the proportion of overlapping area between grid i and grid j including their boundary strip. Then the covariance of \hat{Y}_i and \hat{Y}_j is $Se(\hat{Y}_i) \cdot Se(\hat{Y}_j) \cdot Corr(\hat{Y}_i, \hat{Y}_j)$. With the resulting covariance matrix, generalized nonlinear least squares can be performed to estimate density and strip width as \hat{D} and \hat{W} , respectively. Note that Σ is a function of W , because the amount of overlap between grid i and grid j is a function of the estimated strip width W .

In the most common situation, there is one grid such as our example in the California salt marsh, and one must choose a small number of subgrids to use in this approach. We recommend 4 or more nested grids. The main concern is to keep the

inner grid large enough so that N_1 is well estimated. Other possibilities (not recommended) are to include subgrids consisting of halves and quarters of the total grid, or both.

The basic idea behind this procedure was introduced by MacLulich (1951), but until recently no method of estimating the variances of \hat{D} and \hat{W} had been given (Smith et al. 1975).

In Fig. 19, a complete analysis is presented for the California salt marsh feral mouse study. First note that we have computed the matrix of captures per trap station. That matrix includes the data for both the morning and evening trapping sessions, or for all 10 occasions. Multiple captures per trap cause some entries to be greater than 10. Visual study of this matrix (Fig. 19a) does not disclose any gross trends in mouse density across the grid. Fig. 19b shows the results of 3 chi-square goodness of fit tests, where the capture matrix is first collapsed by rows, then by columns, and finally by rings. Those tests generally tend to reject the null hypothesis of uniform probability of capture by rows or by columns. However, we are more concerned in identifying gross irregularities in mouse density, such as no captures in one corner of the grid, or a strong trend of decreasing probability of capture from right to left. The tests identify a problem of higher trap success at the grid edges. This is indicated by the test of rows, where a large portion of the chi-square value is contributed by the ninth and tenth rows, and by the ring test, where the outside ring of traps had much better success than expected, and the inner ring poorer success than expected. For large trapping grids, the problem could be corrected by not using the data for the outer ring of traps. This is very wasteful of data, however, and we will ignore the problem here.

The next 4 pages of output (Figs. 19c-f) consist of the population estimates for the whole grid and each of the 4 subgrids, computed with the jackknife estimator derived from Model M_h . That model was selected based on the output for the tests

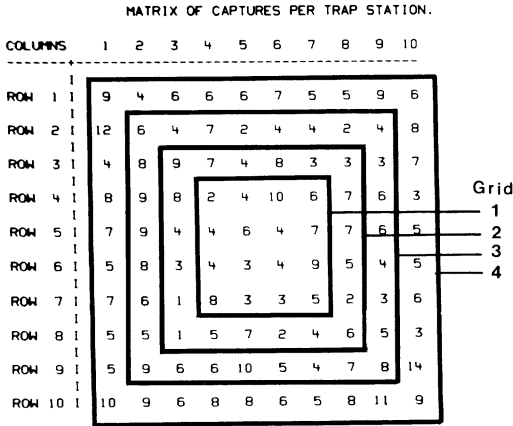


FIG. 19a. Nested subgrids used in the density estimation procedure with feral house mouse data from Coulombe (unpublished thesis). In the matrix, trap coordinates are rounded to the nearest whole integer. In goodness of fit tests, trap coordinates that are not integers and nonrectangular trapping grids will cause spurious results.

mates of D and W . The columns headed NAIVE DENSITY, PERIMETER/AREA, and $PI/AREA$ correspond to \hat{Y}_i , a_i , and b_i , respectively, and the COVARIANCE MATRIX corresponds to $\hat{\Sigma}$ for the initial value of W . The number of iterations required to estimate the parameters is given, and for this example it is 189. The maximum number allowed in program CAPTURE is 300. The approximate number of significant digits of \hat{D} and \hat{W} is also printed. A table listing the grid number, the naive density, and the predicted value is then given. This allows the investigator to get a feel for how well the model fits his data. Note that all the residuals in the example are negative; this is because they are all highly correlated, and hence we expect them to be of the same sign. The multiple correlation coefficient, R , is also printed to help assess the fit of the model. The value of R^2 is interpreted as the proportion of the variance in the data that is explained by the model ($0 \leq R^2 \leq 1$). The computed values of \hat{D} and \hat{W} are highly negatively correlated, indicating the

described in the section on model selection (Fig. 15).

The final page of output (Fig. 19g) gives the generalized least squares esti-

CHI-SQUARE TEST OF UNIFORM DENSITY BY ROWS.

ROW	1	2	3	4	5	6	7	8	9	10
OBSERVED	63	53	56	63	59	50	44	43	74	80
EXPECTED	58.500	58.500	58.500	58.500	58.500	58.500	58.500	58.500	58.500	58.500
CHI-SQUARE	.346	.517	.107	.346	.004	1.235	3.594	4.107	4.107	7.902

TOTAL CHI-SQUARE = 22.26 WITH 9 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .0081

CHI-SQUARE TEST OF UNIFORM DENSITY BY COLUMNS.

COLUMN	1	2	3	4	5	6	7	8	9	10
OBSERVED	72	73	48	57	53	53	52	52	59	66
EXPECTED	58.500	58.500	58.500	58.500	58.500	58.500	58.500	58.500	58.500	58.500
CHI-SQUARE	3.115	3.594	1.885	.038	.517	.517	.722	.722	.004	.962

TOTAL CHI-SQUARE = 12.08 WITH 9 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .2090

CHI-SQUARE TEST OF UNIFORM DENSITY BY RINGS (OUTER RING IS NUMBER 1).

RING	1	2	3	4	5
OBSERVED	247	160	96	65	17
EXPECTED	210.600	163.800	117.000	70.200	23.400
CHI-SQUARE	6.291	.088	3.769	.385	1.750

TOTAL CHI-SQUARE = 12.28 WITH 4 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .0154

FIG. 19b. Chi-square tests of uniform density with feral house mouse data from Coulombe (unpublished thesis).

NUMBER OF TRAPPING OCCASIONS HAS 5
 NUMBER OF ANIMALS CAPTURED, N(I+1), HAS 33
 TOTAL NUMBER OF CAPTURES, N., HAS 50

FREQUENCIES OF CAPTURE, F(I)
 I= 1 2 3 4 5
 F(I)= 18 13 2 0 0

COMPUTED JACKKNIFE COEFFICIENTS

	N(1)	N(2)	N(3)	N(4)	N(5)
1	1.800	2.400	2.800	3.000	3.000
2	1.000	.550	.050	-.250	-.250
3	1.000	1.000	1.133	1.250	1.250
4	1.000	1.000	.992	.992	.992
5	1.000	1.000	1.000	1.000	1.000

THE RESULTS OF THE JACKKNIFE COMPUTATIONS

I	N(I)	SE(I)	.95 CONF. LIMITS	TEST OF N(I+1) VS. N(I) CHI-SQUARE(1 D.F.)
0	33			
1	47.4	5.09	37.4 57.4	2.839
2	52.4	7.57	37.5 67.2	.148
3	53.3	9.51	34.7 72.0	.002
4	53.3	10.62	32.4 74.1	0.000
5	53.3	10.62	32.4 74.1	0.000

AVERAGE P-HAT = .2128

INTERPOLATED POPULATION ESTIMATE IS 47 WITH STANDARD ERROR 4.9234
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 37 TO 57

HISTOGRAM OF F(I)

FREQUENCY	18	13	2	0	0
EACH * EQUALS					
2 POINTS					
18	*				
16	*				
14	*	*			
12	*	*			
10	*	*			
8	*	*			
6	*	*			
4	*	*			
2	*	*	*		

FIG. 19c. Example of population estimation with variable probability of capture by animal under Model M_h with feral house mouse data (Coulombe (unpublished thesis) from inner inner grid of Fig. 19a ($X = 4-7, Y = 4-7$)).

strong inverse relation between density and strip width. This negative correlation makes it difficult to estimate either parameter with a small standard error, because other combinations of \hat{D} and \hat{W} also result in almost as good a fit as the values selected. Finally, a test of whether \hat{W} is significantly different from zero is printed. In that example, the highly significant difference of \hat{W} from zero indicates that the estimate of D is much better than the naive estimate based on the actual grid area.

Discussion

The density estimation problem results because the grid is an artificial entity in the environment, and animals trapped use areas both inside and outside the grid. A practical problem illustrated in the example is that the grid can attract

NUMBER OF TRAPPING OCCASIONS HAS 5
 NUMBER OF ANIMALS CAPTURED, N(I+1), HAS 67
 TOTAL NUMBER OF CAPTURES, N., HAS 110

FREQUENCIES OF CAPTURE, F(I)
 I= 1 2 3 4 5
 F(I)= 37 21 5 4 0

COMPUTED JACKKNIFE COEFFICIENTS

	N(1)	N(2)	N(3)	N(4)	N(5)
1	1.800	2.400	2.800	3.000	3.000
2	1.000	.550	.050	-.250	-.250
3	1.000	1.000	1.133	1.250	1.250
4	1.000	1.000	.992	.992	.992
5	1.000	1.000	1.000	1.000	1.000

THE RESULTS OF THE JACKKNIFE COMPUTATIONS

I	N(I)	SE(I)	.95 CONF. LIMITS	TEST OF N(I+1) VS. N(I) CHI-SQUARE(1 D.F.)
0	67			
1	96.6	7.30	82.3 110.9	10.573
2	109.4	10.91	88.0 130.7	2.231
3	114.3	13.65	87.6 141.1	.769
4	116.0	15.17	86.2 145.7	0.000
5	116.0	15.17	86.2 145.7	0.000

AVERAGE P-HAT = .2056

INTERPOLATED POPULATION ESTIMATE IS 107 WITH STANDARD ERROR 10.1827
 APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 86 TO 127

HISTOGRAM OF F(I)

FREQUENCY	37	21	5	4	0
EACH * EQUALS					
4 POINTS					
36	*				
32	*				
28	*	*			
24	*	*			
20	*	*			
16	*	*			
12	*	*			
8	*	*	*		
4	*	*	*	*	

FIG. 19d. Example of population estimation with variable probability of capture by animal under Model M_h with feral house mouse data (Coulombe unpublished thesis) from the middle inner grid of Fig. 19a ($X = 3-8, Y = 3-8$)).

animals, thus "inducing" a higher density than would otherwise be found. The problem of the grid attracting animals can cause severe bias. If attraction occurs during trapping, then the assumption of population closure will be violated. With removal trapping, over a long enough time the problem is sure to develop. The problem may still occur in livetrapping studies. One possible test for this effect is the "ring" test given in Fig. 19b. Even with some approach like prebaiting to allow this attraction effect to stabilize before trapping, both methods (recapture movements, and direct estimation of W and D) will be invalid if the grid itself attracts animals. In that case, it appears assessment lines would have to be used.

If the above method produces a poor result, an alternative approach is to base an estimate of W on animal movements as determined from recapture locations.

```

NUMBER OF TRAPPING OCCASIONS WAS      5
NUMBER OF ANIMALS CAPTURED, M(I+1), WAS 116
TOTAL NUMBER OF CAPTURES, N., WAS      202

FREQUENCIES OF CAPTURE, F(I)
I= 1 2 3 4 5
F(I)= 55 43 11 7 0

COMPUTED JACKKNIFE COEFFICIENTS

      N(1)  N(2)  N(3)  N(4)  N(5)
1  1.800  2.400  2.800  3.000  3.000
2  1.000  .550  .050  -.250  -.250
3  1.000  1.000  1.133  1.250  1.250
4  1.000  1.000  1.000  .992  .992
5  1.000  1.000  1.000  1.000  1.000

THE RESULTS OF THE JACKKNIFE COMPUTATIONS

      I  N(I)  SE(I)  .95 CONF. LIMITS  TEST OF N(I+1) VS. N(I)
      0  116  .000  .000  CHI-SQUARE(1 D.F.)
1  160.0  8.90  142.6  177.4  6.866
2  173.7  13.20  147.8  199.5  .195
3  175.6  16.64  143.0  208.2  .073
4  174.9  18.62  138.4  211.4  0.000
5  174.9  18.62  138.4  211.4  0.000

AVERAGE P-HAT = .2434

INTERPOLATED POPULATION ESTIMATE IS 166 WITH STANDARD ERROR 10.7335
APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 145 TO 188
    
```

```

HISTOGRAM OF F(I)
-----
FREQUENCY  55  43  11  7  0
EACH * EQUALS  5 POINTS
-----
54 *
48 *
42 * *
36 * *
30 * *
24 * *
18 * *
12 * *
6  * *
    
```

FIG. 19e. Example of population estimation with variable probability of capture by animal under Model M_n with feral house mouse data (Coulombe unpublished thesis) from the middle outer grid of Fig. 19a ($X = 2-9, Y = 2-9$).

Let \bar{W} be such an estimate, for example one-half of the average maximum distance between trapped locations for all animals captured at least twice. The standard error of \bar{W} , $S\hat{e}(\bar{W})$, can be computed from the data themselves. The estimator of density is then

$$\bar{D} = \frac{\hat{N}}{A(\bar{W})}$$

where \hat{N} is based on all the data (i.e., the entire grid) and

$$A(\bar{W}) = A_g[1 + a\bar{W} + b\bar{W}^2],$$

for a and b as defined previously in this section. An estimator of the sampling variance of \bar{D} is given by

$$\hat{V}ar(\bar{D}) = \frac{\hat{V}ar(\hat{N})}{[A(\bar{W})]^2} + (\bar{D})^2 \left(\frac{a + 2b\bar{W}}{1 + a\bar{W} + b\bar{W}^2} \right)^2 \hat{V}ar(\bar{W}),$$

```

NUMBER OF TRAPPING OCCASIONS WAS      5
NUMBER OF ANIMALS CAPTURED, M(I+1), WAS 150
TOTAL NUMBER OF CAPTURES, N., WAS      356

FREQUENCIES OF CAPTURE, F(I)
I= 1 2 3 4 5
F(I)= 45 59 36 15 5

COMPUTED JACKKNIFE COEFFICIENTS

      N(1)  N(2)  N(3)  N(4)  N(5)
1  1.800  2.400  2.800  3.000  3.000
2  1.000  .550  .050  -.250  -.250
3  1.000  1.000  1.133  1.250  1.250
4  1.000  1.000  1.000  .992  .992
5  1.000  1.000  1.000  1.000  1.000

THE RESULTS OF THE JACKKNIFE COMPUTATIONS

      I  N(I)  SE(I)  .95 CONF. LIMITS  TEST OF N(I+1) VS. N(I)
      0  150  .000  .000  CHI-SQUARE(1 D.F.)
1  196.0  8.05  180.2  211.8  .007
2  196.5  11.69  173.5  219.4  2.000
3  189.8  15.15  150.1  219.4  2.897
4  185.1  17.31  151.2  219.0  0.000
5  185.1  17.31  151.2  219.0  0.000

AVERAGE P-HAT = .3670

INTERPOLATED POPULATION ESTIMATE IS 194 WITH STANDARD ERROR 7.6455
APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL 178 TO 209
    
```

```

HISTOGRAM OF F(I)
-----
FREQUENCY  45  59  36  15  5
EACH * EQUALS  6 POINTS
-----
60 *
54 *
48 *
42 * *
36 * *
30 * *
24 * *
18 * *
12 * *
6  * *
    
```

FIG. 19f. Example of population estimation with variable probability of capture by animal under Model M_n with feral house mouse data (Coulombe unpublished thesis) from the entire grid of Fig. 19a ($X = 1-10, Y = 1-10$).

and the standard error of \bar{D} is simply

$$S\hat{e}(\bar{D}) = \sqrt{\hat{V}ar(\bar{D})}$$

These variance formulas are valid whatever the technique for estimating W from recapture locations.

In the example of the feral house mice, \bar{W} is calculated as 3.63 m with $S\hat{e}(\bar{W}) = 0.149$, and compares favorably with the estimated value of $\bar{W} = 4.65$ from Fig. 19g.

The methods outlined in this section require large amounts of data to achieve satisfactory results. Both a large trapping grid and a large number of captures are required. A grid size of 9×9 probably can be considered the minimum; however, a larger grid such as 15×15 is much better. Good trapping success to achieve a large number of captures is necessary to provide a useful population


```

STARTING VALUES FOR DENSITY ESTIMATION--
NUMBER OF GRIDS          4
TRAP INTERVAL            3.00
UNITS CONVERSION         10000.00
INITIAL DENSITY ESTIMATE 5089
INITIAL STRIP WIDTH ESTIMATE 882.6797

GRID  NAIVE DENSITY  PERIMETER/AREA  P1/AREA  STARTING COVARIANCE MATRIX
1      Y(1)          A(1)          B(1)
1  5770.927         44444444         .3879E+01  .369E+06
2  4750.710         26666667         .1396E+01  .273E+06 .205E+06
3  3768.521         1804762          .7124E+02  .148E+06 .108E+06 .592E+05
4  2854.402         1481481          4.359E+02  .625E+05 .468E+05 .294E+05 .110E+05

RESULTS OF ITERATIONS
FUNCTION EVALUATIONS REQUIRED 189
ESTIMATED SIGNIFICANT DIGITS OF PARAMETER VALUES 8

FITTED MODEL COMPARED TO THE DATA
GRID(1)  Y(1)  F(1)
1  5770.927  5505.869
2  4750.710  3983.134
3  3768.521  2675.007
4  2854.402  2511.652

MULTIPLE CORRELATION COEFFICIENT IS .93181

ESTIMATED DENSITY = 1408.934 172.1260 = 175 STANDARD ERROR
ESTIMATED STRIP WIDTH = 4.053 1.0576 = 175 STANDARD ERROR
CORRELATION OF ESTIMATORS = -.9471

TEST OF ESTIMATED STRIP WIDTH GREATER THAN ZERO.
Z-VALUE = 4.3998 PROBABILITY OF LARGER VALUE = .0000

FINAL COVARIANCE MATRIX
.369E+06
.156E+06 .204E+06
.5398E+05 .7085E+05 .5924E+05
.1621E+05 .2127E+05 .1779E+05 .1100E+05

```

FIG. 19g. Example of joint estimation of density and boundary strip width with feral house mouse data from Coulombe (unpublished thesis).

estimate for each grid. This becomes a problem especially with the smaller grids, where only a fraction of the data are used to estimate the population. Besides requiring a good population estimate, a good variance estimate for \hat{N} is required because that quantity is used in constructing the estimate of the weighting matrix $\hat{\Sigma}^{-1}$. Poor data result in a poor weighting matrix, which then results in poor estimates of D and W .

Possible methods to increase the amount of data are (1) to place traps close together to increase recaptures, (2) to place grids in a uniform habitat so that discontinuities in density do not occur, and (3) to increase the number of traps. To repeat, the method requires large amounts of data. A carefully designed study is required to obtain reliable values of \hat{D} and \hat{W} , and only rarely can a typical capture-recapture study be made to yield reasonable estimates.

STUDY DESIGN

The objective of this monograph is to present methods for the thorough analysis of capture data when the target populations are closed except for known re-

movals. However, proper planning, design, and field conduct of such studies is necessary to obtain meaningful data for analysis. Many factors must be considered when planning a capture-recapture or removal study to estimate animal abundance. We consider these as falling into 2 broad categories: statistical design and data recording considerations, and field procedures, although the distinction may sometimes seem a bit arbitrary. Examples of such considerations include (but are not limited to) the following:

Statistical considerations

- number of capture occasions
- time between occasions
- size and shape of trapping grid
- spacing of traps
- number of traps at a point
- numbering of traps

Field procedure considerations

- live vs. removal captures
- choice of trap type
- choice of electrofishing gear
- method of marking or tagging
- method of recapture
- use of bait
- time of day to check traps
- handling the animals

The purpose of this section is to present some suggestions and cautions regarding the aspects of study design. The general themes presented are to conduct the experiment so that (1) assumptions can be tested, (2) the closure assumption is met, (3) the simplest possible model is appropriate, and (4) the number of animals captured is maximized (including recaptures). We begin by mentioning livetrapping versus removal methods, then discuss closure, proceed through ways of eliminating variation due to time, behavior, and heterogeneity, and then discuss sample size considerations such as grid size and number of traps.

For additional discussion of design consideration in grid trapping the reader is referred to Overton and Davis (1969), Smith et al. (1969, 1971, 1975), Tanaka (1970), and Hansson (1974).

It is beyond the intended scope of this monograph to provide guidance on the operational aspects of capture studies, although they are important. For example, if the method of marking (or tagging) is such that marks are lost, then a basic assumption needed for meaningful results is violated. When making decisions about a field study, the scientist should consider the probable effects of the experimental design on assumptions necessary for data analysis. For additional discussion of field procedures the reader is referred to Davis (1956), Southwood (1966), and Taber and Cowan (1969).

Livetrapping Versus Removal Methods

As discussed in the previous sections, removal methods are a special case of livetrapping methods. That is, the removal estimators can be used on livetrapping data. Hence, we recommend that livetrapping methods should be used if possible because of the wider array of options available for the data analysis. A hazard of removal studies is that they disrupt the population, and as substantial animals are removed, immigration may occur; this violates closure.

For livetrapping studies, all possible precautions should be taken to prevent deaths of the animals while in the traps, e.g., shading the traps in summer, or avoiding periods of extreme cold. Similarly, it is assumed (implicitly) that the method of marking will not induce mortality. If substantial mortality cannot be avoided, then one must analyze the study as a removal experiment.

Closure

For the analysis methods presented here, the single most important assumption is closure. Closure is very difficult to test for, yet any violation of this assumption biases the tests and population estimators presented here.

Considerations to help assure closure include, for example, timing the trapping to avoid known migration times, and pe-

riods of recruitment (e.g., juveniles becoming trappable) or immigration. Also keep the duration of the experiment as short as possible. If it is necessary to study the population at such times, open population models should be used for data analysis (Seber 1973, Pollock 1975, Arnason and Baniuk 1977).

Grids are often thought to attract animals. For example, when kill traps are used, animals from the area around the grid will move onto the grid as local animals are removed (cf. Gentry et al. 1968, Smith et al. 1975:38). This violates closure, and the only good solution seems to be to keep the length of the study (in days) very short so that the study will end before significant immigration can occur.

Eliminating Variation Due to Time, Behavior, and Heterogeneity

Given that closure is satisfactorily achieved, the next most important consideration is twofold: (1) achieving a large enough number of captures to obtain reliable results, and (2) achieving a study for which the best model is the simplest possible one (e.g., Model M_0 rather than $M_{(tb)}$, or Model M_b rather than $M_{(tb)}$). In this section, we discuss methods of eliminating variation of capture probabilities due to time, behavioral response to first capture, and heterogeneity of individuals.

Of the 3 factors that affect capture probabilities, time is the one most easily controlled by the biologist. He can select the season of the year the studies are to be conducted, the length of the trapping period, and the time of day when trapping is to be done. In all those decisions, the objective is to reduce variation in capture probabilities over time. Among other things, this means that equal effort should be expended on each occasion. For example, use the same number of traps throughout, trap at the same time of day, and if bait is used, use the same type and amount on all occasions. The study should be done when weather conditions will be as constant as possible, because

variable weather is likely to cause time variation in capture probabilities (Getz 1961).

Behavioral response is common in small mammals, and it is doubtful if much can be done to avoid it. Apparently for this reason, many biologists have studied and used strict removal methods for small mammal population estimation. If there is any choice, a method of capture (for livetrapping) should be used that will not result in a trap response. One approach is to use different methods of capture on each occasion (cf. Overton and Davis 1969). This is an excellent procedure, but its application to multiple capture occasions is severely limited because one cannot usually find 5 or 10 quite different capture methods. We note, however, that the use of different trapping methods probably will result in time variation (the different methods occur of necessity at different times). Consequently, there is a trade-off here: reduced behavioral variation may result in increased time variation.

A common source of heterogeneity is the lack of equal access to traps if traps are far apart relative to home range (Eberhardt 1969a). The above phenomena are part of the reasoning behind our recommendation of 4 traps per home range when we discuss sample size considerations in the next section.

Other sources of heterogeneity are differences in activity or catchability related to measurable characteristics such as size, age, or sex. Such sources can be removed by stratification if the attributes are recorded and sample size permits. Unfortunately, sample sizes are seldom adequate to allow stratification. Heterogeneity may also be due to some unrecognized attribute, and thus cannot be eliminated by stratification. Different methods of capture may increase the number of animals captured in this case, but will not eliminate the basic problem of heterogeneity of capture probabilities.

One possible method of eliminating heterogeneity, and possible trap response, is to locate traps randomly on

each trapping occasion. We are unsure of the usefulness of the technique, and would like to see further studies conducted to see if significant reductions in heterogeneity result. For logistical reasons, the randomization of trap locations on each occasion probably is not feasible.

A valuable check on the livetrapping methods presented in this monograph can be made by using a second method of estimating the proportion of marked animals in the population. A common example is killtrapping or hunting after the capture-recapture experiment has been completed. Then, using the number of marked animals in the population, M_{t+1} , and the ratio of marked to unmarked animals obtained from the killtrapping or hunting, a second and somewhat independent population estimate is made with Chapman's (1951) version of the Lincoln estimate,

$$\hat{N} = \frac{(M_{t+1} + 1) \cdot (\text{all animals killtrapped} + 1)}{(\text{marked animals killtrapped} + 1)} - 1$$

(cf. Seber 1973). Of course traps are traps, and kill traps still present a strange object in the animal's environment. Another possibility for small mammals marked by toe clipping is to use tracks to obtain a ratio of marked to unmarked animals. Biologists have used smoked plates (Justice 1961) or dust covered tiles to obtain an estimate of the proportion of marked animals, and thus a second population estimate to compare with that obtained by livetrapping methods. The use of tracks has the distinct advantage of being a truly different method of sampling the population. Much confidence in the population estimate is obtained when the 2 estimates are similar.

In addition to design considerations, poststratification of the data can be used to create subsets of data which are more homogeneous with respect to capture probabilities. This is nothing more complex than partitioning the data into subsets on variables such as species, sex, age, weight, etc. If there are sufficient

data, such poststratification is a valuable device. The subsets so created are then analyzed separately by the methods described in this monograph. The only additional testing one might do is for homogeneity among these strata (subsets) (see White 1975).

Sample Size

To obtain reliable estimates of population size, a sufficiently large sample must be taken. Typical sample size considerations are not applicable here (e.g., determining the numbers of plots to sample). Rather, "sample size" relates to the number of animals captured. For a live-trapping study, one must have both a large enough number of distinct animals captured and a sufficient number of recaptures (except for Model M_b and M_{bh}). The factors that control expected numbers of captures are (1) grid size, in terms of area covered and numbers of traps used, (2) capture probabilities, and (3) number of trapping occasions. We discuss these 3 factors in relation to the size of the experiment necessary to achieve precise population estimates.

The size of the grid is the first decision to be made. Grid size is a function of trap spacing, s , and the number of rows, r , and columns, c , of traps. Equal trap spacing and a systematic grid layout are suggested. At each grid station there will be one or more traps. If densities are very high, we recommend 2 or more traps per station to avoid competition for traps. Few studies have been made comparing 2 or more traps per station with 1 trap station, and we suggest further research is required to see where multiple traps per station increases the probability of capture for individual animals.

The objective of grid trapping over a short time period is to estimate the population size and usually also the density at the grid site. Because of "edge effect" (as discussed in the section on density estimation), it is necessary to estimate effective trapping area as well as N . In practice, this means we must be able to

estimate strip width W as well as N . This requires that the traps be laid out to cover an area, rather than in a single line. Moreover, we must be able to associate each trap with an arbitrary X - Y coordinate system. For practical reasons, this implies some sort of regular grid layout (often square or rectangular) with equal spacing between traps. This latter aspect of trap layout is not necessary just for estimation of N : for example, if trapping was on a small island, or in an enclosure, knowledge of trap location in a coordinate system would not be needed. Finally, for the density estimate to be meaningful, the grid should be placed in a homogeneous habitat type, to assure uniform density over the grid.

For a choice of r and c when the objective is density estimation, we suggest that both values be greater than or equal to 5; as a minimum we recommend $r + c \geq 20$. Examples are a square grid 10×10 or a rectangle 5×15 . We note much work in the literature relies on 16×16 grids (Gentry et al. 1968, Smith et al. 1971), and we suggest that grids should be at least that large for attempts to use the density estimation method based on nested subgrids. We base this recommendation on the fact that a large number of captures is required in each subgrid; hence, the larger the size of the subgrids the better the chance that large numbers of captures will be achieved.

The next decision to be made is the spacing of traps (i.e., the value of s). Most work with small mammals uses 15-m trap spacing or less (Barbehenn 1974, Smith et al. 1975). The rationale, when one is given, relates to the size of home range. Let s be the spacing between traps, and let $2W$ be the average linear home range size. Home ranges may rarely be circular, but assuming for design purposes that they may be circular, then their radius is W . We suggest at least 4 traps per home range. This implies $s \leq (\sqrt{2})W$. For best results we suggest $s \leq W/2$. Clearly, this implies some knowledge of home range sizes before a good study can be designed. This is not unreasonable to ask;

the biologist should have some behavioral knowledge of the species being studied so that sampling decisions can be made intelligently. In fact, in any statistical sampling problem, a good study cannot be planned without some prior knowledge of the population parameters to be estimated.

The analyses presented in this monograph for estimating population size require sufficient numbers of captures to produce satisfactory results. Defining "sufficient numbers" is an extremely complicated task. Based on our experience with both real and simulated data, however, some crude guidelines may be stated. For instance, experiments in which M_{t+1} is on the order of 10 or 20 animals simply do not provide enough information for the procedures discussed here to perform well. The number of different animals captured needs to be several times larger, and will depend heavily on the probabilities of capture of the population members being studied. That is, a population in which members have an "average" capture probability of 0.40 or 0.50 might only have to be as large as 50 before the estimation and testing techniques become useful, whereas a population size of 200 or so might require an average capture probability of only 0.20. For most studies, a relatively large number of recaptures must be realized before the experiment has a chance to produce useful results, and this again relates to the magnitude of the probabilities of capture involved. In general, the probabilities must be larger for smaller populations, but in no instance should N be less than 25 or average capture probabilities less than 0.10 when trapping small mammals for only a few occasions (say $t \leq 10$). These recommendations do not guarantee that the data can be satisfactorily analyzed, but we have seen enough real and simulated data to say that if the data fail these criteria it is improbable that a precise estimate will be achieved.

Estimation of density by the nested subgrid approach requires even larger sample sizes; the data on the smaller

subgrids will otherwise be too sparse for reliable results. We believe that reliable density estimates using the subgrid approach require a grid of at least a 10×10 and as a *minimum* 75 to 100 different animals caught.

We now consider an approach to determining a minimum grid area on the basis of the above criteria. We start with the relationship $N = D \cdot A(W)$, where $A(W)$ is the effective trapping area. For a rectangular grid, this is

$$N = D[L_r L_c + 2(L_r + L_c)W + \pi W^2],$$

where L_r is the length of a row [$L_r = s \cdot (r - 1)$] and L_c [$L_c = s(c - 1)$] is the length of a column of traps. Thus, the area covered by the grid is $L_r L_c$.

As mentioned above, without some knowledge of D and W , a suitable study cannot be designed. Assume D_0 and W_0 are the best guesses of the values of the parameters. To determine if a grid study is *at all* feasible, set $N = 50$, substitute D_0 and W_0 in the above equation, set $L_r = L_c = L$ (a square grid), and solve for L :

$$L = \sqrt{0.8584(W_0)^2 + N/D_0 - 2W_0}.$$

Then the actual area the grid must cover is L^2 .

This procedure is not difficult; however, one must be careful to use the same basic units for D , L , and W . For example, let W_0 be 100 feet (30.5 m) and assume a density of 1 animal per acre (1/0.4 ha). For compatibility of units put D_0 in terms of square feet, then $D_0 = (1/43,560)$ feet². Solving for L gives

$$\begin{aligned} L &= \sqrt{(0.8584)(100)^2 + (50)(43,560) - 200} \\ &= 1,279 \text{ feet.} \end{aligned}$$

This translates back into 37.5 acres (15.2 ha) as an absolute minimum grid size [$37.5 = (1,279)^2/43,560$].

This is clearly conservative because not all animals will be caught. Improved planning requires us to determine grid size so that a given number of animals M_{t+1} will be caught. But the expected number of distinct animals caught de-

depends upon the true underlying capture probabilities, which are not known. The only practical approach is to make the best guess at the overall average probability of first capture, \bar{p} , applicable during the study and then use the formula

$$M_{t+1} = N[1 - (1 - \bar{p})^t].$$

Because the value of t (number of occasions) has been introduced, the relevant computations for several realistic values of t can now be performed.

For example, assume $\bar{p} = 0.30$, set $M_{t+1} = 50$, and solve for N for several values of t . Given these values of N , solve for the value of L , and hence minimum grid size:

t	N	L	grid size in acres (ha)
4	66	1,498	51.5 (20.8)
6	57	1,378	43.6 (17.6)
8	53	1,323	40.0 (16.2)

In practice, this example means if trapping were only for 4 days, one would need a 16×16 grid, traps spaced 100 feet (30.5 m) apart. For an 8-day period of trapping, the same (expected) data could be obtained with a 14×14 grid of traps spaced 100 feet (30.5 m) apart.

This process can be reversed. Let us say a study is planned with a square grid of 16 traps, spaced 15 m apart ($W = 7.5$ m). Then $L_r = L_c = 225$ m. Substitution in the basic equation gives

$$N = D \cdot 57,552 \quad (\text{m}^2)$$

or

$$N = D \cdot 5.76 \quad (\text{ha}).$$

Assume further the study is to last 5 days and the average capture probability is about $\bar{p} = 0.30$. Then we have

$$M_6 = N(0.83).$$

We need to get at least 100 animals because the smallest subgrid requires 50, so the density should be large enough that $N \geq 120$, or

$$D \geq \frac{120}{5.76} = 20.8 \text{ animals/ha.}$$

Thus, to get reliable results in such a study we would say the true population density should equal or exceed 21 animals/ha. If the biologist has good reason to believe true density is only 10 or 12 animals/ha, the study is not even worth doing.

In addition to controlling the sampling effort through the size of the grid and the number of traps, the biologist can also select the number of trapping occasions. In theory, the more trapping times there are the better, but this ignores the fact that the closure assumption becomes less realistic as more time passes. We recommend a minimum of 5 trapping occasions, but 7 to 10 is better. The interval between occasions should be short. In practice, most trapping of small animals is either once a day (morning) or twice a day (morning and evening). Trapping only once a day is far less likely to introduce time variation. With morning and evening trapping there is very likely to be a difference in capture probabilities between times. If variation of behavior and heterogeneity should also be present, the correct model ends up as M_{tth} for which no suitable estimator is available. Morning and evening trapping may, however, be aimed at different species. Then a workable design would be 5 (or 7) days of trapping in both morning and evening, but with separate analysis of the morning and evening data.

In removal studies, an absolute minimum is $t = 3$ occasions (*not* 2, as is often done) because it is impossible to test for equal capture probabilities when $t = 2$. We recommend that at least 4 removals be done.

Another valuable method of testing the adequacy of the design before going to the field is to simulate the experiment on a computer. Approximate parameter values can be chosen, and the experiment replicated as many times as necessary. Among the criteria that can be observed are selection of the appropriate model, the bias of selected estimators, and the achieved confidence level. Obviously, the validity of the simulations to the field

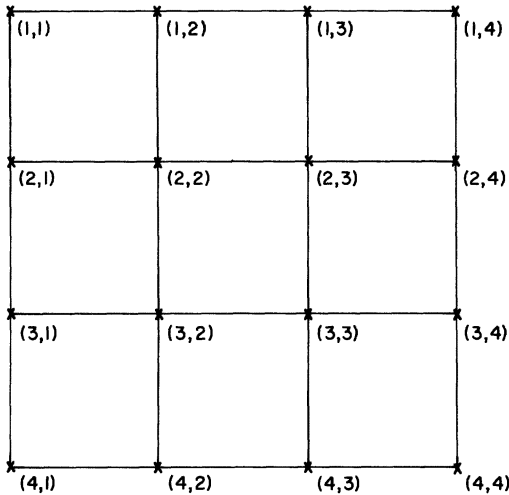


FIG. 20. Example of trap numbering for a 4×4 grid with a standard coordinate system.

study will depend on the similarity of parameters selected to the actual parameter values. However, insights about the experiment can be gained through simulations that cannot be achieved in any other way. We note that program CAPTURE has the capability to conduct these simulation studies.

Recording Data

If density is to be estimated on the basis of grid trapping data, the minimum information that must be taken when a capture occurs includes: (1) animal identification code, (2) trap location, and (3) trapping occasion. Usually, the species, sex, and age are also recorded for each animal, though for analyses given here it has no purpose except to partition the data by species, sex, and age. While analyzing the data separately by species is recommended, there often are not enough data to further partition by sex and age of animal.

For true removal studies (e.g., electrofishing), there is no animal identification code. As suggested by Raleigh (pers. comm.), it is very important in removal studies to record the individual by species and to analyze the data by at least major taxonomic groups.

In recapture studies, each animal should be marked uniquely, otherwise substantial information may be lost and it will be impossible to compute some of the tests for sources of variation. Though it should be obvious, we do mention that it is crucial to correctly record all data (e.g., animal number and trap number).

To know the trap location, traps must be numbered uniquely. Moreover, the biologist must know the relationship between the trap number and its coordinate on some (arbitrary) rectangular X-Y coordinate system. From the standpoint of data analysis, the best approach is to identify the traps in the field by these X-Y coordinates. We strongly recommend use of this system and we stress that density estimation using program CAPTURE requires data to be collected in the context of a coordinate system. We recommend that a corner trap be numbered (1, 1) then the rows become the "X axis" and columns the "Y axis." For example, a 4×4 grid would look like Fig. 20. This system can be extended to cover any regular rectangular grid of r rows and c columns.

When traps are checked both morning and evening, it is necessary to record not only the day of capture, but also the time.

Data recording will be facilitated by using standard field forms and standard conventions for trap numbering and animal identification. One example of a standardized method is presented in Brotzman and Giles (1966).

Data Anomalies

Various anomalies and unplanned events may occur in trapping. For example, (1) several animals may be found in 1 trap, (2) animals may be found dead in traps, (3) released animals may be found further down the grid trapped again on the same occasion, and (4) a trapped animal may escape when one attempts to remove it from the trap. We make the following suggestions regarding these happenings: (1) more than 1 animal per trap presents no problem. Record each animal

separately. This type of data does not invalidate the analyses present here; (2) an animal dead in the trap in a livetrapping study is a more serious problem. If it is the last trapping occasion it does not matter. Otherwise the data analysis must be modified. Some of the methods described here can allow for known removals in a true livetrapping study (specifically Models M_0 and M_t). Because not all tests and estimators can be so modified, we have not dealt here with such modified models. We recommend the following: if trap deaths are less than 5 percent of total captures, remove those data from the total results, run the analyses and add that number of dead animals to \hat{N} , and then multiply the density estimate \hat{D} , by $(1 + \text{proportion dead})$. If such deaths are more than 20 percent of total captures, use the generalized removal method of analysis on first captures. For 5–20 percent trap deaths, the only safe analysis may be the removal method. These modifications give \hat{N} and \hat{D} relative to the first day of the study. If one desires these estimators to apply to the population remaining alive after the study, then simply delete all trap deaths from the data set prior to analysis; (3) the same animal is caught more than once on a given occasion, the only added information provided is on movement. We recommend that both captures be recorded, but only the results of the first capture be used for data analysis; and (4) an animal escapes during handling before it is tagged, or before the mark is read. Do nothing but record the fact. Do not attempt any sort of analysis of this “record.” In essence, it does not become part of the data.

COMPREHENSIVE EXAMPLES

Preceding sections have given the details of the specific models, estimators, and tests that are the basis for our analysis of capture data. The analysis of a set of livetrapping data by these methods will involve the model selection procedure, followed by estimation of N under the selected (or most appropriate model);

OVERALL TEST RESULTS --	
Z-VALUE	-1.084
PROBABILITY OF A SMALLER VALUE	.13925

TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
(FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES	Z-VALUE	PROBABILITY
2	.184	.57306
3	.464	.67863
4	1.208	.88650

FIG. 21a. Example of test procedure for population closure with the Scheme B taxicab data from Carothers (1973b).

density estimation may also be desired. Finally, in the course of these analyses various summary statistics can be computed. Below, we give several complete examples of the entire analysis process of livetrapping data for purposes of estimating population size. We do not include density estimation for all these examples because its essential features are always the same. Also, no further examples are given here of the analysis of removal data because that subject is substantially simpler than the analysis of capture–recapture data.

A Taxicab Example

Carothers (1973b) conducted an ingenious capture–recapture experiment on the taxicab population of Edinburgh, Scotland. Such a study has the advantages of known population size, yet the population is a real one (though not involving animals) as opposed to a computer simulation experiment. Two different sampling methods were used; we have already used Carothers’ Scheme A in the section on Model M_h to illustrate the jackknife estimator (see Fig. 6); the entire 10 days of observations (“trappings”) from Scheme B are used as an example here. In that scheme, observers had fixed stations in the city. This corresponds to a trapping study with 10 days of trapping at fixed trap locations. The true population size was 420, and we can

OCCASION	J=	1	2	3	4	5	6	7	8	9	10
ANIMALS CAUGHT	N(J)=	48	52	47	44	48	45	48	43	47	53
TOTAL CAUGHT	M(J)=	0	48	90	122	146	173	188	203	213	225 241
NEWLY CAUGHT	U(J)=	48	42	32	24	27	15	15	10	12	16
FREQUENCIES	F(J)=	104	67	51	12	6	1	0	0	0	0

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)

CHI-SQUARE VALUE = 7.913 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .09482

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)

CHI-SQUARE VALUE = .095 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = .75743

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)

CHI-SQUARE VALUE = 2.247 DEGREES OF FREEDOM = 9 PROBABILITY OF LARGER VALUE = .98693

4. GOODNESS OF FIT TEST OF MODEL M(H)

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)

CHI-SQUARE VALUE = 2.300 DEGREES OF FREEDOM = 9 PROBABILITY OF LARGER VALUE = .98578

TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
(FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

NUMBER OF CAPTURES CHI-SQUARE D.F. PROBABILITY

NUMBER OF CAPTURES	CHI-SQUARE	D.F.	PROBABILITY
1	6.769	9	.66113
2	6.246	9	.71504
3	4.714	9	.85847

5. GOODNESS OF FIT TEST OF MODEL M(B)

NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)

CHI-SQUARE VALUE = 8.957 DEGREES OF FREEDOM = 16 PROBABILITY OF LARGER VALUE = .91516

5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

CHI-SQUARE VALUE = 7.598 DEGREES OF FREEDOM = 8 PROBABILITY OF LARGER VALUE = .47369

5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

CHI-SQUARE VALUE = 1.359 DEGREES OF FREEDOM = 8 PROBABILITY OF LARGER VALUE = .99480

6. GOODNESS OF FIT TEST OF MODEL M(T)

NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)

CHI-SQUARE VALUE = 188.341 DEGREES OF FREEDOM = 168 PROBABILITY OF LARGER VALUE = .13483

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)

CHI-SQUARE VALUE = 15.148 DEGREES OF FREEDOM = 26 PROBABILITY OF LARGER VALUE = .95458

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL CRITERIA	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
	.93	1.00	.37	.52	0.00	.46	.36	.59

FIG. 21b. Example of model selection procedure based on Scheme B taxicab data from Carothers (1973b). Appropriate model probably is M_h . Suggested estimator is jackknife.

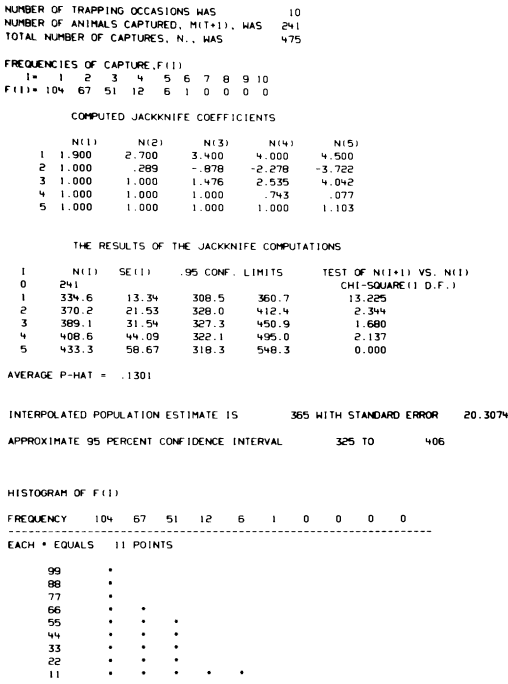


FIG. 21c. Example of population estimation with variable probability of capture by animal under Model M_h with Scheme B taxicab data from Carothers (1973b).

be reasonably certain of the closure assumption.

In this example, the concept of density is not applicable, nor is it meaningful to identify “trap” locations. Consequently, the first page of computer output is the test of closure (Fig. 21a). From Fig. 21a, we have the overall test value $z = -1.084$. This is not significant, consequently we would not reject closure. There were enough resightings 2, 3, and 4 times of the same taxicab to allow specific tests of closure on just those cabs seen that many times. None of the 3 tests are significant.

The next section of output is the model selection procedure (Fig. 21b). After the headings, some summary statistics are presented (daily captures n_j , cumulative marked M_j , new animals u_j , and capture frequencies f_j). The rest of the model selection procedure output (Fig. 21b) is devoted to the 7 tests of assumptions fol-

lowed by the computed selection criteria and a suggested appropriate model. The first 3 tests compare Model M_0 to Models M_h , M_b , and M_t , respectively. From Fig. 21b results are summarized below for those 3 tests:

Test of Model	Chi-square value	Significance level
M_0 versus M_h	7.913	0.09482
M_0 versus M_b	0.095	0.75740
M_0 versus M_t	2.247	0.98693

The only indication of variability in capture probabilities is heterogeneity (significant at the 10% level). The study was designed to achieve constant numbers captured every day; therefore it is not surprising there is no indication of time variability in capture probabilities. Also, we do not find it surprising that there is no indication, from test 2, of behavioral response. One would not expect “trap” responses from taxicabs.

Tests 4, 5, and 6 examine the goodness of fit of Models M_h , M_b , and M_t , respectively. When any of those tests reject, we are saying that model does not appear appropriate for the given study. Test 4 indicates no departure from Model M_h (overall chi-square of 2.3 with 9 df). Nor does test 5 reject Model M_b (overall chi-square of 8.957 with 16 df). The goodness of fit test to Model M_t does not cause us to reject M_t , but does make us suspicious of that model (probability of a larger value = 0.13483). Finally, test 7 gives us no cause to suspect Model M_h should be rejected in favor of Model M_{bh} (chi-square of 16.148 with 27 df).

By itself, none of these 7 tests is definitive in suggesting the appropriate model; rather it is necessary to consider the results of all the tests. In this example we see some evidence of heterogeneity, but no evidence of time or behavioral variations in capture probabilities. Bearing in mind the robustness of the jackknife estimator of Model M_h , we are willing to conclude the appropriate model for these data is M_h . The “model selection criteria” computed by the program suggests M_h as appropriate (with Model M_0 as the

best second choice). Given these results, the user should look next at the jackknife estimate for these data and should consider all other estimators as being inappropriate.

Fig. 21c gives the results of estimation of N from Carothers' Scheme B data. This page of output will always give the summary statistics M_{t+1} , n ., and the capture frequencies. Then the jackknife coefficients are given for the number of trapping occasions used. Next, the results of computing the first 5 jackknife estimates are shown. For example, $\hat{N}_{h1} = 334.6$ with a standard error of 13.34, while $\hat{N}_{h5} = 433.3$, and $\hat{SE}(\hat{N}_{h5}) = 58.67$. In that example, the selected estimate of N is 365 with standard error of 20.3. The approximate 95 percent confidence interval on N is 325 to 406. That confidence interval fails to cover the true $N = 420$, reflecting the previously discussed poor coverage of the confidence intervals associated with the jackknife estimator, but the relative error of the estimator is only 13.1 percent (which is consistent with our simulation results on \hat{N}_h). In terms of real capture-recapture studies, this amount of bias is, in our opinion, acceptable.

Finally, we note that jackknife estimator is better (i.e., nearer to $N = 420$) for Carothers' Scheme B data than the commonly used estimators as Petersen (1896) or Schnabel (1938).

As a further aid in judging the validity of the study, the estimated average capture probability is given. From Fig. 21c, AVERAGE P-HAT = 0.1301. Our studies have indicated that a value of less than 0.10 suggests the capture results may not be trusted to produce good results. A 0.13 average probability is not very high, but it is acceptable when the true population size is as high as 420.

A Pinned Rabbit Study

Edwards and Eberhardt (1967) report the results of a livetrapping study on a penned population of 135 wild cottontails *Sylvilagus floridanus*. To our knowl-

```
OVERALL TEST RESULTS --
Z-VALUE                               -2.132
PROBABILITY OF A SMALLER VALUE       .01650
```

```
TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
(FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)
```

```
NUMBER OF CAPTURES  Z-VALUE  PROBABILITY
-----
                2          3.691          .99989
```

FIG. 22a. Example of test for population closure procedure with cottontail data from Edwards and Eberhardt (1967).

edge there have been few other controlled studies like that done, which is unfortunate because it would be very valuable to have more data sets on real populations where N is known.

In that study, 135 wild cottontails were captured and placed in a 40-acre (16.2 ha) rabbit-proof enclosure. After allowing 4 days for the rabbits to adjust to their new surroundings, livetrapping was conducted for 18 consecutive nights. When program CAPTURE was used to analyze the resultant data, the results were disappointing.

Fig. 22a shows the results of the closure test applied to Edwards and Eberhardt's (1967) data. Because $z = -2.132$ ($P = 0.0165$) one would normally question whether closure was true. In that example, we attribute the result to a time variation in daily capture probabilities, specifically there were fewer captures toward the end of the 18 days than at the start. As mentioned before, this "closure test" can detect only certain types of time variations of individual capture probabilities. It cannot of itself "know" the cause of the variations. Therefore, all the evidence in the data or otherwise available must be used to reach final conclusions about closure, or about other questions such as the presence of behavioral response (which also gets confounded with closure).

A brief data summary and the tests of assumptions are given by the model selection procedure (see Fig. 22b). From tests 1, 2, and 3, we see that Model M_0 is

OCCASION	J=	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
ANIMALS CAUGHT	N(J)=	9	8	9	14	8	5	18	11	4	3	16	5	2	7	9	0	4	10	
TOTAL CAUGHT	M(J)=	0	9	15	18	29	33	34	44	51	52	53	62	62	63	68	74	74	74	76
NEWLY CAUGHT	U(J)=	9	6	3	11	4	1	10	7	1	1	9	0	1	5	6	0	0	2	
FREQUENCIES	F(J)=	43	16	8	6	0	2	1	0	0	0	0	0	0	0	0	0	0	0	

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
 NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)
 CHI-SQUARE VALUE = 11.110 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .01115

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
 NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)
 CHI-SQUARE VALUE = .062 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = .80367

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
 NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)
 CHI-SQUARE VALUE = 46.932 DEGREES OF FREEDOM = 17 PROBABILITY OF LARGER VALUE = .00012

4. GOODNESS OF FIT TEST OF MODEL M(H)
 NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)
 CHI-SQUARE VALUE = 55.502 DEGREES OF FREEDOM = 17 PROBABILITY OF LARGER VALUE = .00001
 TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
 (FREQUENCIES LESS THAN 21 ARE NOT CALCULATED.)

NUMBER OF CAPTURES	CHI-SQUARE	D.F.	PROBABILITY
1	52.023	17	.00002

5. GOODNESS OF FIT TEST OF MODEL M(B)
 NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)
 CHI-SQUARE VALUE = 102.913 DEGREES OF FREEDOM = 31 PROBABILITY OF LARGER VALUE = 0.00000

- 5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME
 CHI-SQUARE VALUE = 47.065 DEGREES OF FREEDOM = 15 PROBABILITY OF LARGER VALUE = .00004

- 5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME
 CHI-SQUARE VALUE = 55.848 DEGREES OF FREEDOM = 16 PROBABILITY OF LARGER VALUE = .00000

6. GOODNESS OF FIT TEST OF MODEL M(T)
 NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)
 EXPECTED VALUES TOO SMALL. TEST NOT PERFORMED.

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
 NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)
 CHI-SQUARE VALUE = 52.023 DEGREES OF FREEDOM = 17 PROBABILITY OF LARGER VALUE = .00002

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL CRITERIA	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
	.15	0.00	.29	.01	1.00	.64	.22	.32

FIG. 22b. Example of model selection procedure based on cottontail data from Edwards and Eberhardt (1967). Appropriate model probably is M_4 . Suggested estimator is Darroch.

not acceptable; moreover, the tests show that there is clearly time variation in average daily capture probabilities and that there probably is some heterogeneity

(tests 3 and 1). Although test 2 fails to suggest that Model M_b is better than Model M_o , we cannot, from that alone, conclude there are no behavioral re-

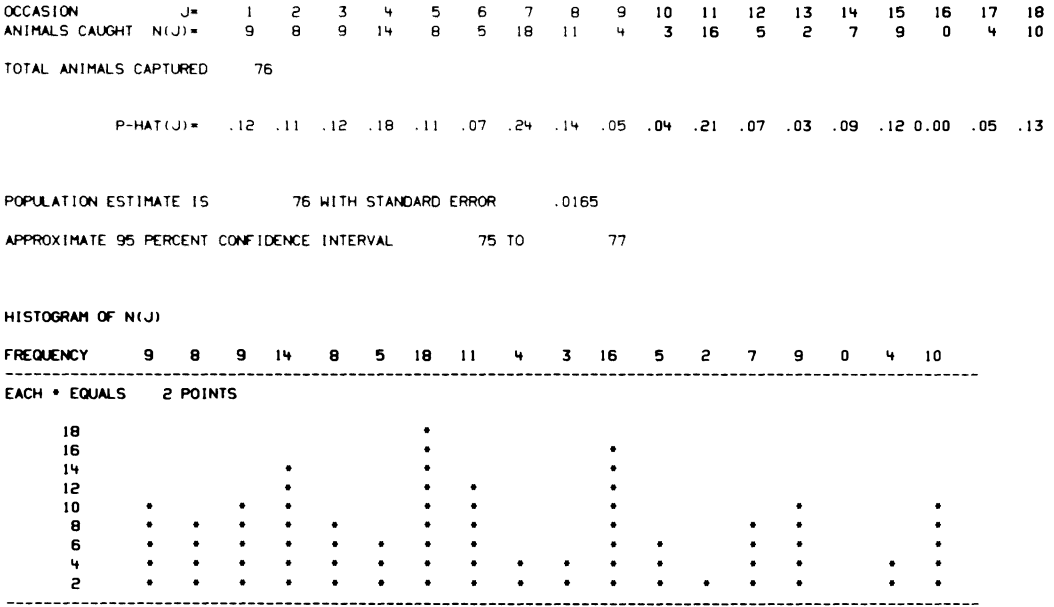


FIG. 22c. Example of population estimation with time specific changes in probability of capture under Model M_t with cottontail data from Edwards and Eberhardt (1967).

sponses. In fact, test 7 ($p = 0.00002$) shows that there are time variations in capture probabilities which *may* be behavioral.

Only 2 of the goodness of fit tests (4 and 5) could be computed. The goodness of fit to Model M_t could not be done because of the small numbers of individuals caught most days. The other 2 tests, however, resulted in rejection, indicating that neither Model M_h nor M_b is a satisfactory model for the data.

When the goodness of fit test cannot be computed for Model M_t , the model selection criteria are computed based on the *assumption* that test 6 would give a significance level of 0.50. This tends to give undue weight to Model M_t , but there is really no good solution to this problem. For Edwards and Eberhardt's data, M_t is the indicated best fitting model, with Model M_h the next best. This is consistent with the observation that capture probabilities do appear affected by both time and heterogeneity.

The estimation of N based on Model M_t is given in Fig. 22c. The point esti-

mate of population is 76, which is the same as the number of distinct animals captured (i.e., $M_{19} = 76$). As further evidence for determining whether \hat{N}_t is a reliable estimate (and model), one should compute the overall average capture probability:

$$\bar{p} = \frac{n}{t\hat{N}}$$

For Edwards and Eberhardt's data:

$$\bar{p} = \frac{142}{(18)(76)} = 0.10380.$$

The advantage of using this formula instead of

$$\bar{p} = \left(\sum_{j=1}^t p_j \right) / (t)$$

is that we know the true population size is greater than or equal to 76. Thus, we know this value of \bar{p} is an *upper bound* on the true expected value of \bar{p} , and hence average capture probability is less than or equal to 0.10. This should raise a red flag to the investigator, average capture probability may well be less than 0.10,

ANIMAL I.D.	NUM. CAP.	MAXIMUM DIST.	AVERAGE DIST.	STANDARD ERROR	ANIMAL I.D.	NUM. CAP.	MAXIMUM DIST.	AVERAGE DIST.	STANDARD ERROR	ANIMAL I.D.	NUM. CAP.	MAXIMUM DIST.	AVERAGE DIST.	STANDARD ERROR
268	6	4.5	1.7	.69	269	3	1.4	.7	.71	272	6	2.0	.8	.37
273	6	1.4	.3	.28	274	5	1.0	.3	.25	276	6	1.4	.7	.29
277	6	2.2	1.5	.30	279	5	8.5	4.0	1.75	280	4	2.2	.8	.42
281	3	5.0	3.2	1.79	282	5	1.4	1.0	0.00	285	5	7.1	3.3	1.45
286	6	2.0	.7	.29	287	5	2.0	.5	.29	288	6	2.2	.8	.20
294	5	1.4	.4	.35	299	4	3.0	1.0	1.00	300	3	2.0	2.0	0.00
163	4	1.0	.7	.33	165	4	4.1	3.5	.63	166	2	0.0	0.0	0.00
167	4	2.0	.7	.67	169	4	2.2	1.2	.65	170	5	2.0	.5	.50
171	5	4.1	1.8	.61	172	2	1.4	1.4	0.00	173	3	5.0	4.1	.92
175	1	0.0	0.0	0.00	176	3	0.0	0.0	0.00	177	3	1.0	1.0	0.00
184	1	0.0	0.0	0.00	187	4	2.2	.7	.75	188	3	0.0	0.0	0.00
189	2	3.0	3.0	0.00	191	4	1.0	.3	.33	192	3	0.0	0.0	0.00
193	3	0.0	0.0	0.00	196	3	1.4	1.2	.21	198	1	0.0	0.0	0.00
199	2	5.0	5.0	0.00	200	2	2.2	2.2	0.00	86	2	0.0	0.0	0.00
89	1	0.0	0.0	0.00	90	1	0.0	0.0	0.00	91	1	0.0	0.0	0.00
92	2	1.0	1.0	0.00	93	2	0.0	0.0	0.00	95	2	4.1	4.1	0.00
98	1	0.0	0.0	0.00	100	1	0.0	0.0	0.00	360	1	0.0	0.0	0.00

NOTE THAT AVERAGE DISTANCE ONLY REFERS TO DISTANCE BETWEEN SUCCESSIVE CAPTURES, WHILE MAXIMUM DISTANCE REFERS TO THE GREATEST DISTANCE BETWEEN ANY TWO CAPTURE POINTS. ALSO DISTANCE IS IN UNITS OF TRAP INTERVALS, I.E., IF THE INTER-TRAP DISTANCE IS 5 METERS, AND THE MAX. DISTANCE IS 1.4, THEN THE MAX. DISTANCE IN METERS IS 1.4*5 OR 7 METERS.

SUMMARY BY FREQUENCY OF CAPTURE OF MAXIMUM DISTANCE BETWEEN CAPTURE POINTS.

NUMBER CAPTURES	SAMPLE SIZE	MEAN OF MAX DIST.	STANDARD ERROR
2	9	1.86	.621
3	10	1.58	.614
4	8	2.23	.360
5	8	3.45	1.018
6	7	2.25	.392
TOTAL	42	2.23	.575

FIG. 23a. Example of animal by animal summary of deer mouse capture data from V. Reid (pers. comm.).

in which case this analysis is not trustworthy when (apparent) population size is around 100.

In fact, if we use the true value of $N = 135$, then average capture probability in this experiment was 0.056. We also point out that even after 18 days of trapping only 53 percent of the population had been captured. This example illustrates our contention that it is very important to have average capture probabilities well above 0.05 or 0.10 for the population sizes typically encountered in capture-recapture studies (50 to 150). Considering that $N = 135$ and the population was penned, we suggest that the “true” situation was as follows. There was significant time variation and heterogeneity but little real behavioral response. The average daily capture probabilities declined in the last 9 (of the 18) days and this caused rejection of both the closure test and test 7. No good estimator of population size is available for data of this type and quality.

An Example of Trap Response

Many data sets we have seen on *Peromyscus maniculatus* have fit Model M_b (trap response). The following is an example supplied by V. Reid (pers. comm.). The data were taken in a 6-day livetrapping study near Wet Swizer Creek, Rio Blanco County, Colorado, August 1975. A rectangular grid of 9×11 traps was used, with 50-foot (15.2-m) trap spacing. One Sherman live trap (for small mammals) was placed at each grid point and trapping was done twice daily (morning and night); we have used only morning captures for this example. The reader should study Figs. 23a–e, which give the basic results, before proceeding.

A summary of movement information from recaptures is given in Fig. 23a. We remind the reader these distances are in units of 1 trap spacing [i.e., 50 feet (15.2 m) here]. Thus, the mean maximum distance of 2.23 implies the average maximum movement was 111.5 feet (34 m).

OCCASION	J=	1	2	3	4	5	6
ANIMALS CAUGHT	N(J)=	16	29	27	29	32	38
TOTAL CAUGHT	M(J)=	0	16	31	38	44	48
NEWLY CAUGHT	U(J)=	16	15	7	6	4	3
FREQUENCIES	F(J)=	9	9	10	8	8	7

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)

CHI-SQUARE VALUE = 37.489 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .00000

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)

CHI-SQUARE VALUE = 24.342 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = .00000

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)

CHI-SQUARE VALUE = 24.773 DEGREES OF FREEDOM = 5 PROBABILITY OF LARGER VALUE = .00015

4. GOODNESS OF FIT TEST OF MODEL M(H)

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)

CHI-SQUARE VALUE = 25.225 DEGREES OF FREEDOM = 5 PROBABILITY OF LARGER VALUE = .00013

5. GOODNESS OF FIT TEST OF MODEL M(B)

NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)

CHI-SQUARE VALUE = 5.978 DEGREES OF FREEDOM = 8 PROBABILITY OF LARGER VALUE = .64968

5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

CHI-SQUARE VALUE = 1.135 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .88866

5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

CHI-SQUARE VALUE = 4.843 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .30379

6. GOODNESS OF FIT TEST OF MODEL M(T)

NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)

EXPECTED VALUES TOO SMALL. TEST NOT PERFORMED.

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)

CHI-SQUARE VALUE = 22.217 DEGREES OF FREEDOM = 7 PROBABILITY OF LARGER VALUE = .00233

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL CRITERIA	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
	.26	.17	1.00	.65	0.00	.28	.55	.27

FIG. 23b. Example of model selection procedure based on deer mouse data from V. Reid (pers. comm.). Appropriate model probably is M_b . Suggested estimator is Zippin.

One estimate of W (strip width) is produced by taking half this average maximum distance (56 feet, 17 m, here), or half of it plus one intertrap distance (giv-

ing 106 feet, 32.3 m, here). Either approach probably would be an improvement over using $W = 0$, but these are essentially ad hoc approaches.

OCCASION	J=	1	2	3	4	5	6	
TOTAL CAUGHT	M(J)=	0	16	31	38	44	48	51
NEWLY CAUGHT	U(J)=	16	15	7	6	4	3	
ESTIMATED PROBABILITY OF CAPTURE, P-HAT =		.319119						
ESTIMATED PROBABILITY OF RECAPTURE, C-HAT =		.677966						
POPULATION ESTIMATE IS		56 WITH STANDARD ERROR		4.2659				
APPROXIMATE 95 PERCENT CONFIDENCE INTERVALS		47 TO		65				
HISTOGRAM OF U(J)								
FREQUENCY	16	15	7	6	4	3		
EACH * EQUALS		2 POINTS						
16	•	•						
14	•	•						
12	•	•						
10	•	•						
8	•	•						
6	•	•	•					
4	•	•	•	•				
2	•	•	•	•	•			

FIG. 23c. Example of population estimation with constant probability removal estimator under Model M_b with deer mouse data from V. Reid (pers. comm.).

In this example, the test for closure gave $z = -1.586$ ($P = 0.05633$). But because the closure test is known to be biased when there is trap response, in the final analysis we conclude Model M_b adequately fits the data and closure is acceptable. A second bit of evidence on closure is given by the “ring” test (see Fig. 23e). That test examines for any apparent attraction of animals by the grid; it is not significant in this example, indicating that there is no basis to think immigration occurred.

From the model selection results (Fig. 23b) we find that tests 1, 2, and 3 all reject. This typically happens when there is strong trap response. Tests 4, 5, and 6 examine the goodness of fit of the 3 Models (M_h , M_b , and M_i , respectively). We see that Model M_h does not fit, but Model M_b does fit the data adequately. Unfortunately, the goodness of fit of M_i is the most difficult test to compute and with the small numbers of animals caught it could not be reliably computed.

Test 7 examines for possible behavioral variations in capture probabilities allowing for any heterogeneity which may be present (Fig. 23b). That test strongly suggests some form of behavioral response is present.

		MATRIX OF CAPTURES PER TRAP STATION.								
COLUMNS		1	2	3	4	5	6	7	8	9
ROW 1	1	0	0	1	3	7	3	5	4	4
ROW 2	1	3	3	4	4	1	5	0	1	2
ROW 3	1	2	2	2	0	0	4	2	4	2
ROW 4	1	3	2	3	2	0	1	1	1	2
ROW 5	1	1	0	0	1	0	0	1	4	5
ROW 6	1	0	2	2	2	2	0	4	0	0
ROW 7	1	6	1	2	2	3	3	1	0	0
ROW 8	1	2	2	4	0	4	0	3	2	3
ROW 9	1	4	0	0	5	0	2	0	0	0
ROW 10	1	0	3	0	1	0	0	0	0	0
ROW 11	1	0	4	2	0	1	2	0	1	1

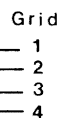


FIG. 23d. Nested subgrids used in the density estimation procedure with deer mouse data from V. Reid (pers. comm.). In the matrix, trap coordinates are rounded to the nearest whole integer. In the following goodness of fit tests, trap coordinates that are not integers and nonrectangular trapping grids will cause spurious results.

The suggested model is M_b , and the corresponding estimator of population size is, essentially, the Zippin removal estimator. (We have used the maximum likelihood estimator of N under the Zippin removal model. Recall that this differs slightly from the estimator usually computed.) Fig. 23c shows the estimate of population size under Model M_b to be $N = 56$, with a standard error of 4.3. The approximate 95 percent confidence interval on N is 51 to 65. Note that the program does not truncate the computed lower limit of 47 back to 51 (the number actually seen), but this could be done in reporting the results. The probability of initial capture is $\hat{p} = 0.32$ and the probability of recapture is $\hat{c} = 0.68$. These are significantly different because Model M_0 was rejected in favor of Model M_b . This sort of dramatic increase in capture probability after initial capture is entirely consistent with the properties of the data, for example, the observed increase over time in the n_j 's (daily numbers captured).

The area covered by the trapping grid was 4.59 acres (1.86 ha); the naive esti-

CHI-SQUARE TEST OF UNIFORM DENSITY BY ROWS.

ROW	1	2	3	4	5	6	7	8	9	10
OBSERVED	27	23	18	15	12	12	18	20	11	4
EXPECTED	15.545	15.545	15.545	15.545	15.545	15.545	15.545	15.545	15.545	15.545
CHI-SQUARE	8.440	3.575	.388	.019	.809	.809	.388	1.276	1.329	8.575

ROW	11
OBSERVED	11
EXPECTED	15.545
CHI-SQUARE	1.329

TOTAL CHI-SQUARE = 26.94 WITH 10 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .0027

CHI-SQUARE TEST OF UNIFORM DENSITY BY COLUMNS.

COLUMN	1	2	3	4	5	6	7	8	9
OBSERVED	21	19	25	15	20	18	17	17	19
EXPECTED	19.000	19.000	19.000	19.000	19.000	19.000	19.000	19.000	19.000
CHI-SQUARE	.211	0.000	1.895	.842	.053	.053	.211	.211	0.000

TOTAL CHI-SQUARE = 3.47 WITH 8 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .9012

CHI-SQUARE TEST OF UNIFORM DENSITY BY RINGS (OUTER RING IS NUMBER 1).

RING	1	2	3	4
OBSERVED	73	42	36	20
EXPECTED	62.182	48.364	34.545	25.909
CHI-SQUARE	1.882	.837	.061	1.348

TOTAL CHI-SQUARE = 4.13 WITH 3 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .2479

FIG. 23e. Chi-square tests of uniform density with deer mouse data from V. Reid (pers. comm.).

mate of density is thus 12.2 mice/acre (30.7/ha). Because of edge effect, we know this is an overestimate. As discussed above, movement data from recaptures is one basis for estimating the strip width W to get an estimate of effective trapping area. A conceptually better approach is to estimate W based on trapping grids of different sizes. Figs. 23e-f give results relevant to this approach for Reid's Wet Swizer Creek data. Four nested subgrids were chosen (as shown in Fig. 23d) as the basis of this procedure; from the number of captures per trap station, there is no apparent evidence of a nonuniform density over the grid. Figure 23e gives some logical tests of this uniform density assumption. However, our experience with the tests is that they often reject when there is, in fact, no real density gradient (i.e., systematic changes in density over the grid). Thus, even though there may be some variation in density by "rows," it is minor. Also, the

choice of subgrids as shown tends to minimize the problem.

For each subgrid, an estimate of the population at risk of capture on that subgrid was obtained using Model M_b . As explained in the density estimation section, we can then estimate W . The results are shown in Fig. 23f. The 4 naive densities are shown, with the necessary computed constants (based on grid sizes), and the initial covariance matrix of the naive densities.

The value of \hat{W} is 105 with a (typically large) standard error of 58. Nonetheless, this value of \hat{W} is significantly greater than zero, so we accept as our density estimate $\hat{D} = 5.54$ animals/acre (13.7/ha), ($\hat{SE}(\hat{D}) = 1.92$).

When this subgrid approach to estimation of D fails, one must use

$$\hat{D} = \hat{N}/A(\hat{W}),$$

$$A(\hat{W}) = A_g[1 + a\hat{W} + b\hat{W}^2],$$

STARTING VALUES FOR DENSITY ESTIMATION--
 NUMBER OF GRIDS 4
 TRAP INTERVAL 50.00
 UNITS CONVERSION 43560.00
 INITIAL DENSITY ESTIMATE 30.6995
 INITIAL STRIP WIDTH ESTIMATE .5228

GRID	NAIVE DENSITY Y(I)	PERIMETER/AREA A(I)	PI/AREA B(I)	STARTING COVARIANCE MATRIX			
1	32.67000	.3000000E-01	.1571E-03	150.			
2	61.71000	.1666667E-01	.5236E-04	507.	.152E+05		
3	17.42400	.1166667E-01	.2618E-04	7.74	232.	14.1	
4	12.19680	.9000000E-02	.1571E-04	1.16	34.7	2.11	.871

RESULTS OF ITERATIONS

FUNCTION EVALUATIONS REQUIRED 87
 ESTIMATED SIGNIFICANT DIGITS OF PARAMETER VALUES 8

FITTED MODEL COMPARED TO THE DATA

GRID(I)	Y(I)	F(I)
1	32.670	32.651
2	61.710	18.460
3	17.424	13.941
4	12.197	11.745

MULTIPLE CORRELATION COEFFICIENT IS .33718

ESTIMATED DENSITY= 5.538 1.9241 = ITS STANDARD ERROR
 ESTIMATED STRIP WIDTH= 105.225 57.9903 = ITS STANDARD ERROR
 CORRELATION OF ESTIMATORS -.9824

TEST OF ESTIMATED STRIP WIDTH GREATER THAN ZERO.
 Z-VALUE = 1.8145 PROBABILITY OF LARGER VALUE = .0348

FINAL COVARIANCE MATRIX

150.0			
918.6	.1521E+05		
18.80	311.4	14.10	
3.359	55.62	2.518	.8714

FIG. 23f. Example of joint estimation of density and boundary strip width with deer mouse data from V. Reid (pers. comm.).

where A_g is the grid area, \tilde{W} derives from either movement data or an independent source, and a and b are constants which can be computed. In fact, the program prints these constants on the density estimation page. From Fig. 23f, we have (for the total grid), $a = 0.009$ and $b = 0.00001571$.

55.75 (half the average maximum distance of 111.5) we get

$$A(\tilde{W}) = 4.59 \times 1.5506 = 7.117,$$

and hence,

$$\hat{D} = \frac{56}{7.117} = 7.87.$$

As an illustration only, if we use $\tilde{W} =$ (For previously given reasons, however,

OCCASION	J=	1	2	3	4	5	
ANIMALS CAUGHT	N(J)=	37	54	58	65	69	
TOTAL CAUGHT	M(J)=	0	37	68	77	98	110
NEWLY CAUGHT	U(J)=	37	31	9	21	12	
FREQUENCIES	F(J)=	34	20	28	15	13	

1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(H)

CHI-SQUARE VALUE = 49.016 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = 0.00000

2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(B)

CHI-SQUARE VALUE = 35.865 DEGREES OF FREEDOM = 1 PROBABILITY OF LARGER VALUE = 0.00000

3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.

NULL HYPOTHESIS OF MODEL M(O) VS. ALTERNATE HYPOTHESIS OF MODEL M(T)

CHI-SQUARE VALUE = 24.071 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .00008

4. GOODNESS OF FIT TEST OF MODEL M(H)

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(H)

CHI-SQUARE VALUE = 25.504 DEGREES OF FREEDOM = 4 PROBABILITY OF LARGER VALUE = .00004

TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
(FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

NUMBER OF CAPTURES	CHI-SQUARE	D.F.	PROBABILITY
1	18.647	4	.00092
2	18.333	4	.00106
3	6.048	4	.19562
4	6.000	4	.19915

5. GOODNESS OF FIT TEST OF MODEL M(B)

NULL HYPOTHESIS OF MODEL M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(B)

CHI-SQUARE VALUE = 13.510 DEGREES OF FREEDOM = 6 PROBABILITY OF LARGER VALUE = .03562

5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

CHI-SQUARE VALUE = 9.220 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .02650

5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

CHI-SQUARE VALUE = 4.289 DEGREES OF FREEDOM = 3 PROBABILITY OF LARGER VALUE = .23187

6. GOODNESS OF FIT TEST OF MODEL M(T)

NULL HYPOTHESIS OF MODEL M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL M(T)

CHI-SQUARE VALUE = 84.121 DEGREES OF FREEDOM = 66 PROBABILITY OF LARGER VALUE = .06562

7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.

NULL HYPOTHESIS OF MODEL M(H) VS. ALTERNATE HYPOTHESIS OF MODEL M(BH)

CHI-SQUARE VALUE = 47.135 DEGREES OF FREEDOM = 10 PROBABILITY OF LARGER VALUE = .00000

MODEL SELECTION CRITERIA. MODEL SELECTED HAS MAXIMUM VALUE.

MODEL	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
CRITERIA	.55	.42	.99	.89	0.00	.55	.88	1.00

we suspect that $\bar{W} = 55.75$ is an underestimate and suggest $\bar{W} = 105$ is a better value to use.)

The sampling variance of \bar{D} is computable by the formula given in the density estimation section. The only missing term is $\hat{\text{Var}}(\bar{W})$ which is obtained from the standard error of mean maximum distance (Fig. 23a)

$$\hat{\text{Var}}(\bar{W}) = \left(\frac{50 \cdot 0.575}{2} \right)^2 = 206.64.$$

Then using the formula for $\hat{\text{Var}}(\bar{D})$ we get $\hat{\text{Var}}(\bar{D}) = 0.98$, or $\hat{\text{SE}}(\bar{D}) = 0.99$.

An Example Where No Model Fits

In previous examples, the model selection procedure usually indicated a model that fit the data, and for which there was an estimator. However, with many real data sets we have found that no model fits the data. In those cases, the procedures given in this monograph serve to warn that no reliable estimator can be computed from the data (unless it can be judged the entire population has been caught).

In this example (Figs. 24a–e), we use livetrapping data collected by S. Hoffman (pers. comm.) in mid-July 1974 on deer mice. The study site was in a sagebrush–greasewood community at 4,500 feet (1,372 m) elevation in Curlew Valley, Idaho. A 12 × 12 grid of Sherman live traps, spaced 15 m apart, was used. Trapping was on 5 consecutive mornings, 110 individuals were caught (i.e., $M_{t+1} = 110$) and there were 283 captures in all.

Fig. 24a shows some summary statistics and the model selection procedure results. The daily captures increased (37, 54, 58, 65, 69), and although the new captures tended to decrease, they varied substantially (37, 31, 9, 21, 12). From these summary statistics, we would judge a substantial part of the catchable popu-

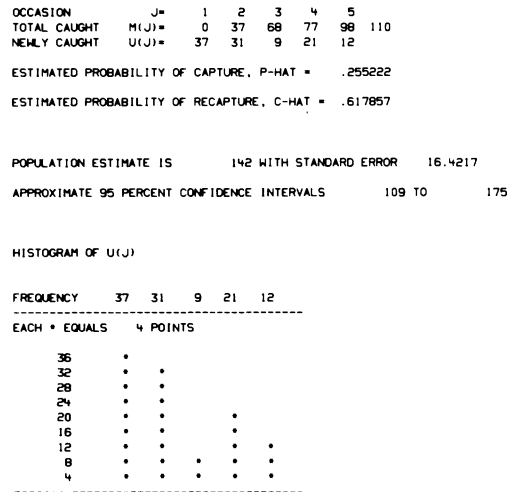


FIG. 24b. Example of population estimation with constant probability removal estimator under Model M_b with deer mouse data from S. Hoffman (pers. comm.).

lation remains untrapped. Phrased differently, there is no evidence here to suggest M_{t+1} is a reliable estimate of N .

From Fig. 24a, it can be seen that tests 1, 2, and 3 reject Model M_0 overwhelmingly. We thus conclude there is some type of variability in capture probabilities, and are suspicious that more than one source is operating. Test 4 also rejects the null hypothesis that Model M_h fits. At this point we can be fairly certain there is some form of time variability in daily capture probabilities, but it may be the often encountered result of behavioral variability.

The goodness of fit test to Model M_b indicates that this model is not a good fit to the data ($P = 0.03562$). Test 6 also suggests Model M_t is not an adequate fit to the data. Test 7 again confirms that some form of behavioral response is present so that Model M_h is inadequate. This leaves us uncertain as to how time and behavior variability are operating. Tests 5a and 5b

←

FIG. 24a. Example of model selection procedure based on deer mouse data from S. Hoffman (pers. comm.). Appropriate model probably is M_{th} or M_b . Suggested estimator is Zippin.

MATRIX OF CAPTURES PER TRAP STATION.

COLUMNS	1	2	3	4	5	6	7	8	9	10	11	12
ROW 1	4	3	5	5	5	3	2	0	3	4	3	1
ROW 2	5	5	0	4	1	0	4	3	0	1	3	0
ROW 3	4	2	0	3	1	0	2	2	1	3	3	1
ROW 4	1	5	3	4	1	0	2	5	2	4	2	3
ROW 5	0	1	4	3	0	1	3	0	0	1	5	2
ROW 6	3	5	0	1	1	0	3	2	0	0	4	5
ROW 7	2	0	2	0	0	0	1	3	0	1	2	1
ROW 8	4	1	4	2	0	0	1	0	0	3	2	0
ROW 9	1	2	4	2	4	0	0	2	4	4	2	4
ROW 10	2	0	5	2	0	0	1	0	1	2	4	4
ROW 11	1	3	1	0	0	0	0	3	4	2	2	2
ROW 12	4	4	3	0	2	0	0	0	1	4	4	2

IN THE ABOVE MATRIX, TRAP COORDINATES ARE ROUNDED TO THE NEAREST WHOLE INTEGER. IN THE FOLLOWING GOODNESS OF FIT TESTS, TRAP COORDINATES THAT ARE NOT INTEGERS AND NON-RECTANGULAR TRAPPING GRIDS WILL CAUSE SPURIOUS RESULTS.

CHI-SQUARE TEST OF UNIFORM DENSITY BY RINGS (OUTER RING IS NUMBER 1).

RING	1	2	3	4	5	6
OBSERVED	107	76	53	32	11	4
EXPECTED	86.472	70.750	55.028	39.306	23.583	7.861
CHI-SQUARE	4.873	.390	.075	1.358	6.714	1.896

TOTAL CHI-SQUARE = 15.31 WITH 5 DEGREES OF FREEDOM. PROBABILITY OF LARGER VALUE = .0091

FIG. 24c. Chi-square tests of uniform density with deer mouse data from S. Hoffman (pers. comm.).

help to clarify matters somewhat. Test 5a is the goodness of fit test to Zippin's (1956) constant probability removal model, i.e., the null hypothesis here is that first capture probability is constant. Test 5b examines the null hypothesis that recapture probabilities are constant. This latter test has $P = 0.23187$ suggesting that recapture probabilities may well be constant over time. By contrast with Test 5a ($P = 0.02650$) we conclude that first capture probabilities probably vary. These 7 tests suggest that behavior is the strongest factor affecting capture probabilities and that time is the next most significant factor.

The model selection criteria are a significant aid to interpreting data. In this example, the criteria values for models M_b , M_{bh} , M_{tb} , and M_{tbh} are 0.99, 0.89, 0.88, and 1.00, respectively. Recall that the most likely model is always given a value of 1, and the other criteria values are scaled accordingly, hence, we cannot say Model M_{tbh} actually fits the data. It is significant that these 4 models all account for behavioral response, thus corroborating our contention that there is a strong behavioral variation in capture probabilities. Because Model M_b is the next most likely model, one might select it as the basis for estimation. But there can be no

SUMMARY BY FREQUENCY OF CAPTURE OF MAXIMUM DISTANCE BETWEEN CAPTURE POINTS.

NUMBER CAPTURES	SAMPLE SIZE	MEAN OF MAX. DIST.	STANDARD ERROR
2	20	1.46	.279
3	28	2.43	.410
4	15	2.26	.388
5	13	2.44	.338
TOTAL	76	2.14	.401

FIG. 24d. Summary by frequency of capture of maximum distance between capture points.

strong reliance on \hat{N}_b or any other estimator here, because none of the models fit the data. From Fig. 24b, the first capture probability (p) and the recapture probability under Model M_b are estimated to be

$$\hat{p} = 0.26 \text{ and } \hat{c} = 0.62.$$

These are known to be significantly different because test 2 rejected Model M_b . From Model M_b the estimate of N is $\hat{N}_b = 142$, with an estimated standard error of 16.4. The approximate 95 percent confidence interval is 109–175. Other alternatives are also basically unsatisfactory. Model M_{bh} could be considered but it does not fit the data for any of its specific cases. This may be because first capture probabilities vary over time. Model M_h does not fit, and there are strong indications that behavior is the dominant source of variation. Based on these factors, we cannot recommend a valid estimation procedure for these data.

If density estimation is desired here, further problems arise. From Fig. 24c, there is no visual evidence of nonuniform density over the 12×12 grid. However, the ring test suggests significantly more animals were caught in the outer traps, which could be evidence of the grid “attracting” animals. This phenomenon does affect density estimation based on nested subgrids; in fact, this procedure failed for these data as the estimated W value was not significantly different from zero. The only recourse is to estimate W from movement data, or to use a value of W known to be representative for deer mice.

Fig. 24d shows the summary of maximum distances moved for recaptured animals. The overall average is 2.14 trap

OVERALL TEST RESULTS --

Z-VALUE	-2.142
PROBABILITY OF A SMALLER VALUE	.01610

TEST OF CLOSURE BY FREQUENCY OF CAPTURE. (FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES Z-VALUE PROBABILITY

NUMBER OF CAPTURES	Z-VALUE	PROBABILITY
2	-1.118	.13178
3	-1.708	.04383
4	-1.054	.14592

FIG. 24e. Example of test for population closure procedure with deer mouse data from S. Hoffman (pers. comm.).

units. Converting this to meters (2.14×15) and dividing by 2 we have 16 m (52.7 feet) as a conservative estimate of W . This is typical of such values seen for deer mice on livetrapping grids with 15-m trap spacing. It is also known to be an underestimate. Using $\hat{W} = 16$ m would give a less biased density estimate than $W = 0$, but it would still probably result in an overestimate of D . The matter is further complicated if animals were in fact attracted to the grid. This phenomenon cannot be adequately dealt with, except by assessment lines (or designing a study so that no attraction occurs). Dropping the outer ring of traps and reanalyzing the data is another possibility, but then we are treating this outer ring as an assessment line.

Normally, we should have looked at the closure test early on in the model assessment procedure (see Fig. 24e); the results are $z = -2.142$, $P = 0.0161$. Because there is strong behavioral variation, we must discount this test; it cannot be relied upon. Thus, the only evidence we have of a closure failure is the ring test of Fig. 24c.

We propose the following as a plausible explanation of these data. From Hoffman (pers. comm.) we know there was no prebaiting of these traps, no time was allowed for the animals to become used to the traps. From the analysis we know there was behavioral response, and probably time variation in first capture prob-

abilities (but not in recapture probabilities). The lack of prebaiting (often recommended in the literature) may have resulted in animals either increasing their first capture probabilities over time as the study progressed, or in an influx occurring during (rather than before) the study started. Prebaiting may solve the first problem. It will not solve the second problem of the grid attracting animals and thereby creating artificially high densities.

COMPREHENSIVE COMPUTER ALGORITHM

The computations necessary to calculate many of the estimates described in the preceding pages are quite lengthy, and in most cases, nearly impossible without a computer. Therefore, to provide methods of population estimation useful to the biologist, a comprehensive FORTRAN computer program, CAPTURE, has been written. The input to the program has been written in a free-form and natural style to make it easy to use.

The program is written in ANSI FORTRAN IV with several small exceptions so that it will function on most brands of digital computers. Statements known to cause compilation errors on IBM, CDC, Burroughs, Univac, Xerox, Honeywell, or DEC equipment have been flagged with comment statements, and the correct statement for the specific brand of computer included on a comment statement. A magnetic tape with the FORTRAN code and the 13 sets of example data illustrated in this monograph are available from

SHARE Program Library Agency
P. O. Box 12076
Research Triangle Park, N.C. 27709

at a cost of \$40.00. Specifications for the tape (e.g., 7 or 9 track, 800 or 1,600 bpi, etc.) should be given when ordering the source program (No. 360D-17.5.002).

The program consists of a main routine

and 54 subroutines. In addition, there are 7 common blocks. Included in the source code are ample comment statements to follow program flow. The total code consists of about 6,000 cards. Core requirements on an IBM 360 are about 200 K for the code without an overlay structure, but this can be reduced considerably with an overlay structure.

The dimensions of the program are presently set to allow up to 2,000 individual animals and 31 trapping occasions. The product of the number of captured animals and the number of trapping occasions must be less than 4,000. This will allow, for example, 30 trapping occasions and 80 animals, or 120 animals and 20 trapping occasions. These values can be changed by changing DIMENSION statements in the program.

SUMMARY

The results of this study provide methods for the estimation of animal population size N and density D from capture experiments. Both capture-recapture and removal studies to estimate population size are treated in detail. The primary focus of the work has been to relax the assumption of equal probability of capture. Three basic types of variation in probabilities of capture were examined: (1) capture probabilities vary with time, Model M_t , (2) capture probabilities vary by behavioral response, Model M_b , and (3) capture probabilities vary by individual animal, Model M_h . Models allowing these assumptions and various combinations of assumptions (i.e., Models M_{th} , M_{tb} , M_{bh} , and M_{tbh}) are treated. Population closure is assumed throughout.

An integrated approach was followed and the result is a blend of practical methods, examples of the analysis of real data, statistical theory, and results of computer simulation studies revealing some small sample properties of the methods. The estimation and testing problem has been treated in a standard and usually rigorous statistical framework. Above all, the necessity of assump-

tions, their importance, and statistical tests of their reasonableness have been emphasized. Assumptions must be carefully evaluated, both biologically and statistically, before a particular estimator can be used justifiably.

A sequence of 8 models that incorporate various assumptions has been developed. Point and interval estimators have been derived for most of the models. Furthermore, a comprehensive and objective (but not optimal) model selection strategy is provided. This is crucial because use of an incorrect model and improper assumptions is apt to produce biased estimates and incorrect inferences. In addition, proper interpretation of the test statistics and their interrelationships is shown to be somewhat difficult.

Although the practical analysis of data from capture experiments has been extended, additional research needs are apparent. Statistical testing within and between models will require more work. Models allowing other sets of assumptions need to be developed. Alternative estimation schemes (e.g., the generalized jackknife as opposed to the standard maximum likelihood approach) need attention, particularly for some of the more complex models. Additional work with incomplete contingency tables may prove fruitful (see Fienberg 1972). Interval estimates in general, and particularly interval estimates for the removal models, require additional research. Better tests for the important closure assumption are needed. Additional controlled studies are needed where population size is known, such as that reported by Edwards and Eberhardt (1967). Behavioral studies aimed at specific animal-trap interactions promise to provide interesting insights.

Research results indicate that acceptable estimates can be obtained if adequate data are collected properly. However, it is clear that biologists have not correctly conceptualized capture experiments over the past 4 decades. These limitations must be corrected if progress is expected in biological experiments in-

volving capture of animals. Biologists have all too often gone into the field with only a few traps hoping to obtain meaningful data. Rarely will this be possible, even if only an estimate of population size is the goal. If density estimation is a goal, then further data requirements must be met.

As a reference point, it probably is appropriate to think in terms of 12×12 , 16×16 , or 20×20 square or rectangular grids and a trapping period of 8–9 days. Further guidelines on effort and sample size are given in the text. Our primary contention is that most capture experiments conducted in the past were quite inadequate in terms of design, effort, and sample size, which virtually prohibits justifiable inferences to be drawn from such studies.

Our computer simulation experiments have examined the small sample properties of the various point and interval estimators and the tests of model assumptions. The results have been informative and provide some basis for cautious optimism. Capture-recapture and removal methods have been overrated in the past and this probably has contributed to the lack of emphasis on design, sample size, and analytical methods. Estimation problems related to \hat{D} and \hat{N} represent difficult subjects. Our results provide hope that a rigorous analysis may often allow useful inferences to be drawn if future experiments are well designed and provide adequate capture and recapture or removal data for analysis.

LITERATURE CITED

- ANDRZEJEWSKI, R., AND W. JEZIEWSKI. 1966. Studies on the European hare: XI. Estimation of population density and attempt to plan the yearly take of hares. *Acta Theriol.* 11:433–448.
- ANONYMOUS. 1971. *Mathematical intrinsics: Information manual.* Burroughs Corp. El Monte, Ca. 37 pp.
- ARNASON, A. N. 1972a. Prediction methods and variance estimates for the parameters of the triple catch—two population model with migration and death. *Univ. Manitoba Sci. Rept.* 54:1–31.
- . 1972b. Parameter estimates from mark-re-

- capture experiments on two populations subject to migration and death. *Res. Pop. Ecol.* 13(2):97-113.
- . 1973. The estimation of population size, migration rates and survival in a stratified population. *Res. Pop. Ecol.* 15(1):1-8.
- , AND L. BANIUK. 1977. User's manual: POPAN-2, A data maintenance and analysis system for recapture data. Univ. Manitoba, Dept. Computer Sci., Winnipeg, Canada.
- BAILEY, G. N. A. 1968. Trap-shyness in a woodland population of bank voles (*Clethrionomys glareolus*). *J. Zool.* 156(10):517-521.
- BAILEY, N. T. J. 1951. On estimating the size of mobile populations from recapture data. *Biometrika* 38(3/4):293-306.
- BARBEHENN, K. R. 1974. Estimating density and home range size with removal grids: The rodents and shrews of Guam. *Acta Theriol.* 19:191-234.
- BIRNBAUM, A. 1969. Concepts of statistical evidence. In S. Morganbesser, et al. (Eds.). *Philosophy, science and method: Essays in honor of E. Nagel*. St. Martin's Press, New York, N.Y.
- BRAATEN, D. O. 1969. Robustness of the DeLury population estimator. *J. Fish. Res. Bd. Can.* 26(2):339-355.
- BRENT, R. P. 1973. Algorithms for minimization without derivatives. Prentice-Hall, Inc., Englewood Cliffs, N.J. 196 pp.
- BROTZMAN, R. L., AND R. H. GILES, JR. 1966. Electronic data processing of capture-recapture and related ecological data. *J. Wildl. Manage.* 30(2):286-292.
- BROWNIE, C., D. R. ANDERSON, K. P. BURNHAM, AND D. S. ROBSON. 1978. Statistical inference from band recovery data: A handbook. U.S. Fish Wildl. Serv., Resour. Publ. 131. 212 pp.
- BURNHAM, K. P., AND W. S. OVERTON. 1969. A simulation study of live-trapping and estimation of population size. Ore. St. Univ., Dept. Stat., Tech. Rept. No. 14. 60 pp. and appendix.
- CAROTHERS, A. D. 1971. An examination and extension of Leslie's test of equal catchability. *Biometrics* 27(3):615-630.
- . 1973a. The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29(1):79-100.
- . 1973b. Capture-recapture methods applied to a population with known parameters. *J. Anim. Ecol.* 42(1):125-146.
- CHAPMAN, D. G. 1948. A mathematical study of confidence limits of salmon populations calculated from sample tag ratios. *Int. Pac. Salmon Fish. Comm. Bull.* 11:69-85.
- . 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. *Univ. Cal. Publ. Stat.* 1(7):131-160.
- . 1952. Inverse, multiple and sequential sample censuses. *Biometrics* 8(4):286-306.
- . 1954. The estimation of biological populations. *Ann. Math. Stat.* 25(1):1-15.
- COOLEY, W. W., AND P. R. LOHNES. 1962. Multivariate procedures for the behavioral sciences. John Wiley & Sons, Inc., New York, N.Y. 211 pp.
- CORMACK R. M. 1966. A test for equal catchability. *Biometrics* 22(2):330-342.
- . 1968. The statistics of capture-recapture methods. *Oceanogr. Mar. Biol. Annu. Rev.* 6:455-506.
- CRAIG, C. C. 1953. On the utilization of marked specimens in estimating populations of flying insects. *Biometrika* 40(1/2):170-176.
- CROWCROFT, P., AND J. N. R. JEFFERS. 1961. Variability in the behavior of wild house mice (*Mus musculus* L.) towards live-traps. *Proc. Zool. Soc. Lond.* 137(4):573-582.
- DARROCH, J. N. 1958. The multiple-recapture census: I. Estimation of a closed population. *Biometrika* 45(3/4):343-359.
- . 1959. The multiple-recapture census: II. Estimation when there is immigration or death. *Biometrika* 46(3/4):336-351.
- . 1961. The two-sample capture-recapture census when tagging and sampling are stratified. *Biometrika* (3/4):241-260.
- DAVIS, D. E. 1956. Manual for analysis of rodent populations. Edwards Brothers, Inc., Ann Arbor, Mich. 82 pp.
- DELURY, D. B. 1947. On the estimation of biological populations. *Biometrics* 3(4):145-167.
- . 1951. On the planning of experiments for the estimation of fish populations. *J. Fish. Res. Bd. Can.* 8(4):281-307.
- . 1958. The estimation of population size by marking and recapture procedure. *J. Fish. Res. Bd. Can.* 15(1):19-25.
- DICE, L. R. 1938. Some census methods for mammals. *J. Wildl. Manage.* 2(2):119-130.
- . 1941. Methods for estimating populations of animals. *J. Wildl. Manage.* 5(4):398-407.
- EBERHARDT, L. L. 1969a. Population estimates from recapture frequencies. *J. Wildl. Manage.* 33(1):28-39.
- . 1969b. Population analysis. Pp. 457-495. In R. H. Giles, Jr., (Ed.). *Wildlife management techniques*. 3rd ed. The Wildlife Society, Wash., D.C.
- , T. J. PETERLE, AND R. SCHOFIELD. 1963. Problems in a rabbit population study. *Wildl. Monogr.* No. 10:1-51.
- EDWARDS, W. R., AND L. L. EBERHARDT. 1967. Estimating cottontail abundance from live trapping data. *J. Wildl. Manage.* 31(1):87-96.
- FELLER, W. 1950. An introduction to probability theory and its applications. Vol. 1. John Wiley & Sons, Inc., New York, N.Y. 509 pp.
- FIENBERG, S. E. 1972. The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* 59(3):591-603.
- FISCHLER, K. J. 1965. The use of catch-effort, catch sampling, and tagging data to estimate a population of blue crabs. *Trans. Amer. Fish. Soc.* 94(3):287-310.

- FISHMAN, G. S. 1973. Concepts and methods in discrete event digital simulation. Wiley-Interscience, New York, N.Y. 385 pp.
- FLETCHER, R. 1972. FORTRAN subroutines for minimization of quasi-Newton methods. Atom. Energy Res. Estab., Rep. 7125. Harwell, Eng. 34 pp.
- FLYGER, V. R. 1959. Movements and home range of the gray squirrel *Sciurus carolinensis*, in two Maryland woodlots. *Ecology* 41(3):365-369.
- GASKELL, T. J., AND B. J. GEORGE. 1972. A Bayesian modification of the Lincoln index. *J. Anim. Ecol.* 41(3):377-384.
- GEIS, A. D. 1955a. Trap response of the cottontail rabbit and its effect on censusing. *J. Wildl. Manage.* 19(4):466-472.
- GENTRY, J. B., F. B. GOLLEY, AND M. H. SMITH. 1968. An evaluation of the proposed International Biological Program census methods for estimating small mammal populations. *Acta Theriol.* 13:313-327.
- GETZ, L. L. 1961. Response of small mammals to live-traps and weather conditions. *Amer. Midl. Nat.* 66(1):160-169.
- GILBERT, R. O. 1973. Approximations of the bias in the Jolly-Seber capture-recapture model. *Biometrics* 29(3):501-526.
- GOODMAN, L. A. 1953. Sequential sampling tagging for population size problems. *Ann. Math. Stat.* 24(1):56-69.
- GRAY, H. L., AND W. R. SCHUCANY. 1972. The generalized jackknife statistic. Marcel Dekker, New York, N.Y. 308 pp.
- GRAYBILL, F. A. 1976. Theory and application of the linear model. Duxbury Press, North Scituate, Mass. 704 pp.
- GREEN, R. G., AND C. A. EVANS. 1940. Studies on a population cycle of snowshoe hares on the Lake Alexander area: I. Gross annual census, 1932-1939. *J. Wildl. Manage.* 4(2):220-238.
- HANSSON, L. 1969. Home range, population structure and density estimates at removal catches with edge effect. *Acta Theriol.* 14:153-160.
- . 1974. Influence area of trap stations as a function of number of small mammals exposed per trap. *Acta Theriol.* 19:19-25.
- HAYNAM, G. E., Z. GOVINDARAJULU, AND F. C. LEONE. 1970. Tables of the cumulative non-central chi-square distribution. Pp. 1-42. In H. L. Harter and D. B. Owen (Eds.). Selected tables in mathematical statistics. Vol. 1. Amer. Math. Soc., Providence, R. I.
- HAYNE, D. W. 1949a. Two methods for estimating populations from trapping records. *J. Mammal.* 30(4):399-411.
- . 1949b. Calculation of size of home range. *J. Mammal.* 30(1):1-18.
- HUBER, J. J. 1962. Trap response of confined cottontail populations. *J. Wildl. Manage.* 26(2):177-185.
- HUDSON, D. J. 1971. Interval estimation from the likelihood function. *J. R. Stat. Soc., B.* 33:256-262.
- IMSL 1976. International mathematical statistical libraries. 5th Ed. Houston, Texas. (pages not consecutively numbered).
- JACKSON, C. H. N. 1933. On the true density of tsetse flies. *J. Anim. Ecol.* 2(2):204-209.
- . 1937. Some new methods in the study of *Glossina morsitans*. *Proc. Zool. Soc. Lond.*, 1936. 811-896.
- . 1939. The analysis of an animal population. *J. Anim. Ecol.* 8(2):238-246.
- . 1940. The analysis of a tsetse fly population. *Ann. Eugen.* 10(4):332-369.
- JENNRICH, R. I., AND F. B. TURNER. 1969. Measurement of non-circular home range. *J. Theoret. Biol.* 22:227-237.
- JOLLY, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigration—stochastic model. *Biometrika* 52(1/2):225-247.
- JUSTICE, K. E. 1961. A new method for measuring home ranges of small mammals. *J. Mammal.* 42(4):462-470.
- KAUFMAN, D. W., G. C. SMITH, R. M. JONES, J. B. GENTRY, AND M. H. SMITH. 1971. Use of assessment lines to estimate density of small animals. *Acta Theriol.* 16:127-147.
- KENDALL, M. G., AND W. R. BUCKLAND. 1971. Dictionary of statistical terms. 3rd edition. Hafner Publ. Co., New York, N.Y. 166 pp.
- , AND A. STUART. 1973. The advance theory of statistics. Vol. 2, 3rd ed. Griffin, London, Eng. 723 pp.
- KETCHEN, K. S. 1953. The use of catch-effort and tagging data in estimating a flat-fish population. *J. Fish. Res. Bd. Can.* 10(4):459-483.
- LESLIE, P. H. 1952. The estimation of population parameters from data obtained by means of the capture-recapture method. II. The estimation of total numbers. *Biometrika* 39(3/4):363-388.
- . 1958. Statistical appendix. *J. Anim. Ecol.* 27(1):84-86.
- , AND D. H. S. DAVIS. 1939. An attempt to determine the absolute number of rats on a given area. *J. Anim. Ecol.* 8(1):94-113.
- , AND D. CHITTY. 1951. The estimation of population parameters from data obtained by means of the capture-recapture method: I. The maximum likelihood equations for estimating the death-rate. *Biometrika* 38(3/4):269-292.
- , ———, AND H. CHITTY. 1953. The estimation of population parameters from data obtained by means of the capture-recapture method: III. An example of the practical applications of the method. *Biometrika* 40(1/2):137-169.
- LINCOLN, F. C. 1930. Calculating waterfowl abundance on the basis of banding returns. *Circ. U.S. Dept. Agric.* No. 118:1-4.
- MACLULICH, D. A. 1951. A new technique of animal census, with examples. *J. Mammal.* 32(3):318-328.
- MANLY, B. F. J. 1970. A simulation study of animal

- population estimation using the capture-recapture method. *J. Appl. Ecol.* 7(1):13-39.
- . 1971. Estimates of a marking effect with capture-recapture sampling. *J. Appl. Ecol.* 8(1):181-189.
- MARTEN, G. G. 1970. A regression method for mark-recapture estimation of population size with unequal catchability. *Ecology* 51(2):291-295.
- MILLER, R. P. 1964. Simultaneous statistical inference. McGraw-Hill Book Co., New York, N.Y. 272 pp.
- MOHR, C. O., AND W. A. STUMPF. 1966. Comparison of methods for calculating areas of animal activity. *J. Wildl. Manage.* 30(2):293-304.
- MOOD, A. M., F. A. GRAYBILL, AND D. C. BOES. 1974. Introduction to the theory of statistics. 3rd ed. McGraw-Hill Book Co., New York, N.Y. 564 pp.
- MORAN, P. A. P. 1951. A mathematical theory of animal trapping. *Biometrika* 38(3/4):307-311.
- NIE, N. H., C. H. HULL, J. G. JENKINS, K. STEINBRENNER, AND D. H. BRENT. 1975. Statistical package for the social sciences. 2nd ed. McGraw-Hill Book Co., New York, N.Y. 675 pp.
- NIXON, C. N., W. R. EDWARDS, AND L. L. EBERHARDT. 1967. Estimating squirrel abundance from live trapping data. *J. Wildl. Manage.* 31(1):96-101.
- O'FARRELL, M. J., D. W. KAUFMAN, AND D. W. LUNDAHL. 1977. Use of live-trapping with the assessment line method for density estimation. *J. Mammal.* 58(4):575-582.
- OMAND, D. N. 1951. A study of populations of fish based on catch-effort statistics. *J. Wildl. Manage.* 15(1):88-98.
- OVERTON, W. S., AND D. E. DAVIS. 1969. Estimating the number of animals in wildlife populations. Pp. 403-455. In R. H. Giles, Jr., (Ed.). *Wildlife management techniques*. 3rd ed. The Wildlife Society, Wash., D.C.
- PAULIK, G. J. 1963. Estimates of mortality rates from tag recoveries. *Biometrics* 19(1):28-57.
- , AND D. S. ROBSON. 1969. Statistical calculations for change-in-ratio estimators of population parameters. *J. Wildl. Manage.* 33(1):1-27.
- PETERSEN, C. G. J. 1896. The yearly immigration of young plaice into the Limfjord from the German Sea. *Rep. Danish Biol. Sta.* 6:1-48.
- POLLOCK, K. H. 1975. A K-sample tag-recapture model allowing for unequal survival and catchability. *Biometrika* 62(3):577-583.
- , D. L. SOLOMON, AND D. S. ROBSON. 1974. Tests for mortality and recruitment in a K-sample tag-recapture experiment. *Biometrics* 30(1):77-87.
- PUCEK, Z. 1969. Trap response and estimation of numbers of shrews in removal catches. *Acta Theriol.* 14:403-426.
- QUENOUILLE, M. 1949. Approximate tests of correlation in time series. *J. Roy. Stat. Soc., Ser. B.* 11(1):68-84.
- . 1956. Notes on bias in estimation. *Biometrika* 43(3/4):353-360.
- RICKER, W. E. 1958. Handbook of computations for biological statistics of fish populations. *Bull. Fish. Res. Bd. Can. No. 119.* 300 pp.
- . 1975. Computation and interpretation of biological statistics of fish populations. *Dept. Environ., Bull.* 191. 382 pp.
- ROBSON, D. S. 1969. Mark-recapture methods of population estimation. Pp. 120-140. In N. L. Johnson and H. Smith, Jr. (Eds.). *New developments in survey sampling*. Wiley-Interscience, New York, N.Y.
- , AND W. A. FLICK. 1965. A nonparametric statistical method for culling recruits from a mark-recapture experiment. *Biometrics* 21(4):936-947.
- , AND H. A. REGIER. 1964. Sample size in Petersen mark-recapture experiments. *Trans. Amer. Fish. Soc.* 93(3):215-226.
- , AND ———. 1968. Estimation of population number and mortality rates. Pp. 124-158. In W. E. Ricker (Ed.). *Methods for assessment of fish production in fresh waters*. IBP Handbook No. 3, Blackwell Scientific Publ., Oxford, Eng.
- , AND W. D. YOUNGS. 1971. Statistical analysis of reported tag-recaptures in the harvest from an exploited population. *Biometrics Unit Rep. BU-369-M*, Cornell Univ., Ithaca, N.Y. 15 pp.
- SAMUEL, E. 1968. Sequential maximum likelihood estimates of the size of a population. *Ann. Math. Stat.* 39(3):1057-1068.
- SCHNABEL, Z. E. 1938. The estimation of the total fish population of a lake. *Amer. Math. Month.* 45(6):348-352.
- SCHUMACHER, F. X., AND R. W. ESCHMEYER. 1943. The estimate of fish population in lakes or ponds. *J. Tenn. Acad. Sci.* 18(3):228-249.
- SEBER, G. A. F. 1962. The multi-sample single recapture census. *Biometrika* 49(3/4):339-349.
- . 1965. A note on the multiple-recapture census. *Biometrika* 52(1/2):249-259.
- . 1970. The effects of trap-response on tag-recapture estimates. *Biometrics* 26(1):13-22.
- . 1973. Estimation of animal abundance and related parameters. Griffin, London, Eng. 506 pp.
- , AND J. F. WHALE. 1970. The removal method for two and three samples. *Biometrics* 26(3):393-400.
- SMITH, H. D., C. D. JORGENSEN, AND H. D. TOLLEY. 1972. Estimation of small mammals using recapture methods: Partitioning of estimator variables. *Acta Theriol.* 17:57-66.
- SMITH, M. H. 1968. A comparison of different methods of capturing and estimating numbers of mice. *J. Mammal.* 49(4):455-462.
- , J. B. GENTRY, AND F. B. GOLLEY. 1969. A preliminary report on the examination of small mammal census methods. Pp. 25-29. In K. Petrusiewicz and L. Ryszkowski (Eds.). *Energy*

- flow through small mammal populations. Polish Scientific Publ., Warsaw, Poland.
- , R. BLESSING, J. G. CHELTON, J. B. GENTRY, F. B. GOLLEY, AND J. T. MCGINNIS. 1971. Determining density for small mammal populations using a grid and assessment lines. *Acta Theriol.* 16:105–125.
- , R. H. GARDNER, J. B. GENTRY, D. W. KAUFMAN, AND M. H. O'FARREL. 1975. Density estimation of small mammal populations. Pp. 25–53. In F. B. Golley, K. Petruszewicz and L. Ryszkowski (Eds.). *Small mammals: Their productivity and population dynamics.* Intern. Biol. Prog. 5., Cambridge Univ. Press, London, Eng.
- SNEDECOR, G. W., AND W. G. COCHRAN. 1967. *Statistical methods.* 6th ed. Iowa St. Univ. Press, Ames, Ia. 593 pp.
- SOUTHWOOD, T. R. E. 1966. *Ecological methods.* Methuen Co., London, Eng. 391 pp.
- STEIN, C. 1962. A remark on the likelihood principle. *J. Roy. Stat. Soc., A.* 125(4):565–568.
- STICKEL, L. F. 1954. A comparison of certain methods of measuring ranges of small mammals. *J. Mammal.* 35(1):1–15.
- SWIFT, D. M., AND R. K. STEINHORST. 1976. A technique for estimating small mammal population densities using a grid and assessment lines. *Acta Theriol.* 21(4):471–480.
- SWINEBROAD, J. 1964. Net-shyness and wood thrush populations. *Bird-Banding* 33(3):196–202.
- TABER, R. D., AND I. M. COWAN. 1969. Capturing and marking wild animals. Pp. 277–317. In R. H. Giles, Jr. (Ed.). *Wildlife management techniques.* 3rd ed. The Wildlife Society, Wash., D.C.
- TANAKA, R. 1951. Estimation of vole and mouse populations on Mount Ishizuchi and on the uplands of Southern Shikoku. *J. Mammal.* 32(4):450–458.
- . 1952. Theoretical justification of the mark-and-release index for small mammals. *Bull. Kochi Women's Coll.* 1:38–47.
- . 1956. On differential response to live traps of marked and unmarked small mammals. *Ann. Zool. Jap.* 29(1):44–51.
- . 1963. On the problem of trap-response types of small mammal populations. *Res. Pop. Ecol.* 2:139–146.
- . 1970. A field study of the effect of prebaiting on censusing by the capture-recapture method in a vole population. *Res. Pop. Ecol.* 12(1):111–125.
- . 1972. Investigation into the edge effect by use of capture-recapture data in a vole population. *Res. Pop. Ecol.* 13(2):127–151.
- , AND M. KANAMORI. 1967. New regression formula to estimate the whole population for recapture-addicted mammals. *Res. Pop. Ecol.* 9(1):83–94.
- TANTON, M. T. 1965. Problems of live-trapping and population estimation for the wood mouse, *Apodemus sylvaticus* (L.). *J. Anim. Ecol.* 34(1):1–22.
- WHITE, E. G. 1975. Identifying population units that comply with capture-recapture assumptions in an open community of alpine grasshoppers. *Res. Pop. Ecol.* 16(2):153–187.
- YOUNG, H., J. NEESS, AND J. T. EMLEN, JR. 1952. Heterogeneity of trap response in a population of house mice. *J. Wildl. Manage.* 16(2):169–180.
- ZIPPIN, C. 1956. An evaluation of the removal method of estimating animal populations. *Biometrics* 12(2):163–189.
- . 1958. The removal method of population estimation. *J. Wildl. Manage.* 22(1):82–90.

APPENDIX A

Notes on Estimation

Justification for the Use of Maximum Likelihood Estimation

Our philosophy throughout this work has been to present a variety of explicit mathematical models for capture data based upon fully specified assumptions. Given these models that incorporate various types of variability in capture probabilities (time, behavior, and heterogeneity), we are concerned with optimal estimation of population size, N , under each model. To achieve that optimality, we turn to the field of mathematical statistics.

A number of very general approaches to optimal statistical inference have been developed during the past 50 years. We feel that not all of them are suitable for use by biologists for the problem at hand (e.g., decision theoretic, or Bayesian approaches would require inputs from the biologist that we doubt they would be willing, or able, to supply). We have chosen to use the method of maximum likelihood (ML) estimation and inference; it is one of the best developed, omnibus tools of mathematical statistics.

The properties of ML estimation are well known (see Mood et al. 1974). For many practical models, ML estimators are optimal in many desirable ways. Because of its general applicability and good properties, ML estimation has been the basis for most modern methods of estimating parameters from animal marking experiments of all kinds (Seber 1973).

The only requirement for application of ML estimation is that one have a well-defined parametric model with fewer parameters than the dimension of the vector of minimal sufficient statistics. Thus, we have been able to use ML estimation for Models M_o , M_t , M_b , M_{bh} , and the removal models, but not for Model M_h . Only for Model M_h do we hesitate to claim there cannot be significant improvements in the estimator.

It is not necessary to understand the mechanics of how a ML estimate is computed; it suffices to understand this is not an ad hoc technique, but rather a well-developed inference method.

For any of the models dealt with here, there is a sampling probability distribution for the basic data (the $\{X_{ij}\}$) that can be represented as a mathematical function of appropriate parameters; symbolically it is

$$P\{X_{ij}|N,\mathbf{p}\},$$

for $N \geq M_{t+1}$, and with \mathbf{p} representing a set of capture probability parameters (e.g., $\mathbf{p} = (p,c)$ under Model M_b). Given an actual sample, we can substitute these data for the X_{ij} variables and treat this formula as strictly a function of the parameters N and \mathbf{p} . This function (of N,\mathbf{p}) is called the likelihood function. The ML estimators \hat{N} and $\hat{\mathbf{p}}$ are those values of N and \mathbf{p} which maximize the function

$$L(N,\mathbf{p}) = P\{X_{ij}|N,\mathbf{p}\}$$

or, equivalently, which maximize the log of $L(N,\mathbf{p})$. In this way, the problem of deriving parameter estimators is reduced to the classic problem of maximizing a given function over a set of possible values of specified variables.

The theory of ML estimation goes much deeper than just deriving point estimators of parameters. It also gives methods for deriving approximate sampling variances for the estimator and provides justification for the usual procedure for confidence interval construction. Also, there is a strong tie between ML theory and the theory of minimal sufficient statistics.

A sufficient statistic is a condensation of the sample into a smaller number of statistics that still contain all the information there is about N and \mathbf{p} . For example, M_{t+1} and n . are the minimal sufficient statistics for N and \mathbf{p} under Model M_0 (representing quite a condensation of the entire sample of X_{ij} values). The ML estimator is always a function of the sufficient statistic, even if one has not explicitly determined the sufficient statistic. Partly because ML estimators are functions of the sufficient statistic, they have, for large samples, the smallest possible sampling variances in the class of consistent estimators. Stated differently, ML estimators are (asymptotically) the most efficient estimators possible under a sampling theory approach to inference. For all the above reasons, we have used likelihood theory as the main basis of estimation in this monograph.

Numerical Methods

Except for the special case of only 2 trapping occasions ($t = 2$), the ML estimators described in the preceding chapters do not have explicit mathematical expressions. Hence, numerical methods must be used to maximize the likelihood functions and thereby find the population estimates on a case by case basis. Conceptually, this turns out to be easy for models to which we can apply ML estimation. From above, the likelihood function is representable simply as $L(N, \mathbf{p})$. In all applicable cases (exclude Model M_h and models where estimation is not possible), we find there is a closed form expression for the value of the capture probabilities which maximize the function $L(N, \mathbf{p})$ for any fixed $N \geq M_{t+1}$. Represent this as $\hat{\mathbf{p}}(N)$. Thus we can write

$$\begin{aligned} \max_N \max_{\mathbf{p}} L(N, \mathbf{p}) &= \max_N L(\hat{\mathbf{p}}(N) | N) \\ &= \max_N \lambda(N), \end{aligned}$$

where $\lambda(N)$ is a function (analytically expressible) only of N . The maximization of $\lambda(N)$ over N must be done numerically, but this is a simple one-dimensional search.

An algorithm by Brent (1973) (also see Fletcher 1972) that does not require analytic derivatives is used by program CAPTURE to maximize $\lambda(N)$. The algorithm searches between 2 end points. Those end points are first taken as the number of animals captured, M_{t+1} , for the lower bound and a linear approximation to \hat{N} plus M_{t+1} for the upper bound. If the upper bound is determined to be the maximum for the range specified, a new set of values lying next to the old values is selected and the search over N continues in this fashion until a true maximum is found. The algorithm searches along the real line although only integer values are appropriate. Because the search is one dimensional, the algorithm is very efficient.

The algorithm does not require the analytic first derivatives of $\lambda(N)$. This means it is not necessary to evaluate the derivative of the log-gamma function with respect to $N(N!$ and $(N - M_{t+1})!$ appear in all the likelihood functions).

Integer \hat{N} and Confidence Bounds

The domain of \hat{N} includes only integer values. Therefore, the integer that produces a maximum in the likelihood function is taken as the estimate. The numerical search is conducted along the real number line; then the value returned is truncated to an integer, and a check is made to see that it is the estimate, and not $\hat{N} + 1$. A somewhat arbitrary rounding to the nearest whole integer is required for the jackknife estimator to maintain consistency with the other estimators.

For all estimators, the variance is computed on the basis of the value of \hat{N} before it is rounded to an integer. We believe this procedure produces a slightly better estimator of $\text{Var}(\hat{N})$, and at the same time avoids numerical problems that would occur when \hat{N} equals M_{t+1} (this situation does occur).

Confidence intervals of 95 percent are computed as $\hat{N} \pm 1.96 \hat{SE}(\hat{N})$. Confidence intervals (but not standard errors) are computed around the integer value of \hat{N} . The lower bound is then truncated to an integer and the upper bound rounded upward to an integer. Those interval end points are thus technically outside the true 95 percent confidence interval. When the lower bound is less than the number of animals captured, the value could be reset to M_{t+1} , a realistic procedure because we know that N is not less than M_{t+1} . The coverage of the interval is not changed by this procedure of using M_{t+1} for the lower bound when the calculated lower bound is less than M_{t+1} .

Admittedly, a confidence interval procedure that can yield a lower bound less than the number of distinct animals captured (M_{t+1}) is not desirable. We investigated alternatives that avoid this problem, but we judged them even less appropriate than the simple approach described above (see Appendix O).

Estimation of Sampling Variance

Maximum likelihood theory includes a general method for obtaining the large sample (i.e., asymptotic) formula for the true sampling variance of \hat{N} . We have denoted this simply as $\text{Var}(\hat{N})$, thus suppressing the fact that it is an approximation to the "true" sampling variance of \hat{N} . The approximation is generally good for large samples; unfortunately, in the capture-recapture context we have no good measure of when a sample is sufficiently large. A further complication enters because the formula for $\text{Var}(\hat{N})$ is almost always a function of the unknown parameters N and \mathbf{p} ; symbolically it is generally of the form

$$\text{Var}(\hat{N}) = Ng(\mathbf{p}),$$

with the form of the function g known. But because N and \mathbf{p} are not known we must estimate $\text{Var}(\hat{N})$ by

$$\hat{\text{Var}}(\hat{N}) = \hat{N}g(\hat{\mathbf{p}}).$$

The properties of $\hat{\text{Var}}(\hat{N})$ can be different (sometimes quite so) from those of $\text{Var}(\hat{N})$, which itself is only an approximation to the appropriate finite sampling variance of \hat{N} . One particular problem that has been noted in the literature concerns the fact that \hat{N} and $\hat{SE}(\hat{N}) = \sqrt{\hat{\text{Var}}(\hat{N})} = \sqrt{\hat{N}g(\hat{\mathbf{p}})}$ can be substantially correlated. This is not surprising, because, of course, \hat{N} and $\sqrt{\hat{N}}$ are highly correlated. The effect of this correlation is to cause an underestimate of $SE(\hat{N})$ when \hat{N} is lower than the true N (which it will be much

of the time). This, in turn, causes the confidence intervals in such cases to be shorter than they should be. An area suitable for further research on capture–recapture statistics (both closed and open models) is the question of improved estimation of sampling variances *and* confidence intervals.

APPENDIX B

Estimation in Model M₀

In Model M₀, parameterized by the parameters N and p, the relevant part of the log-likelihood function is given by

$$\ln L(N, p | \mathbf{X}) = \ln \left(\frac{N!}{(N - M_{t+1})!} \right) + (n.) \ln(p) + (tN - n.) \ln(1 - p),$$

where $p \in [0, 1]$ and $N \in \mathcal{N} = \{M_{t+1}, M_{t+1} + 1, M_{t+1} + 2, \dots\}$. Given the value of N, the ML estimator $\hat{p}(N)$ of p is given as the solution to

$$\frac{\partial}{\partial p} \ln L(p | N, \mathbf{X}) = 0,$$

which reduces to

$$\frac{n.}{\hat{p}(N)} = \frac{tN - n.}{1 - \hat{p}(N)}.$$

This results in the solution

$$\hat{p}(N) = \frac{n.}{tN}.$$

Now, the ML estimator \hat{N}_0 of N satisfies

$$\begin{aligned} \ln L(\hat{N}_0, \hat{p}(\hat{N}_0) | \mathbf{X}) &= \max_{N \in \mathcal{N}} \left[\max_{p \in [0, 1]} \ln L(p | N, \mathbf{X}) \right] \\ &= \max_{N \in \mathcal{N}} \left[\ln L(\hat{p}(N) | N, \mathbf{X}) \right] \\ &= \max_{N \in \mathcal{N}} \left[\ln \left(\frac{N!}{(N - M_{t+1})!} \right) + (n.) \ln(n.) \right. \\ &\quad \left. + (tN - n.) \ln(tN - n.) - tN \ln(tN) \right]. \end{aligned}$$

For a given data set, a search over \mathcal{N} is performed to locate the ML estimate \hat{N}_0 . This value is then used in the calculation of the ML estimate $\hat{p}(\hat{N}_0) = \hat{p}$ via

$$\hat{p} = \frac{n.}{t\hat{N}_0}.$$

The asymptotic variance of \hat{N}_0 is (cf. Darroch 1959)

$$\text{Var}(\hat{N}_0) = N[(1 - p)^{-t} - t(1 - p)^{-1} + t - 1]^{-1}.$$

An estimate of this variance is

$$\hat{\text{Var}}(\hat{N}_0) = \hat{N}_0[(1 - \hat{p})^{-t} - t(1 - \hat{p})^{-1} + t - 1]^{-1}.$$

APPENDIX C

Estimation in Model M_t

Model M_t is parameterized by the $t + 1$ parameters N, p_1, p_2, \dots, p_t . The relevant log-likelihood function for estimation of the parameters is given by

$$\ln L(N, \mathbf{p} | \mathbf{X}) = \ln \left(\frac{N!}{(N - M_{t+1})!} \right) + \sum_{j=1}^t n_j \ln(p_j) \\ + \sum_{j=1}^t (N - n_j) \ln(1 - p_j),$$

where $N \in \mathcal{N} = \{M_{t+1}, M_{t+1} + 1, M_{t+1} + 2, \dots\}$, $\mathbf{p} = \{p_1, p_2, \dots, p_t\}$, $p_j \in [0, 1]$ for $j = 1, 2, \dots, t$. Given the value of N , the ML estimators $\hat{p}_j(N)$ of p_j are given as the solutions to the system of equations

$$\frac{\partial}{\partial p_j} \ln L(\mathbf{p} | N, \mathbf{X}) = 0, \quad j = 1, 2, \dots, t.$$

The j^{th} one of these equations reduces to

$$\frac{n_j}{\hat{p}_j(N)} = \frac{N - n_j}{1 - \hat{p}_j(N)},$$

which results in the solution

$$\hat{p}_j(N) = \frac{n_j}{N}.$$

Now, the ML estimator \hat{N}_t of N satisfies

$$\ln L(\hat{N}_t, \hat{p}_1(\hat{N}_t), \dots, \hat{p}_t(\hat{N}_t) | \mathbf{X}) = \max_{N \in \mathcal{N}} \left[\max_{p_j \in [0, 1]} \ln L(p_1, p_2, \dots, p_t | N, \mathbf{X}) \right] \\ = \max_{N \in \mathcal{N}} \left[\ln L(\hat{p}_1(N), \hat{p}_2(N), \dots, \hat{p}_t(N) | N, \mathbf{X}) \right] \\ = \max_{N \in \mathcal{N}} \left[\ln \left(\frac{N!}{(N - M_{t+1})!} \right) + \sum_{j=1}^t n_j \ln(n_j) \right. \\ \left. + \sum_{j=1}^t (N - n_j) \ln(N - n_j) - tN \ln(N) \right].$$

A search over \mathcal{N} is performed in order to locate the ML estimator \hat{N}_t . It is then possible to calculate the ML estimates $\hat{p}_j(\hat{N}_t) = \hat{p}_j$ of the p_j for $j = 1, 2, \dots, t$ via

$$\hat{p}_j = \frac{n_j}{\hat{N}_t}.$$

The asymptotic variance of \hat{N}_t is given by Darroch (1958) as:

$$\text{Var}(\hat{N}_t) = N \left[\frac{1}{\prod_{j=1}^t (1 - p_j)} + t - 1 - \sum_{j=1}^t (1 - p_j)^{-1} \right]^{-1}.$$

An estimate of this variance is

$$\hat{\text{Var}}(\hat{N}_t) = \hat{N}_t \left[\frac{1}{\prod_{j=1}^t (1 - \hat{p}_j)} + t - 1 - \sum_{j=1}^t (1 - \hat{p}_j)^{-1} \right]^{-1}.$$

APPENDIX D

Estimation in Model M_b

Model M_b is parameterized by the parameters N, p , and c . The part of the log-likelihood necessary for estimation of the parameters is given by

$$\begin{aligned} \ln L(N, p, c | \mathbf{X}) = & \ln \left(\frac{N!}{(N - M_{t+1})!} \right) + M_{t+1} \ln(p) + (tN - M. - M_{t+1}) \ln(1 - p) \\ & + m. \ln(c) + (M. - m.) \ln(1 - c), \end{aligned}$$

where $N \in \mathcal{N} = \{M_{t+1}, M_{t+1} + 1, M_{t+1} + 2, \dots\}$, $p \in [0, 1]$, $c \in [0, 1]$. The ML estimator \hat{c} of c is produced by the equation

$$\frac{\partial}{\partial c} \ln L(N, p, c | \mathbf{X}) = 0,$$

which reduces to

$$\frac{m.}{\hat{c}} = \frac{M. - m.}{1 - \hat{c}}.$$

Solving this equation gives $\hat{c} = m./M.$. Thus, we see that estimation of c is independent of the estimation of N and p . Now, the relevant part of the log-likelihood function for purposes of estimating N and p is

$$\ln L(N, p | \mathbf{X}) = \ln \left(\frac{N!}{(N - M_{t+1})!} \right) + M_{t+1} \ln(p) + (tN - M. - M_{t+1}) \ln(1 - p).$$

Given the value of N , the ML estimator $\hat{p}(N)$ of p is provided by the equation

$$\frac{\partial}{\partial p} \ln L(p | N, \mathbf{X}) = 0,$$

which reduces to

$$\frac{M_{t+1}}{\hat{p}(N)} = \frac{tN - M. - M_{t+1}}{1 - \hat{p}(N)}.$$

The solution to this equation gives

$$\hat{p}(N) = \frac{M_{t+1}}{tN - M.}.$$

Now, the ML estimator \hat{N}_b of N satisfies

$$\begin{aligned} \ln L(\hat{N}_b, \hat{p}(\hat{N}_b) | \mathbf{X}) &= \max_{N \in \mathcal{N}} \left[\max_{p \in [0, 1]} \ln L(p | N, \mathbf{X}) \right] \\ &= \max_{N \in \mathcal{N}} \left[\ln L(\hat{p}(N) | N, \mathbf{X}) \right] \end{aligned}$$

$$= \max_{N \in \mathcal{N}} \left[\ln \left(\frac{N!}{(N - M_{t+1})!} \right) + M_{t+1} \ln(M_{t+1}) + (tN - M. - M_{t+1}) \ln(tN - M. - M_{t+1}) - (tN - M.) \ln(tN - M.) \right].$$

A search over \mathcal{N} is preformed to locate the ML estimate \hat{N}_b . It is then possible to calculate the ML estimate $\hat{p}(\hat{N}_b) = \hat{p}$ of p via

$$\hat{p} = \frac{M_{t+1}}{t\hat{N}_b - M.}$$

The asymptotic variance of \hat{N}_b is given by Zippin (1956) as:

$$\text{Var}(\hat{N}_b) = \frac{N(1 - p)^t [1 - (1 - p)^t]}{[1 - (1 - p)^t]^2 - t^2 p^2 (1 - p)^{t-1}}$$

An estimate of this variance is given by

$$\hat{\text{Var}}(\hat{N}_b) = \frac{\hat{N}_b(1 - \hat{p})^t [1 - (1 - \hat{p})^t]}{[1 - (1 - \hat{p})^t]^2 - t^2 \hat{p}^2 (1 - \hat{p})^{t-1}}$$

APPENDIX E

Estimation in Model M_h

The mathematical details of Model M_h are covered in detail in Burnham (unpublished dissertation) and Burnham and Overton (pers. comm.). The following gives a few of the basic results for this model and the jackknife estimator.

Under Model M_h we assume

$$p_{ij} = p_i \quad i = 1, \dots, N,$$

and we further assume that the p_1, \dots, p_N are a random sample from some (unknown) probability distribution $F(p)$, $p \in [0, 1]$. For any $F(p)$, the MSS contains only the capture frequencies f_1, \dots, f_t , and their distribution is multinomial:

$$P\{f_1, \dots, f_t | F\} = \binom{N}{N - M_{t+1}, f_1, \dots, f_t} (\pi_0)^{N - M_{t+1}} \prod_{j=1}^t (\pi_j)^{f_j},$$

where

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

If one assumes a parametric form for $F(p)$, then standard approaches are possible (e.g., ML). Burnham (unpublished dissertation) investigated the case where $F(p)$ is the class of beta distributions and found that approach basically unacceptable.

The "jackknife" estimator used here was developed by application of the generalized jackknife statistic (Gray and Schucany 1972) to the naive estimator M_{t+1} , assuming the bias of M_{t+1} as an estimator of N is expressible

TABLE E.1.—THE JACKKNIFE ESTIMATORS \hat{N}_{hk} OF POPULATION SIZE, FOR $k = 1$ TO 5

$$\hat{N}_{h1} = M_{t+1} + \left(\frac{t-1}{t}\right)f_1$$

$$\hat{N}_{h2} = M_{t+1} - \left(\frac{2t-3}{t}\right)f_1 - \frac{(t-2)^2}{t(t-1)}f_2$$

$$\hat{N}_{h3} = M_{t+1} + \left(\frac{3t-6}{t}\right)f_1 - \left(\frac{3t^2-15t+19}{t(t-1)}\right)f_2 + \frac{(t-3)^3}{t(t-1)(t-2)}f_3$$

$$\hat{N}_{h4} = M_{t+1} + \left(\frac{4t-10}{t}\right)f_1 - \left(\frac{6t^2-36t+55}{t(t-1)}\right)f_2 + \left(\frac{4t^3-42t^2+148t-175}{t(t-1)(t-2)}\right)f_3 - \frac{(t-4)^4}{t(t-1)(t-2)(t-3)}f_4$$

$$\hat{N}_{h5} = M_{t+1} + \left(\frac{5t-15}{t}\right)f_1 - \left(\frac{10t^2-70t+125}{t(t-1)}\right)f_2 + \left(\frac{10t^3-120t^2+485t-660}{t(t-1)(t-2)}\right)f_3$$

$$- \left(\frac{(t-4)^5-(t-5)^5}{t(t-1)(t-2)(t-3)}\right)f_4 + \frac{(t-5)^5}{t(t-1)(t-2)(t-3)(t-4)}f_5$$

in a power series in $1/t$. The end result of that application is that for the k^{th} order jackknife the estimator is a linear function of the capture frequencies. For example, we have

$$\hat{N}_{h1} = M_{t+1} + \left(\frac{t-1}{t}\right)f_1 \equiv \left(1 + \frac{t-1}{t}\right)f_1 + \sum_{j=2}^t f_j$$

These point estimators of N have been determined for up to the fifth order, and are given in Table E.1.

Because

$$M_{t+1} = \sum_{j=1}^t f_j,$$

any \hat{N}_{hk} is expressible as a linear combination of the capture frequencies, say as

$$\hat{N}_{hk} = \sum_{j=1}^t a_{jk}f_j.$$

Using the fact that the f_j are multinomial random variables an approximate variance estimator of \hat{N}_{hk} is

$$\hat{\text{Var}}(\hat{N}_{hk}) = \sum_{j=1}^t (a_{jk})^2 f_j - \hat{N}_{hk},$$

and confidence intervals can be constructed on the basis of the asymptotic normality of \hat{N}_{hk} .

A procedure for selecting one of the estimators has been suggested by Burnham and Overton (pers. comm.). It involves testing whether $\hat{N}_{h,k+1}$ is significantly different from \hat{N}_{hk} and stopping when no significant difference is found.

APPENDIX F

Discussion of Model M_{tb}

The most general model one might assume in the instance where both behavioral response to capture and variability among trapping occasions are found involves the following set of parameters:

- N = population size
 p_j = probability of capture on trapping occasion j of an animal not previously captured, $j = 1, 2, \dots, t$,
 c^*_{kj} = probability of capture on trapping occasion j of an animal first captured on occasion k , $j = k + 1, k + 2, \dots, t$; $k = 1, \dots, t - 1$.

The corresponding probability distribution of $\{X_\omega\}$ is given by

$$P[\{X_\omega\}] = \frac{N!}{\left[\prod_{\omega} X_\omega! \right] (N - M_{t+1})!} \left[\prod_{j=1}^t p_j^{u_j} (1 - p_j)^{N - M_{t+1}} \right] \\ \cdot \left[\prod_{k=1}^{t-1} \prod_{j=k+1}^t c^*_{kj} R_{kj} (1 - c^*_{kj})^{u_k - R_{kj}} \right],$$

where

- u_j = number of animals first caught on occasion j , $j = 1, 2, \dots, t$,
 M_j = number of marked (previously captured) animals in the population at the time of the j^{th} sample, $j = 2, 3, \dots, t$,
 M_{t+1} = total number of *different* animals captured in the experiment
 (notice $M_{t+1} = \sum_{j=1}^t u_j$),

R_{kj} = number of animals caught on occasion j that were first captured on occasion k , $j = k + 1, k + 2, \dots, t$; $k = 1, \dots, t - 1$.
 A MSS for this distribution is $\{u_1, u_2, \dots, u_t, R_{12}, R_{13}, \dots, R_{1t}, R_{23}, \dots, R_{2t}, \dots, R_{t+1,t}\}$, that has dimension $t(t+1)/2$. There are $t(t+1)/2 + 1$ parameters involved and straightforward ML estimation of N is not possible.

If we restrict the model so that $c^*_{kj} = c_j$, i.e., the probability of recapture on occasion j does not depend on the occasion of first capture, then Model M_{tb} (as previously defined) results. It is easily verified that, when $c^*_{kj} = c_j$, $P[\{X_\omega\}]$ given above reduces to the probability distribution given in the section concerning Model M_{tb} where a discussion of the nonidentifiability of N is presented. Thus, for purposes of estimating N , the assumption that $c^*_{kj} = c_j$ has no utility.

ML estimation of N is theoretically possible if one is willing to further restrict the model so that there is some relationship between c^*_j and p_j . There is no unique restriction, and as an example we consider $c^*_j = \theta p_j$, $j = 2, 3, \dots, t$, where $0 \leq p_j \leq 1$, $j = 1, 2, \dots, t$ and $0 \leq \theta p_j \leq 1$, $j = 2, 3, \dots, t$. With this assumption, the probability distribution of $\{X_\omega\}$ becomes

$$P[\{X_\omega\}] = \frac{N!}{\prod_{\omega} X_\omega! (N - M_{t+1})!} \left[\prod_{j=1}^t p_j^{u_j} (1 - p_j)^{N - M_{t+1}} \right] \\ \cdot \left[\prod_{j=2}^t (\theta p_j)^{m_j} (1 - \theta p_j)^{M_j - m_j} \right]$$

$$= \frac{N!}{\prod_{\omega} X_{\omega}!(N - M_{t+1})!} p_1^{u_1}(1 - p_1)^{N-u_1}(\theta^m) \cdot \prod_{j=2}^t (1 - p_j)^{N-M_{t+1}}(1 - \theta p_j)^{M_t - m_t} p_j^{n_j}.$$

A sufficient statistic for this distribution is $\{u_1, u_2, \dots, u_t, m_2, m_3, \dots, m_t\}$, that has dimension $2t - 1$. Since there are $t + 2$ parameters involved, the model allows all parameters, in particular N , to be identified if $t \geq 3$. Estimation under this model is not considered here, however, due to the questionable assumption that recapture probabilities bear a constant relationship to initial capture probabilities over all trapping occasions, for all animals. Thus, at present we are not able to present an estimator of N that is appropriate under the assumptions of Model M_{th} .

APPENDIX G

Discussion of Model M_{th}

In the section concerned with Model M_n , the random variable X_{ij} was stated to be distributed according to a Bernoulli distribution with parameter p_{ij} , that represents the probability that the i^{th} animal is captured on the j^{th} occasion. Here, we assume that the elements of the set $\{X_{ij}\}$ constitute a mutually independent collection of random variables, and that $p_{ij} = p_i p_j$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, t$. Furthermore, we let the p_i , $i = 1, 2, \dots, N$ be a random sample of size N from some probability distribution function $F(p; \theta)$ defined on $[0, 1]$, and we restrict the p_j , $j = 1, 2, \dots, t$ so that $0 \leq p_i p_j \leq 1$ for all i, j . Under those conditions, the probability distribution of the observed sample $\{X_{ij}\}$, where $i = 1, 2, \dots, M_{t+1}$, $j = 1, 2, \dots, t$ can be written as

$$P[\{X_{ij}\}] = P[\{X_{ij}\} | M_{t+1}] P[M_{t+1}],$$

where

$$P[\{X_{ij}\} | M_{t+1}] = \left(\prod_{j=1}^t p_j^{n_j} \right) \left(\prod_{i=1}^{M_{t+1}} \int_0^1 p^{y_i} \left[\prod_{j=1}^t (1 - p p_j)^{1 - X_{ij}} \right] dF(p; \theta) \right),$$

y_i = the number of times the i^{th} animal is captured,

$P[M_{t+1}]$ = the probability distribution of the number of different animals captured in the experiment. This distribution will involve the parameters N, p_1, p_2, \dots, p_t , and the distribution $F(p; \theta)$.

Obviously, if the form of $F(p; \theta)$ is left unspecified, ML estimation of N will not be possible. If the form of $F(p; \theta)$ is specified, but the vector of parameters θ is not, likelihood estimation will be possible if the dimension of the minimal sufficient statistic is sufficiently large to ensure identifiability of N . As mentioned in the section on Model M_{th} , estimation of N will be possible if $F(p; \theta)$ is completely specified. In that case, a MSS of dimension $t + 1$ is $\{n_1, n_2, \dots, n_t, M_{t+1}\}$, and the number of parameters to be estimated is also $t + 1$, i.e., p_1, p_2, \dots, p_t, N . Therefore, unless the experimenter is willing to make the doubtful assumption that at least the form of $F(p; \theta)$ is known, we can present no satisfactory estimation procedure appropriate under the assumptions of Model M_{th} .

APPENDIX H

Estimation in Model M_{bh}

Pollock (unpublished dissertation) considered Model M_{bh} and found the distribution of the set of possible capture histories $\{X_\omega\}$ under the assumption that (p_i, c_i) , which represents the pair of initial and recapture probabilities, respectively, of the i^{th} animal, is the i^{th} member of a random sample of size N from some bivariate distribution function $G(p, c; \theta)$. Let us consider rewriting the distribution $P[\{X_\omega\} | G(p, c; \theta)]$ as $P[\{u_1, u_2, \dots, u_t\} | G(p, c; \theta)] \cdot P[\{X_\omega\} | \{u_1, u_2, \dots, u_t\}, G(p, c; \theta)]$. The distribution of the removals $P[\{u_1, u_2, \dots, u_t\} | G(p, c; \theta)]$ will be multinomial, with parameter N and cell probabilities that are functions of the moments of $G(p, c; \theta)$. The conditional distribution $P[\{X_\omega\} | \{u_1, u_2, \dots, u_t\}, G(p, c; \theta)]$ will also depend on the moments of $G(p, c; \theta)$ but not on the parameter N . Therefore, the multinomial distribution

$$P[\{u_1, u_2, \dots, u_t\} | G(p, c; \theta)] = \frac{N!}{\left[\prod_{j=1}^t u_j! \right] (N - M_{t+1})!} \left(\prod_{j=1}^t \pi_j^{u_j} \right) (\pi_{t+1})^{N - M_{t+1}},$$

where π_j is a function of the moments of $G(p, c; \theta)$ and

$$\pi_{t+1} = 1 - \sum_{j=1}^t \pi_j,$$

is the relevant distribution for purposes of estimating N . Pollock (unpublished dissertation) showed that if $G(p, c; \theta) = G_1(p; \theta_1) \cdot G_2(c; \theta_2)$, then $\pi_j = E[p(1 - p)^{j-1}]$ and the conditional distribution $P[\{X_\omega\} | \{u_1, u_2, \dots, u_t\}, G(p, c; \theta)]$ depends only upon $G(c; \theta_2)$. Regardless of whether p and c are independent, once we agree to base estimation of N solely on the first capture "removal" type of data, then the problem can be reformulated. The behavioral response is then irrelevant and all we need to consider is the probability distribution of first captures. That is, we assume that p_1, \dots, p_N are a random sample from some distribution $G_1(p; \theta)$.

Consider transforming the parameters $\pi_1, \pi_2, \dots, \pi_t$ into the set of parameters $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_t$ by using the relationship $\pi_j = (1 - \bar{p}_1)(1 - \bar{p}_2) \dots (1 - \bar{p}_{j-1})\bar{p}_j$, $j = 1, 2, \dots, t$. Thus, \bar{p}_j is a conditional probability that represents the average first capture probability of those members of the population that have not yet been captured at the time of the j^{th} trapping occasion. Given this interpretation, it is not unreasonable to assume that $\bar{p}_1 > \bar{p}_2 > \dots > \bar{p}_t$. Furthermore, we assume that $(\bar{p}_1 - \bar{p}_2) > (\bar{p}_2 - \bar{p}_3) > \dots > (\bar{p}_{t-1} - \bar{p}_t)$, so that larger differences in average first capture probability occur initially. (Note: if $G_1(p; \theta)$ is the class of beta distributions both assumptions are easily shown to be true.) The assumptions and the distribution $P[\{u_1, u_2, \dots, u_t\} | G(p, c; \theta)]$, which we abbreviate $P[u_1, u_2, \dots, u_t]$, form the basis of the generalized removal method outlined below.

For $k = 1, 2, \dots, t - 2$,

- (i) Assume $\bar{p}_k = \bar{p}_{k+1} = \dots = \bar{p}_t = \bar{p}$ and that $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_{k-1}$ differ. This reduces the number of parameters involved in the estimation of N to $k + 1$.
- (ii) Estimate N by the ML method. This task is greatly simplified by rewriting

$$P[u_1, u_2, \dots, u_t] \text{ as } \prod_{j=1}^t P[u_j | u_1, u_2, \dots, u_{j-1}], \text{ where } P[u_j | u_1, u_2, \dots, u_{j-1}]$$

is the conditional distribution of the j^{th} removal, given the values of the previous removals.

Then

- (iii) Choose the smallest value of k that produces a sufficient fit to the data, and take as the estimate of N the estimate associated with this value of k . The fit of the data u_1, u_2, \dots, u_t can be measured by the usual size α chi-square goodness of fit test. We have used $\alpha = 0.20$ because of the seriousness of Type II errors. If significance levels of all the tests (one for each value of k) are less than 0.20 we have chosen the value of k corresponding to the largest achieved significance level.

For a given value of k , the asymptotic sampling variance of \hat{N}_{bh} is

$$\text{Var}(\hat{N}_{bh}) = N \left\{ \frac{\sum_{j=1}^t \pi_j}{1 - \sum_{j=1}^t \pi_j} - \frac{\sum_{j=1}^{k-1} \bar{p}_j^2}{\sum_{j=1}^{k-1} \pi_j (1 - \bar{p}_j)} - \frac{(t - k + 1)^2 \bar{p}^2}{(1 - \bar{p}) \left[\prod_{j=1}^{k-1} (1 - \bar{p}_j) + \sum_{j=1}^t \pi_j - 1 \right]} \right\}^{-1}$$

An estimate $\hat{\text{Var}}(\hat{N}_{bh})$ of $\text{Var}(\hat{N}_{bh})$ is obtained by replacing $N, \bar{p}, \bar{p}_1, \bar{p}_2, \dots, \bar{p}_{k-1}, \pi_1, \pi_2, \dots, \pi_t$ by their respective ML estimates.

Finally, we mention that, for a given value of k , it is possible that the experiment “fails,” i.e., valid estimation of N is not possible. Recall that Seber and Whale (1970) provided a failure criterion for the estimator associated with Model M_b . (This model corresponds to the case $k = 1$.) Following their method of proof, it is easily shown that the failure criterion for any value of k is

$$\sum_{j=k}^t (t + k - 2j)u_j \leq 0.$$

If the experiment fails for a value of k , the corresponding model is clearly excluded from those eligible for selection as the appropriate model for estimating N .

APPENDIX I

Discussion of Model $M_{t,bh}$

In Model $M_{t,bh}$, it is assumed that the i^{th} animal in the population has, on the j^{th} trapping occasion, both a unique probability of first capture p_{ij} and a unique probability of recapture c_{ij} . The model therefore requires tN parameters concerning first capture, $(t - 1)N$ parameters concerning recapture (because $c_{i1} = c_{21} = \dots = c_{N1} = 0$), and the parameter N for its complete specification. This totals $(2t - 1)N + 1$ parameters and obviously all parameters are not identifiable for estimation purposes.

The assumption can be made that the $2t - 1$ dimensional vectors $(p_{11}, p_{12}, \dots, p_{1t}, c_{12}, c_{13}, \dots, c_{1t}), (p_{21}, p_{22}, \dots, p_{2t}, c_{22}, c_{23}, \dots, c_{2t}), \dots, (p_{N1}, p_{N2}, \dots, p_{Nt}, c_{N2}, \dots, c_{Nt})$ are a random sample of size N from some probability distribution function $F(p_1, p_2, \dots, p_t, c_2, c_3, \dots, c_t; \theta)$ parameterized by the vector θ and de-

defined on $[0,1]^{2t-1}$. The probability distribution function of $\{X_\omega\}$ can then be written as

$$P[\{X_\omega\}] = \frac{N!}{\left[\prod_\omega X_\omega!\right](N - M_{t+1})!} \prod_\omega \left[E[\pi_\omega]\right]^{X_\omega},$$

where π_ω = a scalar random variable that is a function of the $2t - 1$ dimensional random variable $(p_1, p_2, \dots, p_t, c_2, c_3, \dots, c_t)$ corresponding to the capture history ω , and

$$E[\pi_\omega] = \int_0^1 \cdots \int_0^1 \pi_\omega \, dF(p_1, \dots, p_t, c_2, \dots, c_t; \theta).$$

For instance, if $t = 4$ and $\omega = \{1, 1, 0, 1\}$, then $\pi_\omega = p_1 c_2 (1 - c_3) c_4$. If one is willing to make certain assumptions concerning the family of distributions to which $F(\cdot; \theta)$ belongs, the dependence structure among the variables $\{p_1, p_2, \dots, p_t, c_2, c_3, \dots, c_t\}$ and the vector of parameters θ , then ML estimation of N will theoretically be possible. However, it is not unreasonable to suspect that the amount of numerical computation and the number of assumptions required will prohibit such estimates from being useful in practice.

APPENDIX J

Estimation in Removal Models

For the removal experiment, it is assumed there may be heterogeneity among the capture probabilities of the N members of the population subject to removal. Moreover, the N capture (removal) probabilities $p_i, i = 1, 2, \dots, N$, are a random sample from some probability distribution function $G(p; \theta)$ defined on $[0,1]$ and parameterized by the vector θ . Under those assumptions, the distribution of the vector of removals $\{u_1, u_2, \dots, u_t\}$ is given by

$$P\{u_1, u_2, \dots, u_t\} = \frac{N!}{\left[\prod_{j=1}^t u_j!\right] (N - M_{t+1})!} \left(\prod_{j=1}^t \pi_j^{u_j}\right) \pi_{t+1}^{N - M_{t+1}},$$

where

$$\begin{aligned} \pi_j &= E[p(1 - p)^{j-1}] = \int_0^1 p(1 - p)^{j-1} \, dG(p; \theta), \quad j = 1, 2, \dots, t, \\ \pi_{t+1} &= E[(1 - p)^t] = \int_0^1 (1 - p)^t \, dG(p; \theta) \\ &= 1 - \sum_{j=1}^t E[p(1 - p)^{j-1}]. \end{aligned}$$

Notice that this distribution of $\{u_1, u_2, \dots, u_t\}$ is of the form, and the parameters of the same nature as the distribution of $\{u_1, u_2, \dots, u_t\}$ in Model M_{bh} . Therefore, the generalized removal method developed for Model M_{bh} is also applicable to removal experiments where estimation of N is desired. Although the experimental situations associated with those 2 models are quite different, the fact that the removals are the only statistics relevant for purposes of estimating N , and the nature of the parameters π_j combine to make the generalized removal method appropriate in both cases.

APPENDIX K

Tests of Model Assumptions

The details of the statistical tests described in the section entitled TESTS OF MODEL ASSUMPTIONS are given here. Tests have been numerically identified and correspond to the identification numbers used in program CAPTURE.

Test 1

Since Model M_0 is a special case of Model M_h , a likelihood ratio test of $H_0: p_i = p, i = 1, \dots, N$ versus $H_A: \text{all } p_i \text{ are not equal}$, seems plausible. However, due to the nonidentifiability of parameters in Model M_h a valid likelihood ratio procedure is not possible.

An alternative approach is taken by examining the fit of the observed frequencies f_1, \dots, f_t (recall these are the elements of the MSS for M_h) to their expected values under Model M_0 . The resulting test statistic is intended to be sensitive to departures from Model M_0 in the direction of Model M_h . If H_0 is true, we would expect the test statistic T_1 to be approximately distributed as a chi-square random variable with $t - 2$ degrees of freedom, where

$$T_1 = \sum_{j=1}^t \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j},$$

$$\hat{f}_j = \hat{N}_0 \binom{t}{j} \hat{p}^j (1 - \hat{p})^{t-j},$$

and \hat{N}_0 and \hat{p} are the ML estimates of N and p under Model M_0 .

Test 2

Testing the null hypothesis of Model M_0 versus the alternative of Model M_b is equivalent to testing the null hypothesis $H_0: p = c$ versus the alternative $H_A: p \neq c$. If it is assumed that the bivariate random variable $\{\hat{p} \hat{c}\}$ is distributed as a bivariate normal with mean vector $\{p \ c\}$ and covariance matrix

$$\begin{bmatrix} \text{Var}(\hat{p}) & 0 \\ 0 & \text{Var}(\hat{c}) \end{bmatrix},$$

then, under H_0 , the quantity $T_2' = (\hat{p} - \hat{c})^2 / [\text{Var}(\hat{p}) + \text{Var}(\hat{c})]$ is distributed as a chi-square random variable with 1 degree of freedom (cf. Theorem 4.4.5 in Graybill 1976). Here, \hat{p} and \hat{c} are the ML estimates of p and c under Model M_b (cf. Appendix D), and we use $\text{Var}(\hat{p}) = p^2 q^2 (1 - q) / N [q(1 - q)^2 - p^2 t^2 q^t]$, where $q = 1 - p$ (cf. Seber 1973:312). Furthermore, we approximate $\text{Var}(\hat{c})$ by treating \hat{c} as a binomial variable with M fixed so that $\text{Var}(\hat{c}) = c(1 - c) / M$. Obviously, both $\text{Var}(\hat{p})$ and $\text{Var}(\hat{c})$ will have to be estimated by substituting the ML estimates of N, p , and c under Model M_b , so that the actual test statistic becomes

$$T_2 = \frac{(\hat{p} - \hat{c})^2}{\hat{\text{Var}}(\hat{p}) + \hat{\text{Var}}(\hat{c})}.$$

It follows that T_2 has an approximate chi-square distribution with 1 degree of freedom under H_0 .

If H_0 is false, T_2 has an approximate noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter $\lambda = (p - c)^2/2[\text{Var}(\hat{p}) + \text{Var}(\hat{c})]$ (cf. Theorem 4.4.5 in Graybill 1976). Thus, theoretical approximations to the power of the above test can be found for fixed values of N , p , and c . These approximations, obtained by the use of the noncentral chi-square tables of Haynam et al. (1970), are given for the alternatives involved in Table 8, Appendix N. Finally, it should be mentioned that, due to the conditional nature of $\text{Var}(\hat{c})$, it was necessary to substitute

$$E[M.] = Np \sum_{k=1}^t \sum_{\ell=0}^{k-2} q^\ell$$

for $M.$ to enable evaluation of the noncentrality parameter λ .

Test 3

Testing the null hypothesis of Model M_0 versus the alternative of Model M_1 is equivalent to testing $H_0: p_j = p, j = 1, 2, \dots, t$, against H_A : Not all the p_j are equal. We assume that the t -variate random variable (p_1, p_2, \dots, p_t) is distributed as a t -variate normal distribution with mean vector (p_1, p_2, \dots, p_t) and covariance matrix

$$\begin{bmatrix} \text{Var}(\hat{p}_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \text{Var}(\hat{p}_t) \end{bmatrix}, \text{ where}$$

$\hat{p}_j = n_j/\hat{N}_t = \text{ML estimator of } p_j \text{ under Model } M_t$, and

$$\text{Var}(\hat{p}_j) = p_j q_j / N, \quad q_j = 1 - p_j, \quad j = 1, 2, \dots, t.$$

Now, using Theorem 4.4.5 in Graybill (1976), the random variable

$$T_3' = \sum_{j=1}^t \hat{p}_j^2 / \text{Var}(\hat{p}_j) - \left(\sum_{j=1}^t \hat{p}_j / \text{Var}(\hat{p}_j) \right)^2 / \sum_{j=1}^t \text{Var}(\hat{p}_j)^{-1}$$

has a chi-square distribution with $t - 1$ degrees of freedom under H_0 . As in test 2, the quantities $\text{Var}(\hat{p}_j)$ will have to be estimated using the ML estimates \hat{p}_j and \hat{N}_t . Upon making these substitutions the test statistic T_3' reduces to

$$T_3 = \sum_{j=1}^t n_j / \hat{q}_j - \left(\sum_{j=1}^t n_j / \hat{p}_j \hat{q}_j \right)^2 / \sum_{j=1}^t n_j / \hat{p}_j^2 \hat{q}_j.$$

Under H_0 , T_3 has an approximate chi-square distribution with $t - 1$ degrees of freedom.

An approximation to the power of this test given any alternative is provided by the knowledge that, under H_A , T_3 has an approximate noncentral chi-square distribution with $t - 1$ degrees of freedom and noncentrality parameter

$$\lambda = (N/2) \left[\sum_{j=1}^t p_j / q_j - \left(\sum_{j=1}^t 1 / q_j \right)^2 / \sum_{j=1}^t (p_j q_j)^{-1} \right].$$

Utilizing the tables of Haynam et al. (1970), these approximations were calculated for the alternatives involved in Table 9 of Appendix N, and the results are contained therein.

Test 4

An overall goodness of fit test of Model M_h can be thought of as equivalent to testing $H_0: p_{ij} = p_i$ versus $H_A: \text{not all } p_{ij} = p_i, i = 1, 2, \dots, M_{t+1}, j = 1, \dots, t$. Thus, the null hypothesis assumes heterogeneous individual capture probabilities that do not change over time, and the alternative states that, given heterogeneity, capture probabilities also change over time. Notice that H_A does not state how probabilities change over time, i.e., whether the change is due to behavioral response, variation in trapping occasions, or a combination of both. Thus, if H_0 is rejected, it is not rejected in favor of an alternative Model; rather it is the goodness of fit of Model M_h that is being rejected.

Burnham (unpublished dissertation) has shown that under H_0 the test statistic

$$T_4 = \frac{\sum_{j=1}^t (n_j - n./t)^2}{\sum_{j=1}^t f_j \left(\frac{j}{t}\right) \left(1 - \frac{j}{t}\right)} \cdot \frac{t - 1}{t}$$

has an approximate chi-square distribution with $t - 1$ degrees of freedom and is appropriate for testing H_0 vs. H_A . The test is conditional on the frequency of capture statistics f_1, f_2, \dots, f_t . Burnham also recommended that if f_k is large enough, a statistic appropriate for testing $H_0: p_{ij} = p_i$ versus $H_A: \text{not all } p_{ij} = p_i, \text{ for all } i \text{ such that } y_i = k$, is given by

$$T_{4a} = \frac{\sum_{j=1}^t \left(z_{kj} - \frac{k}{t} f_k\right)^2}{\frac{k}{t} f_k} \cdot \frac{t - 1}{t - k},$$

where z_{kj} = number of animals caught on day j that were captured exactly k times, and

y_i = number of times the i^{th} animal was captured.

Under H_0 , T_{4a} has an approximate chi-square distribution with $t - 1$ degrees of freedom, conditional on the value of f_k . Notice that a test statistic of the form of T_{4a} can be constructed for any $k = 1, 2, \dots, t - 1$ as long as f_k is large enough. We have used the criterion that f_k must be larger than t before the test is performed.

Test 5

An overall goodness of fit test of Model M_b can be constructed by combining the results of 2 independent tests. The first of these tests was introduced by Zippin (1956) for testing $H_0: p_j = p$ versus $H_A: \text{Not all } p_j = p, j = 1, 2, \dots, t$. In the context of Model M_b , p_j represents the probability of first capture on the j^{th} trapping occasion, and thus Zippin's test attempts to determine the constancy of first capture probability over time. The test statistic, which has an approximate chi-square distribution with $t - 2$ degrees of freedom when H_0 is true, is given as

$$T_{5a} = \sum_{j=1}^t \frac{(u_j - \hat{N}_b \hat{p} \hat{q}^{j-1})^2}{\hat{N}_b \hat{p} \hat{q}^{j-1}} + \frac{(\hat{N}_b - M_{t+1} - \hat{N}_b \hat{q}^t)^2}{\hat{N}_b \hat{q}^t},$$

where \hat{N}_b and \hat{p} are the ML estimates of N and p under Model M_b , and $\hat{q} = 1 - \hat{p}$. (Note; we have used the following "pooling" strategy for this chi-square test. If $\hat{N}_b \hat{p} \hat{q}^{t-1} < 2$ and $\hat{N}_b \hat{p} \hat{q}^{s-1} \geq 2$ for $s = 1, 2, \dots, r - 1$, then the "cells" corresponding to $r, r + 1, \dots, t$ are pooled into 1 cell. This will reduce the degrees of freedom associated with T_{5a} to $r - 2$.)

A test (independent of the above procedure) for the constancy of recapture probability over time can be constructed using the so-called variance test for homogeneity of binomial proportions (cf. Snedecor and Cochran 1967: 240). That is, a test statistic appropriate for testing $H_0: c_j = c$ versus $H_A: \text{not all } c_j = c, j = 2, 3, \dots, t$, is given by

$$T_{5b} = \sum_{j=2}^t M_j (\hat{c}_j - \hat{c})^2 / \hat{c}(1 - \hat{c}),$$

where

$$\begin{aligned} \hat{c}_j &= m_j / M_j, \\ \hat{c} &= \text{ML estimate of } c \text{ under } M_b = m / M. \end{aligned}$$

The statistic T_{5a} has an approximate chi-square distribution with $t - 2$ degrees of freedom when H_0 is true.

Because T_{5a} and T_{5b} are independent and have chi-square distributions when Model M_b is true, an overall test statistic for the goodness of fit of Model M_b is given by $T_5 = T_{5a} + T_{5b}$. This test statistic has an approximate chi-square distribution with $2t - 4$ degrees of freedom under Model M_b .

Test 6

Testing the goodness of fit of Model M_t can be interpreted as being equivalent to testing $H_0: p_{ij} = p_j$ versus $H_A: \text{not all } p_{ij} = p_j, i = 1, 2, \dots, M_{t+1}$ and $j = 1, 2, \dots, t$. Thus, the null hypothesis assumes variation in capture probabilities among trapping occasions while asserting that on a given occasion all animals have the same probability of capture. The alternative H_A allows capture probabilities to differ among members of the population on a given occasion, as well as allowing trapping occasions to affect capture probabilities. Notice that H_A does not specify why capture probabilities differ among animals on a given occasion. That is, this difference may be due to behavioral response, individual heterogeneity, or a combination of both. Thus, when H_0 is rejected, it is not in favor of a specific alternative model; rather it is the goodness of fit of M_t that is being rejected. We have chosen to adopt the test procedure proposed by Leslie (1958) for the purpose of testing H_0 versus H_A . The proper test statistic is given by

$$T_6 = \sum_{j=1}^{t-3} \left[\sum_{k=1}^{t-j+1} \frac{f_k^{(j)} (k - 1 - \hat{\mu}_j)^2}{\hat{\mu}_j - \sum_{\ell=j+1}^t R_{j\ell}^2 / u_j^2} \right] I_j,$$

where

- $f_k^{(j)}$ = number of animals captured exactly k times that were first captured on the j^{th} occasion,
- u_j = number of animals first captured on the j^{th} occasion,
- $R_{j\ell}$ = number of animals recaptured on the ℓ^{th} occasion that were first caught on the j^{th} occasion,

$$\hat{\mu}_j = \sum_{k=1}^{t-j+1} f_k^{(j)} (k - 1)/u_j, \text{ and}$$

$$I_j = \begin{cases} 1 & \text{if } u_j \geq 20 \\ 0 & \text{otherwise.} \end{cases}$$

Notice that we have followed Leslie's recommendations that a "cohort" of newly identified animals first captured on the j^{th} occasion not be included in the overall test unless it consists of more than 20 individuals, i.e., u_j must be greater than 20. Furthermore, a new cohort is not included unless it is subject to at least 3 subsequent trapping occasions, i.e., $j = 1, 2, \dots, t - 3$. Under the null hypothesis of "equicatchability," T_6 has an approximate chi-square distribution with

$$\sum_{j=1}^{t-3} (u_j - 1)I_j$$

degrees of freedom.

It should be mentioned that Carothers (1971) proposed an improved version of Leslie's test. However, practical use of the procedure requires some arbitrary trimming of the data and therefore is difficult to use for simulation purposes.

Test 7

Pollock (unpublished dissertation) developed a procedure designed to test H_0 : Model M_n fits the data versus H_A : Model M_{bh} should be used. The test statistic depends on the 2 vectors of statistics

$$\mathbf{f}^* = \{f_1^{(1)}, f_1^{(2)}, \dots, f_1^{(t)}, \dots, f_{t-1}^{(1)}, f_{t-1}^{(2)}, f_{t-1}^{(t)}\} \text{ where } f_k^{(j)} \text{ is the number of animals captured exactly } k \text{ times that were first captured on the } j^{\text{th}} \text{ occasion,}$$

$$\mathbf{f} = \{f_1, f_2, \dots, f_t\}, \text{ where } f_j \text{ is the number of animals captured exactly } j \text{ times.}$$

This is a chi-square goodness of fit test formed by pooling $t - 1$ independent chi-square tests. The k^{th} of these tests is conditional on the value of f_k and has $t - k$ degrees of freedom. The overall test statistic is given by

$$T_7 = \sum_{k=1}^{t-1} \sum_{j=1}^{t-k+1} \frac{\left\{ f_k^{(j)} - \left[\binom{t-j}{t-k-j+1} / \binom{t}{k} \right] f_k \right\}^2}{\left[\binom{t-j}{t-k-j+1} / \binom{t}{k} \right] f_k}$$

Under H_0 , T_7 has an approximate chi-square distribution with

$$\sum_{k=1}^{t-1} (t - k) = t(t - 1)/2$$

degrees of freedom. (Note: for each of the $t - 1$ distributions we have used the same pooling strategy described for the test involving T_{5a} . In this case, the quantity checked for sufficiently large expectation is

$$\left[\binom{t-j}{t-k-j+1} / \binom{t}{k} \right] f_k.$$

Therefore, it may be that in a given case degrees of freedom will not be $t(t-1)/2$, but rather this quantity less the appropriate number of degrees of freedom lost by pooling.)

A Test for Closure

An approach to a closure test suggested by Burnham and Overton (pers. comm.) can be conceptualized by first considering the null hypothesis H_0 : $p_{ij} = p_i$, $j = 1, 2, \dots, t$ for only those animals captured 2 or more times. Essentially, H_0 merely asserts that individual capture probabilities are invariant over time. Now consider, however, an alternative hypothesis H_A that states that for at least some i , i.e., for some animals that were captured at least twice, $p_{i1} = p_{i2} = \dots = p_{ir} = 0$ and/or $p_{is} = \dots = p_{it} = 0$, where $1 \leq r < s \leq t$. H_A states that some members of the population were not present in the population for some initial or terminal part of the study, or both. If that alternative is true, one might expect that the time between first and last capture for animals is, on the average, shorter than one could expect under H_0 . That conjecture is the rationale for the test procedure given below. Note that the test is designed to detect birth-death or immigration-emigration phenomena, or both, that occur only during the initial and latter stages of the study, not phenomena occurring toward the middle of the study. For example, the test would not be appropriate when some animals are present during the initial stages of the study, leave the study area for a time, and then return prior to the termination of the study.

Given that the i^{th} animal was captured exactly y_i times, and that $y_i \geq 2$, let $Q_i = W_i - V_i$, where W_i is the occasion of last capture and V_i is the occasion of first capture. Thus, Q_i is merely the time between first and last capture for the i^{th} animal. Conditional upon the value of y_i , the expectation and variance of Q_i are:

$$\begin{aligned} E(Q_i | y_i = k) &= (k-1)(t+1)/(k+1), \\ \text{Var}(Q_i | y_i = k) &= 2(t-k)(k-1)(t+1)/(k+2)(k+1)^2. \end{aligned}$$

Under H_0 , the statistic

$$\hat{E}(Q | k) = \frac{1}{f_k} \sum_{i=1}^{M_{t+1}} Q_i I_i$$

has the conditional expectation given above

where $I_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ animal was captured exactly } k \text{ times} \\ 0 & \text{otherwise.} \end{cases}$

If f_k is large, the test statistic,

$$C_k = \frac{\hat{E}(Q | k) - (k-1)(t+1)/(k+1)}{\left[\frac{2(t-k)(k-1)(t+1)}{(k+2)(k+1)^2 f_k} \right]^{1/2}}, \quad k = 2, \dots, t-1,$$

can be assumed to be approximately distributed as a standard normal. (We have required that $f_k \geq 10$.) An overall test statistic appropriate for testing H_0 versus H_A is given as

$$C = \frac{\sum_{k=2}^{t-1} [\hat{E}(Q|k) - (k-1)(t+1)/(k+1)]}{\left[\sum_{k=2}^{t-1} \frac{2(t-k)(k-1)(t+1)}{(k+2)(k+1)^2 f_k} \right]^{1/2}}$$

The test statistic C is also approximately distributed as a standard normal if H_0 is true and the number of animals captured at least twice is large (greater than 10). Tests of H_0 versus H_A calculated by using either C_k or C should be one-sided tests since the alternative specifies that the $\hat{E}(Q|k)$ should be smaller than $E(Q_i|y_i = k)$. Thus, closure is rejected only if the test statistic is small. Finally, we emphasize that the test involving C_k is conditional on the value of f_k , and the test using C is conditional on the values of f_2, f_3, \dots, f_{t-1} .

APPENDIX L

Density Estimation Based on Subgrids

The density estimation procedure using nested subgrids is based on Dice's (1938) boundary strip idea. The fundamental aspect of this approach was proposed by MacLulich (1951). Let the trapping grid have area A_g and a perimeter length of P . Then for any convex grid (that includes all rectangular grids) the effective trapping area $A(W)$ is

$$A(W) = A_g + PW/c + \pi W^2/c,$$

where c is a conversion factor to express PW and W^2 in the units of A . From that equation, we derive the expected population size at risk of capture as

$$E(N) = DA(W),$$

where D is the true density of animals. Dividing through by the known area A_g we derive

$$\frac{E(N)}{A_g} = D[1 + aW + bW^2],$$

where

$$a = \frac{P}{A_g(c)}, \quad b = \frac{\pi}{A_g(c)}.$$

Note that the unknown parameters are D and W and that $E(N)$ is estimable from the trapping study.

Given at least 2 grids of different sizes, we can estimate the parameters D and W . Assume there are k different grids (these may be subgrids of 1 overall study). The relevant equations can be written as

$$\frac{N_i}{A_g} = D[1 + a_i W + b_i W^2] + \epsilon_i, \quad i = 1, \dots, k.$$

By assumption, $E(\epsilon) = \mathbf{0}$. Let the variance covariance matrix of ϵ be Σ . We are assuming the same density (D) and strip width (W) apply to all grids. This seems especially reasonable when the grids are nested (see section on DENSITY ESTIMATION).

Given estimates \hat{N}_i , the above equations put estimation of D and W in the framework of generalized nonlinear regression. All we need to carry out the estimation is a knowledge of the variance-covariance matrix $\hat{\Sigma}$. Letting $Y_i = \hat{N}_i/A_g$, we have

$$\text{Var}(Y_i) = \text{Var}(\hat{N}_i)/A_g^2,$$

which is estimable. Also we know that

$$\text{Cov}(Y_i, Y_j) = r_{ij} \text{SE}(\hat{Y}_i) \text{SE}(\hat{Y}_j),$$

where r_{ij} is the correlation of \hat{N}_i and \hat{N}_j . This correlation is not known. Burnham and Cushwa (pers. comm.) suggested the following way of approximating r_{ij} .

The correlation of \hat{N}_i and \hat{N}_j will depend in large part on the overlap of the 2 populations of size N_i and N_j that in turn depends on the overlap of the areas $A_i(W)$ and $A_j(W)$. Thus let

$$r_{ij} = \text{corr}(\hat{N}_i, \hat{N}_j) = \frac{A_i(W) \cap A_j(W)}{A_i(W) \cup A_j(W)},$$

that is, r_{ij} is the area of the intersection of $A_i(W)$ and $A_j(W)$ divided by the area of their union.

Using this formula, we can arrive at an estimator of $\hat{\Sigma}$ that seems reasonable; hence, from the live trapping data we can compute $\hat{Y}_1, \dots, \hat{Y}_k$ and $\hat{\Sigma}$.

We are now in a position to obtain weighted nonlinear least squares estimators of D and W as \hat{D} , \hat{W} satisfying

$$\begin{aligned} \min_{D, W} (\hat{Y} - \mathbf{f})' \hat{\Sigma}^{-1} (\hat{Y} - \mathbf{f}), \quad \text{where} \\ f_i = D[1 + a_i W + b_i W^2], \quad i = 1, \dots, k. \end{aligned}$$

An approximate variance-covariance matrix for the estimators \hat{D} and \hat{W} is given by the 2×2 matrix $V' \hat{\Sigma}^{-1} V$ where V is the Jacobian matrix

$$V = \left[\frac{\partial \mathbf{f}}{\partial D}, \frac{\partial \mathbf{f}}{\partial W} \right].$$

The elements of the columns of the matrix V are, respectively,

$$\begin{aligned} \frac{\partial f_i}{\partial D} &= 1 + a_i W + b_i W^2, \\ \frac{\partial f_i}{\partial W} &= D[a_i + b_i 2W], \quad i = 1, \dots, t. \end{aligned}$$

Because $\hat{\Sigma}$ depends upon W (but not D), an iterative procedure is needed wherein an initial value of \hat{W}_0 is chosen, $\hat{\Sigma}$ is computed based on it, and the new \hat{D}_1 , \hat{W}_1 obtained. Iteration is continued until stable estimates of D and W are obtained.

APPENDIX M

General Simulation Methods

The various estimators and tests described in this monograph were simulated to study their small sample properties and operating characteristics. The asymptotic properties of the tests and estimators are known in most cases, but not the finite sample properties. Monte Carlo simulations provide a method by which our estimation procedures may be evaluated under the exact model from which they were derived. Because we would not expect real data to fit any model exactly, by simulating data, we can study the properties of the procedures without interference due to the data not fitting the model.

Pseudo-random uniform (0,1) variables were generated using the Burroughs 6700 FORTRAN intrinsic function RANDOM. The mixed congruential method is used by this function (Anonymous 1971).

Beta variables for the simulation of heterogeneous probabilities of first capture were generated using subroutine GGBTA from the IMSL (1976) package, with the required uniform variables furnished by RANDOM. That routine uses a rejection method (Fishman 1973).

In the most general simulations, the trapping process is simulated for each animal in the population for each trapping occasion. The probability of capture for a particular animal on a particular day is compared against a uniform (0,1) variable. If the value of the random variable is less than the probability of capture, then the animal is assumed captured, and that element of the \mathbf{X} matrix is set to unity. Otherwise, the value in the \mathbf{X} matrix is set to zero, indicating no capture. When the \mathbf{X} matrix is completely filled, the necessary MSS are computed and the tests and estimation completed.

APPENDIX N

Simulation Results

Simulation Results Regarding Estimation Procedures

In developing this material, we used 6 different models to generate simulated data: all capture-recapture models except M_{tb} and M_{tth} . For each of the data sets, various estimation procedures considered here were applied to gain insight into the operating characteristics of such procedures. The results of this simulation study are presented here in tabular form.

For each model there are 2 corresponding tables. Obviously, for any given model one may choose any number of sets of parameter values needed to completely specify the model. Thus, the first table for each model lists all the different sets of parameter values (each of which is called a Trial) used in the simulation study. For example, Trial 1 in Table N.1.a indicates that there were 400 animals in the population and that every animal had a 0.30 probability of capture on each trapping occasion. That information is all that is required to specify an example of Model M_0 .

The second table for each model presents the simulation results for each estimation procedure used on data generated from the Trials of that model. Column headings in the tables are:

Estimator—identifies the estimation procedure used; e.g., “ \hat{N}_b ” indicates that the estimation procedure described in Appendix C for Model M_b was applied to the data.

R—the number of replications (data sets).

t—the number of trapping occasions.

Ave(\hat{N})—the average value of \hat{N} taken over all replications, i.e.,

$$\text{Ave}(\hat{N}) = \frac{1}{R} \sum_{k=1}^R \hat{N}_k.$$

$\sigma(\hat{N})$ —the “sample” standard error of \hat{N} taken over all replications, i.e.,

$$\sigma(\hat{N}) = \sqrt{\sum_{k=1}^R (\hat{N}_k - \text{Ave}(\hat{N}))^2 / (R - 1)}.$$

Ave $\sqrt{\hat{\text{Var}}(\hat{N})}$ —the average value of $\sqrt{\hat{\text{Var}}(\hat{N})}$ taken over all replications, i.e.,

$$\text{Ave}\sqrt{\hat{\text{Var}}(\hat{N})} = \frac{1}{R} \sum_{k=1}^R \sqrt{\hat{\text{Var}}(\hat{N}_k)}.$$

Coverage—the proportion of replications in which the constructed 95% confidence interval contained the true population size N.

Trial—identifies which population (set of parameter values) generated the data.

Two points concerning $\sigma(\hat{N})$ and Ave $\sqrt{\hat{\text{Var}}(\hat{N})}$ should be mentioned. First the value of $\sigma(\hat{N})$ estimates the true standard error of the estimator \hat{N} in any given Trial and Model so that a measure of the relative bias of the estimator $\sqrt{\hat{\text{Var}}(\hat{N})}$ is given by $|\text{Ave}\sqrt{\hat{\text{Var}}(\hat{N})} - \sigma(\hat{N})|/\sigma(\hat{N})$. Second, the average confidence interval width in a given Trial and Model is given by $2 \cdot (1.96) \text{Ave}\sqrt{\hat{\text{Var}}(\hat{N})}$.

As an example, consider the first line of Table N.1.b. We see that 200 data sets, each consisting of 5 trapping occasions were generated according to Model M_0 , Trial 1 and that the estimation procedure associated with Model M_0 was applied to each set. For those 200 data sets, \hat{N}_0 was essentially unbiased ($\text{Ave}(\hat{N}_0) - N = -0.7$), as was the estimator of the standard error of \hat{N}_0 ($\text{Ave}\sqrt{\hat{\text{Var}}(\hat{N}_0)} - \sigma(\hat{N}_0) = 0.05$). Moreover, 95 percent of the constructed confidence intervals covered the true value of N. This achieved confidence coefficient of 0.95 corresponds exactly with the stated level of the interval, namely 0.95. Such excellent performance of the confidence interval procedure is due in large part to the fact that \hat{N}_0 and $\sqrt{\hat{\text{Var}}(\hat{N}_0)}$ appear to be essentially unbiased. Finally, note that the average width of the 200 constructed confidence intervals is given by $2(1.96)(11.96) = 46.88$.

Simulation Results Regarding the Size and Power of Testing Procedures

Because the testing of assumptions plays a vital role in the analysis of capture-recapture data, it is important to have some insight into the operating characteristics of such tests. Simulation studies of the tests T_1 through T_7 defined in Appendix K were carried out on data simulated from all 8 models to provide such insight, and the results are presented in Tables N.7 through N.13. In addition, Tables N.8 and N.9 give results of an approximation to the power of tests T_2 and T_3 based on theoretical results in Appendix K.

The column headings of the tables are:

α —indicates the nominal size of the test.

Data Model—indicates the Trial and Model that were used to generate the data.

R—indicates number of replications (data sets).

t—indicates number of trapping occasions.

The reader will note that in some instances the number of replications R appears to be strange. The explanation is that a few of the estimators and tests simulated are subject to “failure,” i.e., a particular data set may have characteristics that do not allow calculation of the desired statistics. In those cases, the data sets are excluded from the reported simulation results.

Finally, in Table N.14 we give a description of the Trials of Models M_{tb} and M_{tbh} . Those Trials were involved in the simulation of some of the tests of model assumptions.

TABLE N.1.a.—DESCRIPTION ON THE TRIALS OF MODEL M_0

N	p	Trial
400	0.30	1
400	0.10	2
400	0.05	3
200	0.25	4
100	0.20	5
100	0.15	6

TABLE N.1.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_0

Estimator	N	Ave(\hat{N})	$\sigma(\hat{N})$	Ave $\sqrt{\widehat{\text{Var}}(\hat{N})}$	Coverage	R	t	Trial
\hat{N}_0	400	399.30	11.91	11.96	0.950	200	5	1
\hat{N}_{00}	400	456.93	219.31	160.25	0.922	500	5	3
\hat{N}_{01}	400	405.97	52.84	55.54	0.930	200	5	2
\hat{N}_{02}	100	101.70	19.43	—	—	200	5	6
\hat{N}_{03}	400	397.34	46.56	—	—	200	5	2
\hat{N}_{04}	400	453.82	203.75	—	—	372	5	3
\hat{N}_{05}	400	405.73	52.54	—	—	200	5	2
\hat{N}_{06}	200	199.60	6.58	—	—	100	7	4
\hat{N}_{07}	100	113.46	17.25	13.74	0.830	200	5	6
\hat{N}_{10}	400	386.89	27.65	27.93	0.920	200	5	2
\hat{N}_{11}	400	389.12	29.56	28.02	0.898	400	5	2
\hat{N}_{12}	200	226.38	10.77	10.07	0.260	100	7	4
\hat{N}_{13}	400	319.25	204.51	—	—	372	5	3
\hat{N}_{14}	400	485.90	387.42	—	—	198	5	2
\hat{N}_{15}	200	198.59	11.66	10.47	0.880	100	7	4
\hat{N}_{16}	400	407.71	61.81	57.45	0.940	400	5	2
\hat{N}_{17}	400	439.96	161.42	148.13	0.940	500	5	3
\hat{N}_{18}	200	199.37	6.51	6.85	0.950	100	7	4

TABLE N.2.a.—DESCRIPTION OF THE TRIALS OF MODEL M₁

N	(p ₁ , p ₂ , . . . , p _d)	Trial
400	(0.55, 0.55, 0.50, 0.45, 0.45)	1
400	(0.05, 0.10, 0.15, 0.15, 0.05)	2
400	(0.10, 0.10, 0.10, 0.10, 0.01)	3
400	(0.01, 0.01, 0.02, 0.03, 0.03)	4
400	(0.04, 0.05, 0.03, 0.07, 0.06)	5
400	(0.60, 0.40, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10, 0.10, 0.10)	6
400	(0.50, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10)	7
400	(0.60, 0.40, 0.20, 0.10, 0.10)	8
400	(0.20, 0.40, 0.30, 0.10, 0.20, 0.30, 0.20)	9
200	(0.30, 0.40, 0.10, 0.40, 0.30)	10
800	(0.02, 0.01, 0.03, 0.03, 0.01)	11
100	(0.05, 0.05, 0.10, 0.15, 0.15)	12

TABLE N.2.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M₁

Estimator	N	Ave(\hat{N})	$\sigma(\hat{N})$	Ave $\sqrt{\text{Var}(\hat{N})}$	Coverage	R	t	Trial
\hat{N}_t	400	399.30	202.39	348.63	0.839	155	5	4
\hat{N}_t	400	408.57	73.91	74.19	0.949	198	5	3
\hat{N}_t	100	108.44	37.18	35.77	0.942	989	5	12
\hat{N}_t	800	1,015.62	630.46	748.23	0.902	194	5	11
\hat{N}_t	400	400.33	54.50	56.00	0.955	200	5	2
\hat{N}_t	400	399.28	3.55	3.86	0.960	200	5	1
\hat{N}_t	400	442.66	187.00	151.93	0.930	200	5	5
\hat{N}_t	400	397.14	206.97	341.61	0.880	292	5	4
\hat{N}_t	400	398.59	15.64	16.45	0.940	100	7	7
\hat{N}_t	400	398.81	8.03	7.57	0.950	100	10	6
\hat{N}_t	200	198.70	8.98	7.98	0.880	100	5	10
\hat{N}_t	400	399.41	10.16	9.96	0.930	100	7	9
\hat{N}_t	400	399.45	12.80	11.94	0.940	100	5	8
\hat{N}_t	400	349.17	27.27	26.54	0.515	200	5	3
\hat{N}_t	400	100.83	27.37	12.93	0.000	200	5	4
\hat{N}_t	100	98.70	14.43	13.77	0.924	1,000	5	12
\hat{N}_t	800	219.60	27.94	20.79	0.000	200	5	11
\hat{N}_t	400	244.46	24.04	22.17	0.000	200	5	5
\hat{N}_t	400	98.56	30.27	12.43	0.000	400	5	4
\hat{N}_t	400	506.72	27.76	22.00	0.000	100	5	8
\hat{N}_t	400	516.22	31.59	26.91	0.000	100	7	7
\hat{N}_t	200	456.68	50.28	15.91	0.000	100	10	6
\hat{N}_t	400	235.94	15.78	12.54	0.170	100	5	10
\hat{N}_t	200	459.60	18.32	15.30	0.010	100	7	9
\hat{N}_{bh}^1	400	199.70	21.08	18.08	0.810	100	5	10
\hat{N}_{bh}^1	400	410.30	54.16	70.10	0.860	100	7	9
\hat{N}_{bh}^1	400	387.07	21.14	27.63	0.540	100	10	6
\hat{N}_{bh}^1	400	364.06	55.82	65.24	0.510	100	7	7
\hat{N}_{bh}^1	400	342.56	9.48	5.72	0.040	100	5	7

¹ Note, we use \hat{N}_{bh} and \hat{N}_R as equivalent notation.

TABLE N.3.a.—DESCRIPTION OF THE TRIALS OF MODEL M_b

N	p	c	Trial
400	0.30	0.10	1
400	0.20	0.50	11
400	0.20	0.05	2
400	0.10	0.30	3
400	0.10	0.15	4
200	0.30	0.50	5
200	0.25	0.25	6
200	0.20	0.10	7
100	0.20	0.05	8
100	0.20	0.50	9
100	0.10	0.30	10
100	0.40	0.20	12

TABLE N.3.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_b

Estimator	N	Ave(\hat{N})	$\sigma(\hat{N})$	Ave $\sqrt{\widehat{\text{Var}}(\hat{N})}$	Coverage	R	t	Trial
\bar{N}_b	100	82.00	36.73	64.76	0.708	161	5	10
\bar{N}_b	200	208.09	38.53	40.27	0.930	100	5	7
\bar{N}_b	200	198.36	13.32	13.13	0.900	100	5	5
\bar{N}_b	400	401.23	25.43	19.91	0.920	100	5	1
\bar{N}_b	100	109.83	48.05	—	—	199	5	8
\bar{N}_b	400	400.22	22.38	—	—	200	5	1
\bar{N}_b	400	511.04	408.42	—	—	193	5	3
\bar{N}_b	400	403.56	58.70	—	—	200	5	2
\bar{N}_b	200	197.88	13.32	—	—	200	5	5
\bar{N}_b	400	461.48	269.25	—	—	196	5	4
\bar{N}_b	200	206.99	37.11	—	—	200	5	7
\bar{N}_b	200	198.59	11.66	10.47	0.880	100	7	6
\bar{N}_b	100	82.39	6.75	5.09	0.140	100	5	9
\bar{N}_b	100	176.49	14.40	18.82	0.000	100	5	8
\bar{N}_b	400	347.67	29.55	26.16	0.495	200	5	4
\bar{N}_b	100	71.75	4.80	2.43	0.000	100	5	9
\bar{N}_b	100	299.25	193.37	115.92	0.630	100	5	8
\bar{N}_b	100	51.38	7.06	—	—	161	5	10
\bar{N}_b	100	293.88	109.76	—	—	199	5	8
\bar{N}_b	400	720.39	57.05	—	—	200	5	1
\bar{N}_b	400	212.48	13.56	—	—	193	5	3
\bar{N}_b	400	1,070.94	176.01	—	—	200	5	2
\bar{N}_b	200	175.18	5.80	—	—	200	5	5
\bar{N}_b	400	305.73	33.32	—	—	196	5	4
\bar{N}_b	200	311.67	48.50	—	—	200	5	7

TABLE N.4.a.—DESCRIPTION OF THE TRIALS OF MODEL M_h

N	$p_i, i = 1, 2, \dots, N$	Trial
400	$p_1 = 0.05, i = 1, 200; p_1 = 0.15, i = 201, 300; p_1 = 0.50, i = 301, 400.$	1
400	$p_1 = 0.01, i = 1, 100; p_1 = 0.05, i = 101, 200; p_1 = 0.10, i = 201, 300; p_1 = 0.20, i = 301, 400.$	2
400	$p_1 = 0.10, i = 1, 100; p_1 = 0.20, i = 101, 200; p_1 = 0.25, i = 201, 300; p_1 = 0.30, i = 301, 400.$	3
400	$p_1 = 0.01, i = 1, 50; p_1 = 0.15, i = 51, 200; p_1 = 0.25, i = 201, 300; p_1 = 0.30, i = 301, 400.$	4
400	$p_1 = 0.20, i = 1, 100; p_1 = 0.30, i = 101, 200; p_1 = 0.40, i = 201, 300; p_1 = 0.50, i = 301, 400.$	5
200	$p_1 = 0.05, i = 1, 50; p_1 = 0.15, i = 51, 150; p_1 = 0.25, i = 151, 200.$	6
200	$p_1 = 0.15, i = 1, 50; p_1 = 0.20, i = 51, 100; p_1 = 0.25, i = 101, 150; p_1 = 0.30, i = 151, 200.$	7
100	$p_1 = 0.05, i = 1, 40; p_1 = 0.10, i = 41, 80; p_1 = 0.30, i = 81, 100.$	8
100	$p_1 \sim \beta(3, 22), i = 1, 100.^1$	9
100	$p_1 \sim \beta(1, 22/3), i = 1, 100.^1$	10
100	$p_1 \sim \beta(3/22, 1), i = 1, 100.^1$	11
100	$p_1 = 0.10, i = 1, 40; p_1 = 0.20, i = 41, 80; p_1 = 0.30, i = 81, 100.$	12
400	$p_1 = 0.05, i = 1, 50; p_1 = 0.10, i = 51, 200; p_1 = 0.15, i = 201, 300; p_1 = 0.25, i = 301, 400.$	13
200	$p_1 = 0.05, i = 1, 50; p_1 = 0.10, i = 51, 150; p_1 = 0.25, i = 151, 200.$	14

¹ Indicates that the probabilities p_i were a random sample of size 100 from a beta probability distribution with the indicated parameter values. $E[p] = 0.12$ in those cases.

TABLE N.4.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_h

Estimator	N	Ave(\hat{N})	$\sigma(\hat{N})$	Ave $\sqrt{\widehat{\text{Var}}(\hat{N})}$	Coverage	R	t	Trial
\hat{N}_n	100	35.41	9.36	4.38	0.000	200	5	11
\hat{N}_n	100	79.46	15.15	10.91	0.545	200	5	10
\hat{N}_n	100	84.66	14.32	11.56	0.675	200	5	8
\hat{N}_n	400	331.06	23.99	22.69	0.180	200	5	1
\hat{N}_n	400	298.06	24.52	24.22	0.000	200	5	2
\hat{N}_n	100	100.17	21.15	10.54	0.675	200	10	8
\hat{N}_n	400	460.10	32.31	26.77	0.395	200	5	3
\hat{N}_n	200	226.78	19.82	11.94	0.440	100	7	7
\hat{N}_n	400	443.95	19.71	15.20	0.170	100	5	5
\hat{N}_n	400	461.22	32.64	26.92	0.410	100	5	3
\hat{N}_n	400	417.08	30.91	24.92	0.850	100	5	4
\hat{N}_n	100	107.57	16.80	12.01	0.810	100	5	12
\hat{N}_n	200	207.34	16.14	11.59	0.870	100	10	6
\hat{N}_o	100	26.83	4.79	—	—	200	5	11
\hat{N}_o	100	62.68	12.39	—	—	200	5	10
\hat{N}_o	100	68.90	12.43	—	—	200	5	8
\hat{N}_o	400	227.83	10.97	—	—	200	5	1
\hat{N}_o	400	260.29	30.93	—	—	200	5	2
\hat{N}_o	100	73.72	7.28	—	—	200	10	8
\hat{N}_o	400	365.12	17.15	—	—	200	5	3
\hat{N}_t	200	193.61	9.70	7.69	0.770	100	7	7
\hat{N}_t	400	376.18	10.00	7.67	0.240	100	5	5
\hat{N}_t	400	364.31	16.44	17.01	0.460	100	5	3
\hat{N}_t	400	333.17	17.92	16.19	0.060	100	5	4
\hat{N}_t	100	87.87	12.07	10.19	0.690	100	5	12
\hat{N}_t	200	173.64	8.68	6.83	0.120	100	10	6
\hat{N}_t	400	228.02	10.73	7.40	0.000	100	5	1

TABLE N.5.a.—DESCRIPTION OF THE TRIALS OF MODEL M_{th} . THE PROBABILITY THAT THE i^{th} ANIMAL IS CAUGHT ON THE j^{th} TRAPPING OCCASION IS EQUAL TO $p_i p_j$, $j = 1, \dots, t$ AND $i = 1, \dots, N$

N	$p_i, i = 1, 2, \dots, N$	$p_j, j = 1, 2, \dots, t$	Trial
400	$p_1 = 0.15, i = 1, 100; p_i = 0.33, i = 101, 200; p_i = 0.67, i = 201, 300; p_i = 1.00, i = 301, 400.$	$p_1 = 0.30, p_2 = 0.60, p_3 = 0.10, p_4 = 0.30, p_5 = 0.60.$	1
400	$p_1 = 0.40, i = 1, 100; p_i = 0.60, i = 101, 200; p_i = 0.80, i = 201, 300; p_i = 1.00, i = 301, 400.$	$p_1 = p_2 = 0.25, p_3 = 0.50, p_4 = 0.15, p_5 = 0.25.$	2
400	$p_1 = 0.30, i = 1, 150; p_i = 0.40, i = 151, 250; p_i = 1.00, i = 251, 400.$	$p_1 = 0.15, p_2 = 0.25, p_3 = 0.05, p_4 = 0.10, p_5 = 0.30.$	3
400	$p_1 = 0.03, i = 1, 100; p_i = 0.35, i = 101, 200; p_i = 0.55, i = 201, 300; p_i = 1.00, i = 301, 400.$	$p_1 = p_2 = p_3 = 0.35, p_4 = 0.40, p_5 = 0.60.$	4
200	$p_1 = 0.40, i = 1, 50; p_i = 0.60, i = 51, 100; p_i = 0.80, i = 101, 150; p_i = 1.00, i = 151, 200.$	$p_1 = 0.50, p_2 = p_3 = 0.30, p_4 = 0.60, p_5 = 0.20, p_6 = 0.40, p_7 = 0.30.$	5
200	$p_1 = 0.25, i = 1, 50; p_i = 0.50, i = 51, 150; p_i = 1.00, i = 151, 200.$	$p_1 = 0.15, p_2 = 0.25, p_3 = 0.05, p_4 = 0.10, p_5 = 0.30.$	6
400	$p_1 = 0.45, i = 1, 100; p_i = 0.55, i = 101, 200; p_i = 0.65, i = 201, 300; p_i = 0.75, i = 301, 400.$	$p_1 = 0.65, p_2 = 0.75, p_3 = 0.55, p_4 = 0.35, p_5 = 0.60, p_6 = 0.50, p_7 = 0.80.$	7
100	$p_1 = 0.35, i = 1, 25; p_i = 0.45, i = 26, 50; p_i = 0.55; i = 51, 75; p_i = 0.65, i = 76, 100.$	$p_1 = 0.65, p_2 = 0.75, p_3 = 0.55, p_4 = 0.35, p_5 = 0.60.$	8

TABLE N.5.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES APPLIED TO DATA GENERATED ACCORDING TO MODEL M_{th}

Estimator	N	Ave(\hat{N})	$\sigma(\hat{N})$	Ave $\sqrt{\hat{V}ar(\hat{N})}$	Coverage	R	t	Trial
\hat{N}_t	400	303.60	14.13	11.74	0.000	100	5	1
\hat{N}_t	400	369.90	21.69	19.69	0.600	100	5	2
\hat{N}_t	400	309.04	37.46	35.66	0.340	100	5	3
\hat{N}_t	400	272.77	12.53	10.32	0.000	100	5	4
\hat{N}_t	200	187.93	5.68	5.29	0.360	100	7	5
\hat{N}_t	200	164.17	34.05	29.11	0.590	100	5	6
\hat{N}_h	400	402.88	31.42	23.46	0.850	100	5	1
\hat{N}_h	400	478.73	33.55	29.20	0.220	100	5	2
\hat{N}_h	400	353.90	25.02	26.62	0.590	100	5	3
\hat{N}_h	400	352.78	29.38	20.68	0.430	100	5	4
\hat{N}_h	200	217.72	8.42	9.46	0.560	100	7	5
\hat{N}_h	200	178.21	21.29	18.67	0.760	100	5	6
\hat{N}_{bh}	400	Failed in all 100 replications				100	5	1
\hat{N}_{bh}	400	272.56	11.65	6.27	0.000	100	5	2
\hat{N}_{bh}	400	Failed in all 100 replications				100	5	3
\hat{N}_{bh}	400	489.06	275.72	319.02	0.840	100	5	4
\hat{N}_{bh}	200	179.41	9.27	9.16	0.340	100	7	5
\hat{N}_{bh}	200	235.91	156.90	337.63	0.793	92	5	6

TABLE N.6.a.—DESCRIPTION OF THE TRIALS OF MODEL M_{bh}

N	$p_i, c_i; i = 1, 2, \dots, N$	Trial
400	$p_i = 0.05, i = 1,200; p_i = 0.15, i = 201,300; p_i = 0.50, i = 301,400.$	1 ¹
400	$p_i = 0.01, i = 1,50; p_i = 0.15, i = 51,200; p_i = 0.25, i = 201,300; p_i = 0.30, i = 301,400.$	2 ¹
400	$p_i = 0.10, i = 1,100; p_i = 0.20, i = 101,200; p_i = 0.25, i = 201,300; p_i = 0.30, i = 301,400.$	3 ¹
400	$p_i = 0.20, i = 1,100; p_i = 0.30, i = 101,200; p_i = 0.40, i = 201,300; p_i = 0.50, i = 301,400.$	4 ¹
200	$p_i = 0.05, i = 1,50; p_i = 0.15, i = 51,150; p_i = 0.25, i = 151,200.$	5 ¹
200	$p_i = 0.15, i = 1,50; p_i = 0.20, i = 51,100; p_i = 0.25, i = 101,150; p_i = 0.30, i = 151,200.$	6 ¹
100	$p_i = 0.10, i = 1,40; p_i = 0.20, i = 41,80; p_i = 0.30, i = 81,100.$	7 ¹
400	$p_i = 0.10, i = 1,50; p_i = 0.25, i = 51,200; p_i = 0.35, i = 201,300; p_i = 0.45, i = 301,400; c_i = \max[(p_i - \text{ran}(i)/4), 0].$	8 ²
400	$p_i = 0.05, i = 1,50; p_i = 0.10, i = 51,200; p_i = 0.15, i = 201,300; p_i = 0.25, i = 301,400; c_i = \min[(p_i + \text{ran}(i)/4), 1].$	9 ²
400	$p_i = 0.05, i = 1,50; p_i = 0.10, i = 51,200; p_i = 0.15, i = 201,300; p_i = 0.25, i = 301,400; c_i = \max[(p_i - \text{ran}(i)/4), 0].$	10 ²
100	$p_i \sim \beta(1, 22/3), i = 1, 100^3; c_i = \min[(p_i + \text{ran}(i)/2), 1]$	11 ²
100	$p_i \sim \beta(1, 22/3), i = 1, 100^3; c_i = \max[(p_i - \text{ran}(i)/2), 0]$	12 ²
100	$p_i \sim \beta(1, 22/3), i = 1, 100^3; c_i = \min\{\max[p_i + (\text{ran}(i) - 0.5)/2, 0], 1\}$	13 ²
400	$p_i = 0.15, i = 1,100, p_i = 0.25, i = 101,300; p_i = 0.35, i = 301,400. c_i = 0.5 p_i$	14
100	$p_i = 0.15, i = 1,33; p_i = 0.25, i = 34,66; p_i = 0.35, i = 67,100. c_i = 0.5 p_i$	15

¹ Indicates the trial was used for estimation purposes only. Because the performance of the estimator associated with Model M_{bh} depends only upon N and the probabilities of first capture p_i , recapture probabilities; c_i need not be specified.

² The function $\text{ran}(i)$ produces a random value of a variable distributed uniformly over the interval $[0, 1]$.

³ $p_i \sim \beta(1, 22/3)$ indicates that the random variable p_i has a beta distribution with parameters 1 and 22/3.

TABLE N.6.b.—SIMULATION RESULTS OF ESTIMATION PROCEDURES USED ON DATA GENERATED ACCORDING TO MODEL M_{bh}

Estimator	N	Ave(\hat{N})	$\sigma(\hat{N})$	Ave $\sqrt{\widehat{\text{Var}}(\hat{N})}$	Coverage	R	t	Trial
\hat{N}_{bh}^1	400	246.59	38.04	32.26	0.120	100	5	1
\hat{N}_{bh}^1	400	340.83	57.57	47.90	0.360	100	5	2
\hat{N}_{bh}^1	400	366.43	41.87	35.63	0.600	100	5	3
\hat{N}_{bh}^1	400	383.00	21.30	15.89	0.560	100	5	4
\hat{N}_{bh}^1	200	175.51	16.26	13.52	0.380	100	10	5
\hat{N}_{bh}^1	200	193.72	13.90	13.72	0.780	100	7	6
\hat{N}_{bh}^1	100	94.14	43.26	41.48	0.620	100	5	7

¹ Note, we use \hat{N}_{bh} and \hat{N}_R as equivalent notations.

TABLE N.7.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_1 OF MODEL M_0 VS. MODEL M_b

α			Data model	R	t
0.01	0.05	0.10			
0.0100	0.0300	0.0550	M_0 , Trial 2	200	5
0.0100	0.0150	0.0450	M_0 , Trial 6	200	5
0.9900	0.9900	0.9900	M_b , Trial 1	200	5
0.0900	0.1150	0.1400	M_b , Trial 2	200	5
0.0850	0.1350	0.1900	M_b , Trial 3	200	5
0.1350	0.1650	0.1950	M_b , Trial 8	200	5
0.3950	0.5650	0.6250	M_b , Trial 8	200	10
0.1700	0.1750	0.2500	M_b , Trial 10	200	5
0.7450	0.7750	0.8400	M_b , Trial 11	200	5

TABLE N.8.—SIMULATED AND THEORETICAL RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_2 OF MODEL M_0 VS. MODEL M_b

Method	α			Data model	R	t
	0.01	0.05	0.10			
Simulation	0.0100	0.0550	0.0860	M_0 , Trial 2	198	5
Theoretical	0.0100	0.0500	0.1000	M_0 , Trial 2	198	5
Simulation	0.0280	0.0830	0.1240	M_0 , Trial 3	372	5
Theoretical	0.0100	0.0500	0.1000	M_0 , Trial 3	372	5
Simulation	0.0600	0.1300	0.1400	M_0 , Trial 4	100	7
Theoretical	0.0100	0.0500	0.1000	M_0 , Trial 4	100	7
Simulation	1.0000	1.0000	1.0000	M_b , Trial 1	200	5
Theoretical	0.9990	0.9990	0.9940	M_b , Trial 1	200	5
Simulation	0.9150	0.9700	0.9800	M_b , Trial 2	200	5
Theoretical	0.9480	0.9880	0.9947	M_b , Trial 2	200	5
Simulation	0.8800	0.9640	0.9790	M_b , Trial 3	193	5
Theoretical	0.8480	0.9500	0.9750	M_b , Trial 3	193	5
Simulation	0.0050	0.0660	0.1890	M_b , Trial 4	200	5
Theoretical	0.0510	0.1550	0.2450	M_b , Trial 4	200	5
Simulation	0.9700	0.9900	1.0000	M_b , Trial 5	200	5
Theoretical	0.9460	0.9870	0.9980	M_b , Trial 5	200	5
Simulation	0.2800	0.5100	0.6350	M_b , Trial 7	200	5
Theoretical	0.2610	0.4900	0.6140	M_b , Trial 7	200	5
Simulation	0.4220	0.6280	0.7140	M_b , Trial 8	199	5
Theoretical	0.4570	0.6940	0.7950	M_b , Trial 8	199	5
Simulation	0.0190	0.1990	0.3230	M_b , Trial 10	161	5
Theoretical	0.2200	0.4370	0.5620	M_b , Trial 10	161	5

TABLE N.9.—SIMULATED AND THEORETICAL RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_3 OF MODEL M_0 VS. MODEL M_t

Method	α			Data model	R	t
	0.01	0.05	0.10			
Simulation	0.0150	0.0525	0.1100	M_0 , Trial 2	400	5
Theoretical	0.0100	0.0500	0.1000	M_0 , Trial 2	400	5
Simulation	0.0120	0.0500	0.1060	M_0 , Trial 3	500	5
Theoretical	0.0100	0.0500	0.1000	M_0 , Trial 3	500	5
Simulation	1.0000	1.0000	1.0000	M_t , Trial 2	200	5
Theoretical	0.9998	0.9999	0.9999	M_t , Trial 2	200	5
Simulation	0.3800	0.6550	0.7350	M_t , Trial 5	200	5
Theoretical	0.4269	0.6635	0.7703	M_t , Trial 5	200	5
Simulation	0.6138	0.8028	0.7984	M_t , Trial 12	989	5
Theoretical	0.5925	0.7984	0.8751	M_t , Trial 12	989	5

TABLE N.10.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE GOODNESS OF FIT TEST T_4 OF MODEL M_h

α			Data model	R	t
0.01	0.05	0.10			
0.0200	0.0500	0.0900	M_0 , Trial 2	100	5
0.0000	0.0400	0.0900	M_h , Trial 3	100	5
0.0000	0.0500	0.1000	M_h , Trial 8	100	10
0.0000	0.0800	0.1300	M_h , Trial 10	100	5
1.0000	1.0000	1.0000	M_{bh} , Trial 14	100	7
0.0900	0.2300	0.3700	M_{bh} , Trial 15	100	5
1.0000	1.0000	1.0000	M_{th} , Trial 7	100	7
0.5400	0.7900	0.8800	M_{th} , Trial 8	100	5
1.0000	1.0000	1.0000	M_{tbh} , Trial 1	100	7
0.1600	0.3300	0.4200	M_{tbh} , Trial 2	100	5

TABLE N.11.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE GOODNESS OF FIT TEST T_5 OF MODEL M_b

α			Data model	R	t
0.01	0.05	0.10			
0.0000	0.0200	0.0200	M_b , Trial 11	100	7
0.0000	0.0500	0.1200	M_b , Trial 12	100	5
0.0000	0.0600	0.1000	M_{bh} , Trial 14	100	7
0.0000	0.1000	0.1600	M_{bh} , Trial 15	100	5
1.0000	1.0000	1.0000	M_{tb} , Trial 1	100	7
0.3300	0.5700	0.7100	M_{tb} , Trial 2	100	5
1.0000	1.0000	1.0000	M_{tbb} , Trial 1	100	7
0.0600	0.1800	0.3000	M_{tbb} , Trial 2	100	5

TABLE N.12.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE GOODNESS OF FIT TEST T_6 OF MODEL M_t

α			Data model	R	t
0.01	0.05	0.10			
0.0000	0.0200	0.1100	M_t , Trial 13	100	7
0.0000	0.0300	0.0600	M_t , Trial 14	100	5
0.0000	0.0100	0.0900	M_{tb} , Trial 1	100	7
0.0100	0.0700	0.1200	M_{tb} , Trial 2	100	5
0.0600	0.2800	0.3600	M_{tb} , Trial 7	100	7
0.0300	0.0700	0.2000	M_{tb} , Trial 8	100	5
0.9300	0.9800	0.9900	M_{tbb} , Trial 1	100	7
0.0200	0.0600	0.1000	M_{tbb} , Trial 2	100	5

TABLE N.13.—SIMULATION RESULTS CONCERNING THE SIZE AND POWER OF THE TEST T_7 OF MODEL M_h VS. MODEL M_{bh}

α			Data model	R	t
0.01	0.05	0.10			
0.0100	0.0800	0.1500	M_h , Trial 10	100	5
0.0100	0.0300	0.0800	M_h , Trial 13	100	5
0.0600	0.0700	0.0800	M_h , Trial 14	100	10
0.4000	0.6700	0.7600	M_{bh} , Trial 8	100	5
0.1500	0.4000	0.5100	M_{bh} , Trial 9	100	5
0.0900	0.2300	0.3300	M_{bh} , Trial 10	100	5
0.2800	0.5000	0.6100	M_{bh} , Trial 11	100	5
0.0200	0.1400	0.1800	M_{bh} , Trial 12	100	5
0.0100	0.0500	0.1100	M_{bh} , Trial 13	100	5

TABLE N.14.—DESCRIPTION OF THE TRIALS OF MODELS M_{tb} AND M_{tbb}

M_{tb} ¹					
N	$P_{i,j} = 1, 2, \dots, t$	c	Trial		
400	$p_1 = 0.05, p_2 = 0.35, p_3 = 0.25, p_4 = 0.15,$ $p_5 = 0.20, p_6 = 0.10, p_7 = 0.30.$	2.5	1		
100	$p_1 = 0.35, p_2 = 0.25, p_3 = 0.15, p_4 = 0.20,$ $p_5 = 0.30.$	0.5	2		
M_{tbb} ²					
N	$P_{i,i} = 1, 2, \dots, N$	$P_{i,j} = 1, 2, \dots, t$	c	Trial	
400	$p_i = 0.15, i = 1, 100; p_i = 0.25,$ $i = 101, 200; p_i = 0.35, i = 201, 300;$ $p_i = 0.45, i = 301, 400.$	$p_1 = 0.65, p_2 = 0.75, p_3 = 0.55, p_4 = 0.45,$ $p_5 = 0.60, p_6 = 0.50, p_7 = 0.70.$	2.5	1	
100	$p_i = 0.15, i = 1, 25; p_i = 0.25, i = 26, 50;$ $p_i = 0.35, i = 51, 75; p_i = 0.45,$ $i = 76, 100.$	$p_1 = 0.65, p_2 = 0.75, p_3 = 0.55, p_4 = 0.45,$ $p_5 = 0.60.$	0.75	2	

¹ The probability of any animal being captured on the j^{th} occasion, $j = 1, 2, \dots, t$, is p_j if the animal has not previously been caught and cp_j if the animal is being recaptured.

² The probability of the i^{th} animal being captured on the j^{th} occasion is p_{ij} if the animal has not previously been caught and $p_{ij}c$ if the animal is being recaptured.

APPENDIX O

Interval Estimation

Use of the Central Limit Theorem (cf. Mood et al. 1974:195) in setting so-called “normal theory” confidence intervals for parameters is widespread. The theorem usually is cited as justification for asserting that, for “large samples,” a 95 percent confidence interval for the parameter of interest θ is given by $P\{\hat{\theta} - 1.96\hat{\sigma}(\hat{\theta}) \leq \theta \leq \hat{\theta} + 1.96\hat{\sigma}(\hat{\theta})\} = 0.95$, where $\hat{\sigma}(\hat{\theta})$ represents an estimate of the standard deviation of $\hat{\theta}$ (also see Seber 1973:134 for an example of an indirectly constructed confidence interval based on asymptotic normality). Further impetus is given to this argument when ML estimators are involved by the knowledge that, under certain regularity conditions, those estimators are BAN (Best Asymptotically Normal). Furthermore, large sample (normal theory) confidence intervals based on ML estimators are known to have smaller expected width than intervals constructed using any other estimator (cf. Mood et al. 1974:393).

Unfortunately, it is generally true that assertions concerning the operating characteristics of large sample intervals cannot be made if sample sizes are small or regularity conditions are not met or both. Even more unfortunate is the fact that small or moderate sample sizes are more often the rule than the exception in capture–recapture experiments. Therefore, it was not unexpected that initial simulation of capture–recapture experiments revealed that “normal theory” confidence intervals often exhibit undesirable properties. For example, the lower limit of a given interval for population size N may be less than the number of different animals captured in the experiment. Also, coverage of the confidence interval, i.e., the percentage of simulated intervals that contain the true value N , is often significantly less than the nominal level of 0.95. Because of such problems, 2 alternative interval estimation procedures were investigated in the hope that a more satisfactory procedure could be suggested for practical use.

The first procedure is based upon the supposition that the distribution of the estimator \hat{N}^{-1} is more symmetric (hence closer to normality) than the distribution of \hat{N} (cf. Cormack 1968). Thus, the following procedure was proposed. First, calculate the ML estimates \hat{N} and $\hat{\text{Var}}(\hat{N})$. Construct a 95 percent confidence interval for \hat{N}^{-1} of the form $P\{\hat{N}^{-1} - 1.96\hat{N}^{-2}\sqrt{\hat{\text{Var}}(\hat{N})} \leq N^{-1} \leq \hat{N}^{-1} + 1.96\hat{N}^{-2}\sqrt{\hat{\text{Var}}(\hat{N})}\}$. Finally, invert the interval in the obvious manner to arrive at a confidence interval for N . Simulation results involving Model M_1 showed that, although the distribution of \hat{N}^{-1} was nearly normal, the above procedure did not represent a significant improvement over the usual large sample interval. Coverage of the 2 procedures was roughly the same, but the "reciprocal" procedure had, on the average, greater width than the usual large sample procedure. Moreover, lower confidence limits for \hat{N}^{-1} sometimes were less than zero and thus, upon inversion, the upper limit for \hat{N} was negative.

The second alternative method for interval estimation involved the use of only the likelihood function of the sample, and is based largely upon the likelihood principle (cf. Kendall and Stuart 1973:226). That principle asserts that the likelihood function provides all the information necessary for making statistical inferences concerning the data and has been subjected to serious theoretical questioning (Stein 1962, Birnbaum 1968, Kendall and Stuart 1973). Nevertheless, we felt that it would be beneficial to simulate the operating characteristics of these "likelihood intervals" in capture-recapture experiments to evaluate their practical utility.

Briefly, the mechanics of constructing a likelihood interval are as follows. (The reader is referred to Hudson [1971] for a thorough explanation.) For a given data set \mathbf{X} , form the log-likelihood function $\ln L(\theta|\mathbf{X})$. (Assume θ is a scalar for simplicity of presentation.) Under the assumption that $\hat{\theta}$, the ML estimator of θ , is unique, and that the likelihood function is unimodal, the likelihood interval $I(\theta)$ is defined as $I(\theta) = \{\theta: \ln L(\theta|\mathbf{X}) \geq \ln L(\hat{\theta}|\mathbf{X}) - 2\}$. $I(\theta)$ consists of all those θ for which $\ln L(\theta|\mathbf{X})$ (the log-likelihood function evaluated at θ) is no more than 2 units away from the maximum value of the likelihood function $\ln L(\hat{\theta}|\mathbf{X})$. The assumption is that these values of θ are "plausible" values of θ , in view of the data observed. That is, these values of θ produce values of the likelihood function that are not "far" from its maximum, and thus they cannot be discounted. The use of the value 2 may seem arbitrary, but Hudson (1971) argued that this value leads to asymptotic 95 percent confidence intervals.

Likelihood intervals for population size N were constructed from data simulated according to the 2-sample removal experiment treated by Seber and Whale (1970). Varying the values of N and p (probability of removal) did not appear to significantly affect coverage probability of the intervals, although it is theoretically true that the coverage probability varies at least slightly with the true values of the parameters. Moreover, lower limits of the intervals did not extend below the number of animals seen. Average coverage of the likelihood intervals (95.5%) was approximately the same as that of the "normal theory" confidence intervals (92.2%) constructed from the same data. However, average interval width for the likelihood intervals was consistently greater than that of the normal confidence intervals; on the average they were approximately 10 percent greater. In view of these somewhat mixed results, a second simulation study was performed on data from Model M_1 . That study revealed similar results, in that coverage of the 2

procedures was on the average approximately the same (96.0% for likelihood vs. 93.2% for “normal” intervals), and the average width for the likelihood intervals was always greater than that of the normal confidence intervals. In one case, the average width of the likelihood interval was more than 3 times the average width of the normal confidence interval. Finally, in virtually every simulation involving likelihood intervals, we noted that approximately half of those intervals that did not contain N were too low (i.e., the upper limit of the interval was less than N) and half were too high (i.e., the lower limit of the interval was greater than N). This is in sharp contrast to the results of simulating normal confidence intervals, where we have found that the large majority of intervals that do not contain N are too low.

On the basis of the results described above, we recommend continued use of the usual large sample confidence interval procedure rather than either of the 2 alternative procedures discussed. All 3 procedures possess both attractive and unattractive operating characteristics, and the choice therefore cannot be clear-cut. The decision to continue the use of large sample confidence intervals was made for 2 basic reasons. First, biologists in general tend to be more familiar with the computation and use of that procedure. Second, and more important, is the fact that more is known about the theoretical properties of the large sample procedure than is known about the 2 alternatives, and, therefore, it may be possible to assess theoretically the small sample behavior of the procedure in certain capture–recapture situations. Finally, we wish to reemphasize that large sample “normal” confidence intervals should be used with great caution in many capture–recapture experiments, and that much more theoretical work appears necessary before more adequate procedures are available.