

WILEY



Estimation of Density from Line Transect Sampling of Biological Populations

Author(s): Kenneth P. Burnham, David R. Anderson and Jeffrey L. Laake

Source: *Wildlife Monographs*, No. 72, Estimation of Density from Line Transect Sampling of Biological Populations (Apr., 1980), pp. 3-202

Published by: [Wiley](#) on behalf of the [Wildlife Society](#)

Stable URL: <http://www.jstor.org/stable/3830641>

Accessed: 17-03-2015 20:42 UTC

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.org/stable/3830641?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

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.



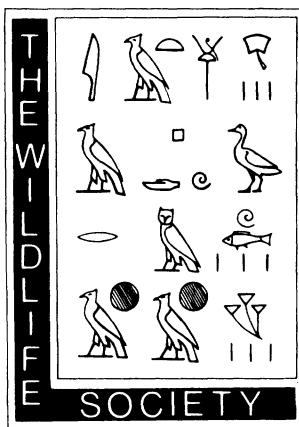
Wiley and Wildlife Society are collaborating with JSTOR to digitize, preserve and extend access to *Wildlife Monographs*.

<http://www.jstor.org>

WILDLIFE MONOGRAPHS

(ISSN: 0084-0173)

A Publication of The Wildlife Society



ESTIMATION OF DENSITY FROM LINE TRANSECT SAMPLING OF BIOLOGICAL POPULATIONS

by

KENNETH P. BURNHAM, DAVID R. ANDERSON,
AND JEFFREY L. LAAKE

APRIL 1980

No. 72



FRONTISPIECE. Field experimentation with methodology for line transect sampling to estimate abundance of the greater prairie chicken *Tympanuchus cupido* in northeastern Colorado. (Photo courtesy of the Colorado Division of Wildlife)

ESTIMATION OF DENSITY FROM LINE TRANSECT SAMPLING OF BIOLOGICAL POPULATIONS

Kenneth P. Burnham,¹ David R. Anderson,² and Jeffrey L. Laake³

CONTENTS

PREFACE -----	8
PART 1	
BASIC CONCEPTS AND BACKGROUND INFORMATION	
INTRODUCTION -----	10
ACKNOWLEDGMENTS -----	11
READER'S GUIDE	11
<i>Biological Technicians Involved Only in Planning and Conducting Surveys</i>	11
<i>Field Biologists without Substantial Quantitative Training</i>	12
<i>Biologists with Substantial Quantitative Training</i>	12
<i>Biometricians and Statisticians</i>	12
CONCEPTUAL BACKGROUND	12
<i>The Abstract Physical Problem</i>	13
<i>Assumptions</i>	13
<i>The Detection Function</i>	14
<i>Estimating the Density of Objects</i>	14
<i>Major Features of Line Transect Sampling</i>	15
Basic Measured Data	15
Data Truncation	15
Grouping Perpendicular Distance Data	15
Replicate Lines	15
Clustered Objects	15
Ancillary Data	15
Variable Sighting Probability	15
Asymmetry of the Detection Function	15
Movement	15
Mode of Detection	15
Studies of Long Duration	15
Estimation of Density versus Numbers	15
Measurement Units	15
<i>History of Concepts</i>	16
<i>Summary</i>	16
FIELD SAMPLING PROCEDURES AND STUDY DESIGN	16
<i>Examples</i>	10
<i>Simulated Data</i>	10
<i>Stake Data</i>	10
<i>Hemingway's Data</i>	10
<i>Summary</i>	10
PART 2	
ROBUST ESTIMATION METHODS	
ANALYSIS AND INFERENCE PROCEDURES FOR PERPENDICULAR DISTANCES, UNGROUPED DATA	56
<i>Fourier Series Estimator</i>	56
<i>Estimation Equations</i>	56
<i>Stopping Rule for the Selection of m</i>	57
ANALYSIS AND INFERENCE PROCEDURES FOR TRANSECT SEGMENTS AS REPLICATES	57

¹ Present address: U.S. Fish and Wildlife Service, Office of Biological Services, Fort Collins, Colorado 80526.

² Present address: Utah Cooperative Wildlife Research Unit, Utah State University, Logan, Utah 84322.

³ Present address: Inter-American Tropical Tuna Commission, LaJolla, California 92037.

PERPENDICULAR DISTANCES, GROUPED DATA	66	Stake Data	84
<i>The Multinomial Distribution</i>	68	Hemingway's Data	86
<i>Estimation with the Fourier Series Model</i>	69	Discussion	88
<i>Selection of m</i>	69	ILLUSTRATIVE EXAMPLES	90
Likelihood Ratio Test	70	<i>Example 1, Ungrouped Perpendicular Distance Data</i>	91
The Chi-square Goodness of Fit Test	73	<i>Example 2, Analysis of Sighting Angles and Distances</i>	94
<i>Examples</i>	73	<i>Example 3, Grouped Perpendicular Distance Data</i>	101
Stake Data	76	<i>Example 4, Transects of Unequal Length</i>	103
Duck Nests	76	<i>Example 5, Measurement Errors</i>	105
Porpoise Study	78	<i>Example 6, Data Truncation</i>	108
<i>Discussion</i>	78	<i>Example 7, Variance Estimates Using Replicate Lines</i>	111
ANALYSIS AND INFERENCE PROCEDURES FOR SIGHTING ANGLES AND SIGHTING DISTANCES	79	<i>Example 8, Noncircular Flushing Envelopes</i>	113
<i>Modeling Sighting Angle and Distance Data</i>	79	<i>Example 9, Other Estimators</i>	117
<i>The Hayne Estimator</i>	80	ANALYSIS AND INFERENCE PROCEDURES FOR MOBILE POPULATIONS	120
Assumptions	81	<i>Estimators Robust to Movement</i>	125
Estimation	81	Monotonically Decreasing Estimators	127
<i>The Generalized Hayne Estimator</i>	82	Criticalness of Movement Assumption	129
Assumptions	82	Recommendations	130
Estimation	83	COMPREHENSIVE COMPUTER PROGRAM	131
<i>The Modified Hayne Estimator</i>	83	DISCUSSION AND SUMMARY	132
Assumptions	83	Discussion of Assumptions	132
Estimation	84	Summary of Important Points	133
<i>Examples</i>	84		

PART 3 STATISTICAL ESTIMATION THEORY FOR ROBUST METHODS

NUMERICAL MAXIMUM LIKELIHOOD ESTIMATION	134	Reversed Logistic Model	145
<i>Ungrouped Perpendicular Distance Data</i>	134	Incomplete Gamma Model	146
<i>Grouped Perpendicular Distance Data</i>	135	A Comment	146
FOURIER SERIES	136	METHODS FOR SIGHTING DISTANCE AND SIGHTING ANGLE DATA	147
<i>Ungrouped Data</i>	136	<i>The Hayne Estimator</i>	148
<i>Grouped Data</i>	139	<i>Modified Hayne Estimator</i>	148
GENERALIZED LINEAR MODELS	139	<i>Generalized Hayne Estimator</i>	149
PARAMETRIC MODELS	141	THE RELATIVE EFFICIENCY OF VARIOUS ESTIMATORS OF $f(0)$	149
<i>Numerical Methods for Parametric Models</i>	141	<i>Introduction</i>	149
Ungrouped Data	141	<i>The Fourier Series</i>	151
Grouped Data	142	<i>Models Investigated</i>	152
<i>Simple Parametric Models</i>	142	Exponential Power Series Model	152
Negative Exponential Model	142	Exponential Quadratic Model	153
Half-Normal Model	143	<i>Results and Discussion</i>	154
<i>Generalized Parametric Models</i>	143	COMPUTER SIMULATION RESULTS	156
Exponential Power Series	144	Simulation Study Design	156
Exponential Quadratic	144	Simulation Results	157
A Mixture Model	145		

PART 4 COMMENTS ON OTHER ESTIMATORS

METHODS BASED ON PERPENDICULAR DISTANCES	164	METHODS BASED ON SIGHTING ANGLES AND DISTANCES	171
<i>Ad Hoc Methods</i>	164	LITERATURE CITED	173
<i>Parametric Methods</i>	165		
<i>Nonparametric Methods</i>	168		

PART 5
APPENDIXES

APPENDIX A	175	Fourier Series Fit to a Linearly Decreasing $f(0)$	191
SYMBOLS, DEFINITIONS AND ABBREVIATIONS	175	Fourier Series Fit to a Strictly Concave True pdf	191
<i>List of Symbols</i>	175	Comment	192
<i>List of Abbreviations</i>	177	APPENDIX D	192
APPENDIX B	177	COMMENTS ON ADDITIONAL LINE TRANSECT TOPICS AND RELATED SUBJECTS	192
SAMPLE HISTOGRAMS OF VARIOUS SIMULATED DATA SETS	177	<i>Clustered Objects</i>	192
APPENDIX C	184	<i>Other Ancillary Data</i>	194
EXAMPLES OF DETECTION FUNCTION SHAPES AND THE FOURIER SERIES MODEL	184	<i>Plotless Sampling</i>	194
<i>Construction of Figures, General Considerations</i>	184	Point Transects	195
<i>The Simulation Study</i>	185	Point to Plant Distance Sampling	196
<i>The Exponential Power Series Model</i>	185	Random Line Lengths and Interobservational Distances	196
<i>The Exponential Quadratic Model</i>	186	Nonuniform Distribution of Objects About the Line	198
<i>The Reversed Logistic Model</i>	188	Modeling Animal Density over the Area Sampled	199
<i>The Incomplete Gamma Model</i>	188	Pooling Robustness and Area Stratification	200
<i>A Comment</i>	189	Finite versus Infinite Population Sampling Theory Approaches	201
<i>Some Fourier Series Models</i>	189		
Fourier Series Fit to the Half-Normal Model	191		

PREFACE

This monograph is intended to provide a rigorous, comprehensive, and practical reference on line transects as they are used in the estimation of density of biological populations. The emphasis is on general inference procedures: point and interval estimates and hypothesis testing. Our work has been focused on inference procedures that are robust to the failure of certain underlying assumptions.

The material presented in this monograph is the result of our interest in the subject over several years and presents the major findings of our research. In particular, we emphasized analytical procedures (in PART 2) that are robust, practical to use, and have good properties. We have also made a concerted effort to present the many complications and facets that arise in line transect sampling and methods for the analysis of data that account for those complications.

Our interest in line transects was stimulated by the paper of Robinette et al. (1974). Their surprising results showed that several seemingly good methods performed poorly while 2 "naive" methods performed quite well. Shortly thereafter, papers on line transects by Sen et al. (1974) and Kovner and Patil (1974) appeared in *Biometrics*. By then, we realized that Seber (1973) had presented several relevant results in his book.

Two of us began to research the subject and our work culminated by presenting a general theoretical framework for line transect estimation (Burnham and Anderson 1976). The framework included parametric and nonparametric estimation for truncated and untruncated data, grouped and ungrouped data, and perpendicular distance data as well as sighting angle and sighting distance data. Our work stopped for 18 months, during which time we both changed positions and locations within the U.S. Fish and Wildlife Service, Burnham going to Colorado and Anderson to Utah.

Research on the subject resumed through a contract with the Western Energy and Land Use Team, Office of Bio-

logical Services. Through Dr. Robert G. Streeter of the Coal Program, we obtained funds to explore density estimation from line transect sampling. An intensive team effort centered at the Utah Cooperative Wildlife Research Unit in Logan began in 1976. Laake joined the team as a graduate student in the Department of Wildlife Science at Utah State University that summer.

Dr. Bradford R. Crain joined us for 10 months as a Visiting Professor in the Unit in the fall of 1976. Crain, a faculty member on leave from the Department of Mathematics at the University of Oklahoma, initiated work on the Fourier series and exponential polynomial estimators and continued their development for ungrouped data. He participated in our effort in innumerable ways and provided a very mathematical outlook for our already diverse group. If we have been successful in attaining some of our objectives, we attribute the success to the team effort.

This monograph combines theory and practice in one package—a philosophy about which we feel strongly. Based on the general theory of line transects, we conduct extensive theoretical and empirical investigations of methods for estimation of density and for dealing with other aspects of data analysis problems that arise in line transect sampling. In some parts of the monograph, we provide many examples and figures to illustrate concepts and calculations; in other parts, we document statistical theory. Our goal in addressing statistical considerations has 2 main components, corresponding to 2 perceived audiences:

- (1) to provide biologists with reliable, practical procedures for the design, conduct, and analysis of line transect studies; and
- (2) to provide statisticians with both the basic general statistical theory for line transects and the specific theory for the new methods recommended here.

These objectives are not mutually exclusive; however, the separation of ap-

plication from theory guided our structuring of this monograph. We included substantial theory partly so that others may use it as a basic reference for developing additional methods for line transect analysis. We encourage biologists to read and understand as much of the methods and theory as possible. We ask the biostatisticians to continue to be concerned with the thorny practical problems in line transect sampling of biological populations. We believe that the involvement of both disciplines is prerequisite to further progress.

We prefer that the reader study the entire monograph. However, recognizing that this will not always be possible, we have organized the material into 5 parts, including the Appendixes. We have provided a detailed READER'S GUIDE to allow readers with various levels of quantitative training to use the monograph. Because of the amount of material covered and the length of the monograph, we encourage everyone to follow the READER'S GUIDE, at least on the initial reading.

All too often, practice and theory are published separately; we feel this approach is counterproductive to good science. For example, PART 3 presents the specific statistical theory underlying the methods presented in PART 2. PART 3 also presents the results of our studies on estimation efficiency and the small sample properties of the estimation procedures. These are important, but difficult, subjects. We have included approximately 18 examples to illustrate the various methods and technical details fully. Several of the examples use the same set of

data to illustrate and compare different methods, approaches, and problems. We have made a concerted effort to present a wide array of examples to aid the reader in understanding the subject. We encourage all readers to make use of those examples.

Our goal of producing a comprehensive monograph on line transects motivates the inclusion of PART 4 and the Appendixes. Those sections are largely oriented to both biologists and statisticians. In particular, anyone wanting to understand past developments in line transect methodology should read PART 4.

All the basic analytical methods we recommend are available in a FORTRAN computer program we call TRANSECT and many of the examples presented here are TRANSECT outputs. That program is easy to use and should be a major help to biologists in analyzing real data and to statistical theorists in further studies of estimation methods. The availability, procurement, and documentation of program TRANSECT are discussed.

This monograph should be most helpful to biologists who have had two or three courses in applied statistics; and we have assumed that the reader is familiar with such concepts as random variables, estimators, sampling variance, confidence intervals, bias, and chi-square test statistics. We make no apologies for those fundamental requirements.

Kenneth P. Burnham,
David R. Anderson, and
Jeffrey L. Laake

December 1979

PART 1

BASIC CONCEPTS AND BACKGROUND INFORMATION

INTRODUCTION

Line transect sampling has been used since at least the early 1930s to obtain estimates of wildlife abundance. Sufficient experience with the method now exists to show that it is practical, efficient, and relatively inexpensive. Despite the potential usefulness of line transect sampling, it is a relatively unexplored methodology (including field methods), especially in comparison to capture-recapture or catch-effort methods. In fact, we know of only about 24 serious papers on line transects and there exists no single comprehensive reference on the subject. This monograph provides what we believe to be a rigorous, comprehensive, and practical reference on line transect sampling as it is used in estimating the abundance of wildlife populations.

The most direct way to estimate the abundance of a biological population is to count all individuals in a known area. In such an approach, the size of the area to be sampled is known and, therefore, an estimate of population density can be obtained simply by dividing the number counted by the size of the area censused. Methods based upon that approach usually are called quadrat, plot, or strip sampling methods. Establishing a plot and then counting all entities of interest within it can be very time consuming. Moreover, such an approach is often impractical, if not impossible; for example, if the target population is mobile or if individuals are difficult to detect or widely scattered (i.e., the density of the population is low). Also, plots may be unsuitable for use in some environments (e.g., surveying marine mammals in the open ocean). For whatever the reasons, plot methods are often unsuitable for use in estimating population abundance, especially for wildlife populations.

Various plotless methods have been devised for estimating animal abun-

dance; they constitute a second major approach to the problem. The use of line transects may be considered a plotless method. For the purpose of this monograph, we distinguish line transects by the following 2 features:

- (1) All objects that are potentially detectable from the transect will not be seen; some objects will be missed, and the farther the object is from the transect center line the greater is the probability that it will be missed.
- (2) Quantitative data are taken, relative to the transect center line, regarding the location of all objects detected. Those data can take several different forms; the 2 basic types are either perpendicular distance from the transect line, or sighting distance and sighting angle.

As with any method for estimation of abundance, line transect sampling is not appropriate for all biological populations. In comparison to many other methods, however, line transect sampling is widely applicable to many species and environments. In fact, as should be evident from the 2 criteria given above, the nature of the biological population being sampled is not the distinguishing feature of line transects. For example, line transects have been applied to the following biological populations: grouse (Rusch and Keith 1971, Myrberget 1976), quail (Overton and Davis 1969), live deer (Kelker 1945, unpublished doctoral dissertation, University of Michigan, Ann Arbor, Michigan), dead deer (Robinette et al. 1954), duck nests (Anderson and Pospahala 1970), African ungulates (Hemingway 1971), nongame birds (Emlen 1971, Jarvinen and Vaisanen 1975), rabbits (Webb 1942, Gross et al. 1974), whales (Doi 1970, 1974, 1975; Schweder 1974, unpublished doctoral dissertation, University of California, Berkeley, California), porpoise (Quinn 1977, unpub-

lished master's thesis, University of Washington, Seattle, Washington; Burdick pers. comm.), seals (Eklund and Atwood 1962), and lizards (Eberhardt 1978a).

Although line transect sampling has a number of advantages as a method for estimating population abundance, careful design of the study is required and the data should receive rigorous statistical analysis. Numerous statistical considerations deserve attention when a line transect survey is planned and when the subsequent data analysis is performed. Only recently, have discussions in any depth of such design considerations begun to appear in the literature. By way of contrast, 30 or more estimators of density for line transect data have been proposed over the past 50 years. Most of those proposed estimators are poor and some are totally unfounded. Furthermore, most reported statistical analyses of line transect data are, in our opinion, inadequate. It is our objective to present a basis for adequate, rigorous statistical analysis of line transect data.

ACKNOWLEDGMENTS

Numerous people have contributed in many ways to this work through discussions, and other communications. There was often disagreement, and such disagreement helped to focus our attention on the critical issues. In particular, we gratefully acknowledge the following people for their stimulation of our thoughts: E. G. Barham, D. L. Burdick, B. R. Crain, L. L. Eberhardt, C. E. Gates, R. S. Holt, K. H. Pollock, J. E. Powers, T. J. Quinn, and F. L. Ramsey.

In developing this monograph and our ideas in general about line transects, the interaction with real data has been extremely beneficial. The following people provided data; we learned much from analyzing them. We have been allowed to use some of these data as the basis for examples presented in this monograph: W. W. Fox and J. E. Powers (1977 porpoise data), P. Hemingway (African ungulates), J. J. Hickey and S. A. Mikol

(breeding bird data), D. A. Kitchen (grouse data), L. W. Robinette (field simulated data), and T. D. Smith (1974 porpoise data). Dr. P. J. Zwank aided in the conduct of the table top surveys used under ILLUSTRATIVE EXAMPLES.

Drs. R. S. Holt, K. H. Pollock, and G. C. White provided extensive technical reviews of this monograph at the final stage of our writing; K. Redman provided a thorough editorial review. Those reviews improved this monograph materially; we thank those people for their help.

We thank G. C. White for producing the final computer runs of our examples based on program TRANSECT. Furthermore, he aided us substantially in other ways, including computer simulations to investigate confidence interval coverage, and preparation of figures involving output from program TRANSECT. We also express our appreciation to R. G. Streeter, U.S. Fish and Wildlife Service, who provided encouragement during this investigation, and to B. Johnson, D. Radtke, C. Snelling, and S. Talbot for their typing of the many drafts of the manuscript.

This work was performed under Contract 14-16-0008-1223 of the Coal Program, Office of Biological Services, U.S. Fish and Wildlife Service, to the Utah Cooperative Wildlife Research Unit. Funds for the 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 Office of Biological Services, U.S. Fish and Wildlife Service.

READER'S GUIDE

This section is intended to provide information on how to study the material presented here based on a person's interest and background. As might be expected in a monographic treatment, the degree of technical difficulty is highly

variable. A field biologist should not be deterred by the mathematics in certain sections of this monograph. The material below is offered as a reader's guide for various user groups. Within each reading list, those sections preceded by an asterisk (*) are suggested as being the more important, while other sections are somewhat optional.

*Biological Technicians Involved
Only in Planning and Conducting
Surveys*

PART 1—Basic Concepts and Background Information

Introduction

Conceptual Background

- *The Abstract Physical Problem
- *Assumptions
- The Detection Function
- Estimating the Density of Objects
- *Major Features of Line Transect Sampling
- History of Concepts
- *Summary

Field Sampling Procedures and Study Design

- *Measurements and Methods of Searching
- *Applications
- *Survey Design
- Sample Size Guidelines
- *Summary

PART 2—Robust Estimation Methods

Illustrative Examples (Nos. 1, 4–6)

Field Biologists without Substantial Quantitative Training

Same reading list as above with the addition of the following:

PART 1—Basic Concepts and Background Information

Conceptual Background

- *The Detection Function

PART 2—Robust Estimation Methods

Analysis and Inference Procedures for Perpendicular Distance—Ungrouped Data

- *Fourier Series Estimator

*Examples

*Summary

Analysis and Inference Procedures for Perpendicular Distances—Grouped Data

Estimation with the Fourier Series Model

*Examples

*Discussion

Analysis and Inference Procedures for Sighting Angles and Sighting Distances

The Hayne Estimator

The Generalized Hayne Estimator

The Modified Hayne Estimator

Examples

Discussion

*Illustrative Examples (Nos. 1, 3–7)

Comprehensive Computer Program

*Discussion and Summary

APPENDIX B. Sample Histograms of Various Simulated Data Sets

APPENDIX C. Examples of Detection Function Shapes and the Fourier Series Model (figures only)

Biologists with Substantial Quantitative Training

Same reading list as above with the addition of the following:

PART 1—Basic Concepts and Background Information

*Estimating the Density of Objects

*History of Concepts

Mathematical Background (all subsections)

*Statistical Background (all subsections)

PART 2—Robust Estimation Methods

The Multinomial Distribution

*Selection of m

Modeling Sighting Angle and Distance Data

*Illustrative Examples (all)

Analysis and Inference Procedures for Mobile Populations

PART 3—Statistical Estimation Theory for Robust Methods

The Relative Efficiency of Various Estimators of $f(0)$

Computer Simulation Results

PART 4—Comments on Other Estimators

(all sections)

APPENDIX A. Symbols, Definitions and Abbreviations*Biometricalians and Statisticians***PART 1—Basic Concepts and Background Information***Introduction***Conceptual Background* (all subsections)**Mathematical Background* (all subsections)**Statistical Background* (all subsections)***PART 2—Robust Estimation Methods** (all sections)***PART 3—Statistical Estimation Theory for Robust Methods** (all sections)**APPENDIX A. Symbols, Definitions and Abbreviations****APPENDIX C. Examples of Detection Function Shapes and the Fourier Series Model****APPENDIX D. Comments on Additional Line Transect Topics and Related Subjects****CONCEPTUAL BACKGROUND***The Abstract Physical Problem*

In this section, the basic concept of line transect sampling is introduced. Numerous considerations and complications can arise in line transect studies and many of the possible ramifications are described. First, the basic idea is presented in the simplest context.

Assume that one has an area of known boundaries and size A and that it is necessary to estimate the abundance of some biological population in the area. The use of line transect sampling requires that at least 1 line of travel be established in the area. That line is traversed (e.g., walked, flown, ridden on horseback) with the objective of detecting a sample of the target objects. Such a situation is diagrammed in Fig. 1, which shows the perpendicular distances from the detected objects to the transect line.

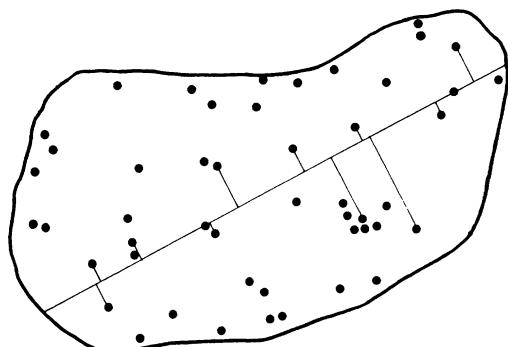


FIG. 1. Diagrammatic representation of a line transect sample. The area, with known boundaries, has points distributed throughout it. Points detected show the perpendicular distance to the line. Points without lines to indicate the perpendicular distances are not detected. Note that points on the line are always detected while those at greater distances have a low probability of being detected. In this example, $n = 13$, including the 2 points detected on the line.

It is not sufficient just to record the number of observed objects; instead, certain quantitative measurements must be taken when an object is detected. A simple geometric diagram of measurements that can be taken is given in Fig. 2. The following notation is used (also see APPENDIX A for a list of the basic notation used):

x = perpendicular distance from the line to a detected object (often an individual animal, nest, burrow, center of a covey, etc.),

r = sighting distance, i.e., distance from the observer to the object at the moment of detection, and

θ = sighting angle, i.e., the angle between the line of travel and the line of sight to the object at the moment of detection.

The line is of predetermined length L . When that line is traversed, a sample of objects will be detected; let n be the sample size. The corresponding sample of potential data is indexed by i (i.e., $x_i, \theta_i, r_i, i = 1, \dots, n$). In practice, all 3 of the possible measurements have rarely been taken; usually, either the perpendicular

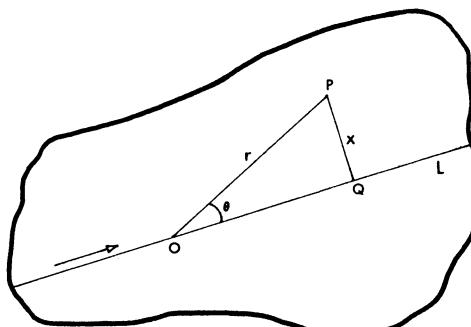


FIG. 2. Diagram of the possible measurements that can be made for a detected object. The observer is at position O when an object is detected at position P, and Q is that point on the line perpendicular to the object. The sighting distance is r , the sighting angle is θ , and the perpendicular distance from the object to the center line is x . Note that the direction of the observer's travel, as shown by the arrow, is from O to Q.

distance (x) or both sighting distance and sighting angle (r and θ , respectively) have been recorded.

The fundamental property and advantage of line transect sampling is that not all individuals will be detected; moreover, objects near the transect center line are more likely to be detected than objects far from the line, as illustrated in Fig. 1.

Assumptions

Because statistical inference will be made on the basis of the data collected, it is essential to formalize a set of basic assumptions that enable one to produce a general model for data analysis. First, the abstract physical situation is formalized:

- (a) Objects (idealized as points) are distributed in the area to be sampled according to a stochastic process (random mechanism) with rate parameter D (density of points per unit area).
- (b) Let a line of known length L be *randomly* located in the area. The line is traversed and a sample of n objects (on either side of the line) is recorded along with either (1) perpendicular distances

or (2) sighting distances and angles for each observed object. In practice, several (replicate) random lines usually will be used.

These are not to be considered as assumptions. That there are objects to be detected in the area of interest is a statement of fact; also, it is not necessary to assume anything about the spatial distribution of the objects. (However, available knowledge, if any, about their distribution should affect the study design with respect to the placement of transect lines.) In particular, it is not necessary to assume that objects are randomly, or even independently, distributed throughout the area. Although it has been stated that randomness is necessary (cf. Gates et al. 1968, Seber 1973), that assumption is unnecessary if the line itself is *randomly located* (Hayes 1977, unpublished master's thesis, University of Edinburgh, Edinburgh, Scotland; Quinn, unpublished thesis; Eberhardt 1978a). It is critical that the line be placed randomly with respect to the distribution of objects; it is that random line placement that justifies extrapolating results to an area of interest larger than just the immediate vicinity of the transect line.

Four assumptions are critical to the achievement of reliable estimates of population abundance from line transect sampling; they are given in order from most to least critical:

- (1) Points directly on the line will never be missed (i.e., they are seen with probability 1).
- (2) Points are fixed at the initial sighting position; they do not move before being detected and none are counted twice.
- (3) Distances and angles are measured exactly; thus, neither measurement errors nor rounding errors occur.
- (4) Sightings are independent events.

Any of these assumptions can be violated in a variety of ways in practical line transect sampling. Which ones might be violated, and why, vary according to the particulars of a given study. The investi-

gator must seriously consider possible violations and design the study and the data analysis to minimize such violations and their effects.

The Detection Function

The data from a line transect survey are a set of distances (and possibly angles) and the resultant sample size itself. To infer animal abundance from those data, one must have a conceptual model that relates the data to the abundance parameter to be estimated. The basic idea underlying such a general model is that the probability of detecting an animal decreases as its distance from the line increases. Mathematically, this idea is represented by a function (or curve) $g(x)$ called the detection function. We were unaware that Schweder (unpublished dissertation) had adopted that name for $g(x)$, when we also chose to use it.

The detection function $g(x)$ is the conditional probability of observing an object given that the object is at perpendicular distance x from the line. It is important to conceptualize the function $g(x)$ clearly. In probability notation,

$g(x) = \Pr\{\text{observing a randomly selected object given } x = \text{the perpendicular distance from the object to the line}\}$,

or, more compactly,

$g(x) = \Pr\{\text{object observed} \mid x\}$.

The function $g(x)$ defines the detection probability of the points as a function of their distance from the line. Assumption (1) in the preceding section specifies that if points are directly on the line (0 distance), they will be seen with probability 1. Mathematically, that is stated as $g(0) = 1$. The idea that the detection probability decreases as perpendicular distance increases means it is assumed that $g(x)$ is a monotonically decreasing function of x (distance from transect line). Although that is not a necessary assumption, it is a very reasonable one and it helps guide subsequent choices of models for $g(x)$. In Fig. 3 and APPENDIX C, several possible

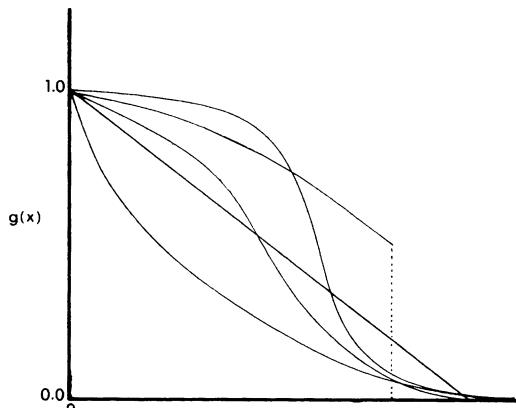


FIG. 3. Diagrammatic representation of several possible shapes of the detection curve $g(x)$. Note that data would be recorded only to distance w in one of the curves; we refer to that as truncation. The bottom curve is convex, the top 3 curves are concave.

shapes of the detection function are diagrammed.

We have introduced $g(x)$ as the probability of sighting an animal given that it is at perpendicular distance x . Note that in so doing it is not assumed that there is just one simple curve that is identical for all objects over the duration of the study; such an assumption would be both extremely restrictive and surely false. Numerous factors affect the probability of detecting objects: for example, weather conditions, the alertness, interests and training of observers, habitat conditions, time of year, time of day, group size (for clustered populations), species, sex, and age, among others. If line transect sampling strictly required the detection probability as a function of x to be some simple function of x for all members of the target population over the entire study course, line transect sampling would be of little value because such a condition is impossible. Fortunately, analytical methods exist that are not affected by variations in detection probability over the population or over the course of the study. Mathematically, the detection function is allowed to be a mixture of many simple functions: $g(x) = g_1(x) + g_2(x) + \dots + g_k(x)$, where each $g_i(x)$, ap-

plies to a particular population segment or set of detection conditions. Thus, it is not necessary, for example, to stratify the data on variables that can affect detection probability. The reason for this is fairly technical; the full explanation appears in *Criteria for Robust Estimation of f(0)*.

Although the detection function is related to, or affected by, many factors, the spatial distribution of objects is one factor to which it is totally unrelated. In particular, whether or not objects have a random spatial distribution has no logical effect on the detection function.

It is noted that the adoption of this conceptual model of a detection function is valid whether or not the observer is recording perpendicular distance. Moreover, it is valid whether detection is due to the observer actively searching for (possibly) nonresponsive objects or to animals flushing as the observer draws near (cf. Eberhardt 1978a). Mode of detection has bearing neither on the validity of the general model nor on the need for the 4 basic assumptions previously stated.

Note also that $g(x)$ is not defined in terms of the right or left side of the line. This point, which is related to whether the detection probability is the same on both sides of the line, is discussed under *Major Features of Line Transect Sampling*.

The sample of perpendicular distances x_1, \dots, x_n is directly related to the detection curve. If one were to take a large sample (which generally implies a long line) and plot a histogram of the distance data and then draw a smooth curve through it, that curve would look much like the true underlying detection curve. In practice, samples usually are too small to give such a smooth histogram. Fig. 4 diagrams an expected histogram and Fig. 5 shows an example of what the actual data might be like for a hypothetical, but reasonably shaped, detection function. Additional examples of histograms of simulated data are given in APPENDIX B; because the true detection functions are known for those simulated data, the

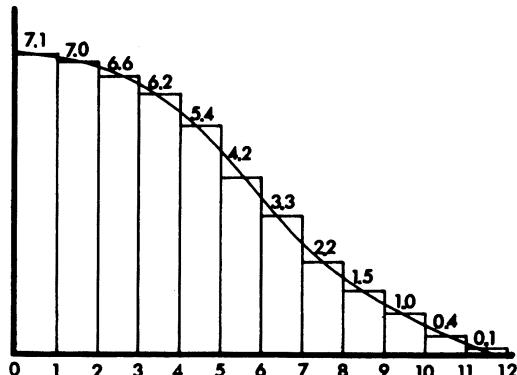


FIG. 4. Expected histogram of perpendicular distance data (arbitrary units) if $g(x)$ has the shape shown for a sample size of $n = 45$, $g(x)$ is not to scale. The vertical axis is the number of individuals.

range of variability to be expected in real data can be inferred.

Estimating the Density of Objects

The detailed mathematical background for estimation of animal density is given under MATHEMATICAL BACKGROUND. The conceptual nature of all line transect estimation is, however, quite simple. Any estimator of the density (\hat{D}) of objects is expressible as

$$\hat{D} = \frac{n}{2La}; \quad (1.1)$$

an equivalent form is

$$\hat{D} = \frac{nb}{2L}, \quad (1.2)$$

with $b = 1/a$ (cf. Seber 1973:29). The heart of the estimation problem is to estimate the 1 unknown parameter, a . For statistical reasons, it is advantageous to estimate $1/a$ (i.e., b) rather than a directly.

The parameter a (or its reciprocal) can be given a variety of interpretations. In particular, a is often referred to as one-half the effective strip width (ESW) of the transect. More important than this unnecessary interpretation of a is the relationship between a and the detection function:

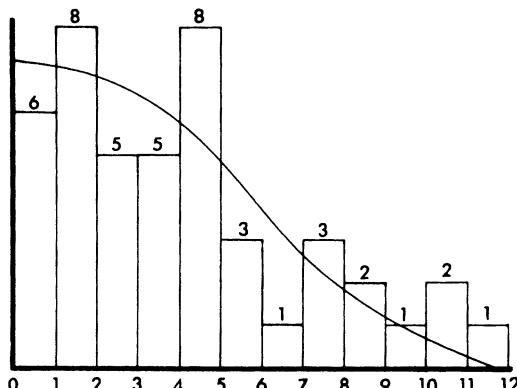


FIG. 5. Example of a histogram of actual perpendicular distance data (arbitrary units) for one sample, $n = 45$, $g(x)$ is not to scale. The vertical axis is the number of individuals.

$$a = \int_0^w g(x) dx.$$

In that equation, w is the maximum perpendicular distance at which observations are made (i.e., the transect width). The investigator may specify a value of w in advance, and then not seek (or record) sightings when their perpendicular distance would exceed w . Such a “trun-

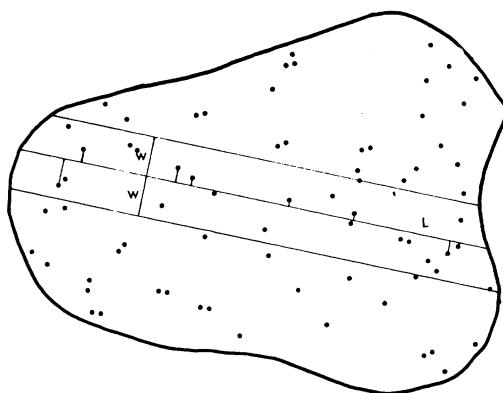


FIG. 6. Diagrammatic representation of line transect line sampling when a finite boundary, w , is placed on the area to be searched. Points without lines perpendicular to the transect represent undetected objects. Only points detected within $2w$ are recorded, $n = 10$. The population is $N = 74$ of which 24 occur within a strip of width $2w$.

cated” sampling scheme amounts to sampling only a strip of width $2w$ centered about the transect line as shown in Fig. 6. It is equally valid to have no designated maximum perpendicular distance, in which case w is taken as infinite ($w = \infty$).

When w is finite, the identifiable strip associated with the line transect has area $2Lw$. The average probability of detecting an object in that strip is given by $P_w = a/w$. Let N_w be the total number of objects in the sampled strip. The expected number of detected objects is $E(n) = N_w P_w$, hence an estimator of N_w is

$$\hat{N}_w = \frac{n}{\hat{P}_w} = \frac{nw}{\hat{a}}.$$

The estimated density of objects from this line transect is just

$$\hat{D} = \frac{\hat{N}_w}{\hat{A}_w} = \frac{\hat{N}_w}{2Lw} = \frac{nw}{2Lw\hat{a}}$$

or

$$\hat{D} = \frac{n}{2L\hat{a}}.$$

Two points are worth making here: (1) the value of w drops out of the formula for \hat{D} , hence it does not matter whether w is finite (fixed) or infinite; and (2) the parameter a is related, but not identical, to the average detection probability over the area searched.

The basic problem in estimating the density of objects is to estimate the parameter a , or, equivalently, $1/a$. To give our approach to this problem, yet another interpretation of $1/a$ must be presented; specifically, there is a relation between the probability density function of the observed perpendicular distance data (x) and the reciprocal of a .

Underlying any continuous random variable, such as detection distance, there is a probability density function. This is the mathematical expression on which probability statements about the random variable are based. Examples that may be familiar to the reader are the

normal and negative exponential probability distributions. Let $f(x)$ denote the probability density function of the perpendicular distance data. That probability function is related to $g(x)$ (Burnham and Anderson 1976) as

$$f(x) = \frac{g(x)}{a},$$

thus $f(x)$ is just $g(x)$ scaled to integrate to 1 (hence be a valid probability density function). Note that $g(x)$ and $f(x)$ have exactly the same shape (see APPENDIX C for graphs of some of those functions).

By virtue of our first assumption (objects on the line are seen with certainty), $g(0) = 1$, hence it follows that $f(0) = (1/a)$. Therefore, equations (1.1) and (1.2) can be written in terms of $f(0)$:

$$\hat{D} = \frac{n\hat{f}(0)}{2L}. \quad (1.3)$$

We have chosen to use Eq. (1.3) throughout this manuscript because it is convenient to formulate the basic problem as one of estimating $f(0)$ as the unknown parameter rather than using $a = 1/f(0)$. It is noted that the units of $f(0)$ are the reciprocal of the units of the perpendicular distance x . The quantity $f(0)$ is statistically well defined and the problem of estimating it has been investigated extensively. The actual line transect sampling process dictates that the key modeling problem is to specify $g(x)$. However, modeling $g(x)$ and $f(x)$ are equivalent problems from a mathematical standpoint, and it is often easier to directly model $f(x)$. Because one then immediately knows the form of $f(0)$, the most useful general equation is (1.3).

The critical nature of the model adopted for the detection function is emphasized. If the model used for $g(x)$ is incorrect, the subsequent estimate of animal density may be biased, possibly severely so. Much of this monograph is concerned with the appropriate modeling of $g(x)$, or equivalently $f(x)$, and the subsequent estimation of $f(0)$ (from a modeling viewpoint, $g(x)$ and $f(x)$ are equivalent; from

a data analysis viewpoint, one is concerned only with estimating $f(x)$.

Some insight into that formulation, including especially Eq. (1.3), may be gained by noting that strip transects are a special case of line transects. In a strip transect, one records any objects within a predetermined, known distance w on both sides of the transect line; and the assumption is made that *no* objects in that strip of area $2Lw$ are missed. That implies a detection function $g(x) = 1$ for $0 \leq x \leq w$, which translates into a probability density function $f(x) = 1/w$, $0 \leq x \leq w$. Thus, $\hat{f}(0) = 1/w$ and applying Eq. (1.3) one has the estimated density from a strip transect as $\hat{D} = n/(2Lw)$, which is the appropriate estimator to use.

Major Features of Line Transect Sampling

The introduction to line transects given above presents only the general features of the subject and the minimal, abstract assumptions and formulations common to all line transect studies. Numerous variants of the basic theory exist, however, and many factors can be combined to specify the particular nature of any given line transect study. The order of presentation of those features does not reflect any priority attached to them.

Basic Measured Data

The basic data that can be taken are perpendicular distance (x), sighting distance (r), and sighting angle (θ). All 3 rarely have been taken independently; typically, investigators have recorded either perpendicular distance, or both the sighting angle and sighting distance, from which the perpendicular distance can be computed. The relation among the 3 measurements is $x = r \cdot \sin(\theta)$. Estimation of density can be based either on perpendicular distance, or on sighting distances and angles. Although the Hayne (1949) estimator can be computed from just sighting distance, it is based on an assumption about sighting angles.

That assumption must be tested and can be done only if sighting angles are also taken. Other methods based on sighting distance also require sighting angles. It is emphasized that no valid estimation is possible based solely on sighting distances, but that valid inferences can be based solely on perpendicular distances.

Data Truncation

Transects are sometimes run with a fixed maximum observation distance (w) perpendicular to the lines (Figs. 3, 6). The alternative case of an unbounded observation "strip" is also used frequently and is mathematically equivalent to w being infinite.

Another form of truncation is often necessary in the actual data analysis. Let w^* be a perpendicular distance such that all recorded perpendicular distances greater than w^* (if any) are not used in the data analysis. It is possible to analyze the data without any such truncation; however, in our experience, it has often been desirable to eliminate 1–3 percent of the most extreme observations ("outliers"; they often represent weird cases). Clearly, if the data are recorded truncated (all observations are $\leq w$), then one must have $w^* \leq w$. Aside from that constraint, w^* can be selected after the data are collected.

Finally, some analytical methods require that a truncation point, w^* , be established (e.g., the Fourier series estimator). However, because w^* can be taken as the largest perpendicular distance observed, no data need be lost.

Grouping Perpendicular Distance Data

It is often difficult to measure distances accurately. Consequently, data for perpendicular distances are sometimes recorded only by distance category. Thus, the exact distance of an animal detected anywhere between, say 0–10 m from the line will not be recorded, but only that it was in that distance category. The resultant data are a set of frequency counts by

specified distance categories rather than a set of exact distances.

In general, let c_i denote a designated perpendicular distance from the line, and assume that for k such distances we have $0 < c_1 < c_2 \dots < c_k \equiv w$. That gives k distance classes, and the basic data are the frequencies of points falling in each distance class (see e.g., Fig. 5). Let n_i = the number of points in distance class i corresponding to the interval $(c_{i-1}, c_i]$; such k intervals can be of unequal length. Note, we let $c_0 \equiv 0$. Also, $w \equiv c_k$ can be finite or infinite. Finally, it is noted that $\sum_{i=1}^k n_i = n = \text{total sample size}$. There are valid ways to analyze grouped perpendicular distance data. No methods have been proposed yet to analyze grouped sighting distance and angle data, although such methods could be developed.

Replicate Lines

It would be rare to have a major line transect study with only 1 line, run only once. In fact, we anticipate that replicate lines will be sampled, over either space or time, or both. That is, an actual field study may consist of numerous lines allocated to the area of interest according to some design. If ℓ_1, \dots, ℓ_R denotes the (possibly unequal) lengths of R replicate lines, then the total length of line is $L = \ell_1 + \dots + \ell_R$. Correspondingly, data will be recognized as deriving from each line segment. True replicate lines must be physically distinct, as opposed to "pseudo" replication in which one line is broken into separate legs and data are recorded separately for each leg. Replication over time may also occur; (i.e., the same line may be sampled more than once, at different times).

It is very desirable to have some form of replicate lines, and to record the data (e.g., sample size and distances) separately for each line. One possible layout of replicate lines is illustrated in Fig. 7; see FIELD SAMPLING PROCEDURES AND STUDY DESIGN for more examples.

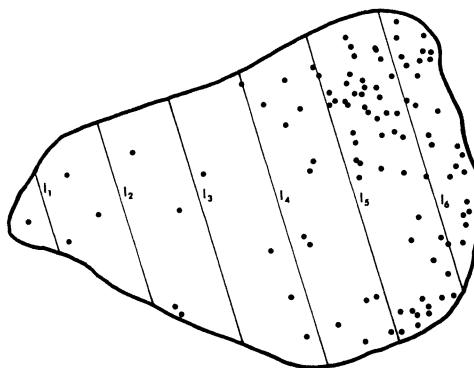


FIG. 7. Example of a study area with 6 replicate lines of unequal length, $\ell_1, \ell_2, \dots, \ell_6$, evenly spaced over the area. Note that the density of objects is not uniform over the entire area, a typical situation.

Clustered Objects

The objects sought often exist not as individuals but as groups or clusters (e.g., herds of antelope, coveys of quail, schools of porpoise). Let group size be denoted by s (s is a random variable, $s = 1, 2, \dots$). If one wants to know just the density of the groups, groups may be treated as the primary objects of interest, in which case measured distances should be to the geometric center of the group. If the density or number of individuals is also desired, then group size must also be recorded accurately. In the subsequent analysis of data, one must first estimate the density of groups by using line transect theory, then estimate average group size. Unless the probability of sighting a group is independent of group size, it is not an easy matter to get a valid estimate of average group size.

Ancillary Data

In addition to distances and angles, much ancillary data (covariates) may be of interest. Group size is one such ancillary variable. Among the many attributes of the objects sought that may be of interest are species, sex, age, size of individuals, and the species composition of groups. Other data that can be taken are

various measurements of factors that can affect the detection probability; they may be features of the objects themselves or of environmental conditions (e.g., reason for the observation, state of the weather, habitat type). See APPENDIX D for more discussion of ancillary data.

Variable Sighting Probability

It is not assumed that the probability of detecting an object given it is at perpendicular distance x is constant either throughout the survey or for all objects. Numerous factors (e.g., weather, time, group size, fatigue of observers) can vary over individual transect lines, during the entire survey, or over the population being sampled. Those variables are likely to influence the detection curve. The analytical methods presented here are robust to those sorts of variations in the detection function. That robustness has significant implications; in particular, it is not necessary to stratify the data on variables that can affect detection probability.

Asymmetry of the Detection Function

It is not necessary for the detection probability always to be the same on both sides of the transect line. It is assumed that data are analyzed without reference to the side of the line on which the object was seen. Thus the data and corresponding detection function are "folded over" (right and left sides pooled).

Movement

Almost all existing methods for the analysis of line transect data assume that the objects sought do not move before being detected. When the target objects are live animals that assumption is unrealistic. However, movement has no detrimental effect on the analyses presented here unless it is in response to the approaching observer and before the animals are detected. The problem most frequently encountered in applying line

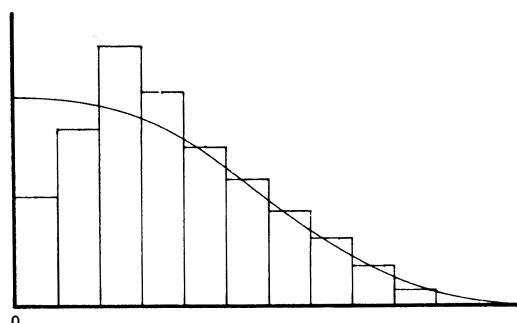


FIG. 8. Diagrammatic representation of the frequency histogram of perpendicular distance data when animals exhibit evasive movement in response to the observer. Units of measurement and sample size are arbitrary. The true detection function (the smooth curve) is shown for comparison.

transects to wildlife populations is that animals will move to avoid the observer. As the observer moves along the line, getting closer to a given animal, that animal may move away, often at right angles to the approaching observer. Such behavior tends either to increase the perpendicular distance of the animal from the line or otherwise cause it to be missed. Then, if the animal is spotted, it will be at a point further from the line than its original position; this violates Assumption (2). The characteristic pattern one can expect to see in the data histogram if evasive movement occurs is shown in Fig. 8. We present results on the analysis of data when such movements have taken place in the section ANALYSIS AND INFERENCE PROCEDURES FOR MOBILE POPULATIONS. With that one exception, all analytical methods in this monograph are for situations with no significant induced movement before detection.

Less commonly, animals may be attracted toward the observer, thus biasing the observed perpendicular distances by rendering them shorter than they would otherwise have been (e.g., some species of porpoise are attracted toward boats). In that case also, the problem again is that the movement is in response to the observer and occurs before the animals are detected. That type of movement (attrac-

tion) cannot be discovered by examination of the histogram of the data.

Movement is not a practical problem if it is both random with respect to the observer and slow relative to the observer's progression down the line (cf. Schweder unpublished dissertation). For example, in an aerial line transect study of porpoise (J. E. Powers pers. comm.) the aircraft moved at more than 140 nautical miles per hour (nmph) and the porpoise moved at less than 10 nmph. Also, the porpoise were not aware of the plane, hence they exhibited no movements in response to the plane (i.e., to the observer). In such a situation, movements are not a problem; for all practical purposes, Assumption (2) is true. By contrast, in a line transect ship survey of porpoise, problems were encountered due to movement because porpoise reacted to the ship, either fleeing from or being attracted to it, and that movement started before the porpoise were detected. Because movement of the ship was barely faster than that of porpoise, porpoise movements cannot be ignored in the data analysis; however, no existing theory enables valid analysis of line transect data contaminated by such movement.

In any application of line transects, the possibility of movement, and its nature, must be considered. If the objects being sampled are immobile (e.g., deer carcasses, plants, duck nests), there is no problem. If the subject of the study is a highly mobile animal (such as a passerine bird), serious problems due to movement can arise, often to the extent of rendering line transect sampling useless for such species.

Mode of Detection

Eberhardt (1978a) has elaborated on 2 approaches to estimation in line transect sampling according to the mode of detection of objects: either the object is unresponsive to the observer, and detection depends upon active (visual) searching for objects, or detection occurs because the object (animal in this case) reacts to the observer and "flushes." A simple way

to distinguish between them is that either "you see it" or "it sees you." Nothing in our assumptions addresses the mode of detection, and that choice was deliberate. The general theory presented is equally applicable to both modes, or to data that represent a mixture of modes; we believe that in real transect studies some observations will be due to passive detection of the animal and others will be the result of an animal flushing in response to the observer.

Unless detection is entirely dependent on a sudden, conspicuous response of the animal, the sighting distance and angle will not be unique. In searching for passive objects that are difficult to detect, the exact moment when one is detected may not be certain. In an aerial survey, for example, 2 independent observers could easily record different sighting distances and angles, but if their data are accurate, the computed perpendicular distances should be the same.

Studies of Long Duration

If a study extends over a long period, further complications in the analysis may result because population density can be expected to change over time. Consequently, some stratification of the data with respect to time becomes necessary. The data set from each such time interval should be analyzed by the methods described here. After numbers or density have been estimated by time interval, one can test for changes over time. It is emphasized that data taken over long periods are not identical in spirit to spatially replicated lines, and may require analytical methods beyond those discussed here.

Estimation of Density versus Numbers

The literature on line transects is ambiguous about whether one is estimating the numbers of objects in a given area or their density. If there is a well-defined reference area of known size A, the choice of estimating numbers N, or den-

sity D, is irrelevant because they are related as $N = DA$. Hence, if either \hat{N} or \hat{D} is computed, the other estimate is readily available.

We have chosen to write in terms of estimating density (primarily for convenience). Because the formula for \hat{D} does not include the known area A, we avoid always showing that symbol. One can always compute \hat{N} , given \hat{D} (and A). All other quantities of interest are related in the same simple way. For example, the standard error of \hat{D} , $\hat{s}_e(\hat{D})$, is related to $\hat{s}_e(\hat{N})$ by $\hat{s}_e(\hat{N}) = A \cdot \hat{s}_e(\hat{D})$.

A minor reason for concentrating on D is that studies may sometimes be conducted in a generally identified area without a specific fixed boundary. In such a case, because there is no well-designated area A to which the density estimate applies, there is no valid concept of numbers N to be estimated and just density is meaningful. We do not recommend that sampling studies be done in that way. However, such line transect sampling is sometimes part of a long-term environmental monitoring program, with the same lines being sampled repeatedly to detect temporal trends in density in a vaguely defined area.

Measurement Units

The units of line length often differ from units of distance. For example, L may be in kilometers or miles and x in meters, feet, or yards. That causes no problem, but must be accounted for in the estimation of animal abundance. Hence, it is worth repeating here that the units of $f(0)$ are the reciprocal of the units of perpendicular distance x. Care must be taken about mixing units of measurement and line length. If they are mixed, \hat{D} must be multiplied by a conversion factor to obtain acceptable units (such as objects per square kilometer, hectare, or acre). Such conversion factors are needed only after the data analysis is done entirely in the original units (computer program TRANSECT automatically handles that conversion).

History of Concepts

Several recent works cover the historical development of line transects: Seber (1973), Hayes (unpublished thesis), Eberhardt (1978a), Gates (1979a). Consequently, only a broad overview of certain conceptual developments in the modeling and analysis of line transect data is presented.

In their initial use, line transects were often thought of as very long, narrow quadrats, i.e., as strip transects (e.g., the pioneering work of Gross in 1906 reported in Forbes and Gross 1921). Line transects that involve the recording of distances and subsequent estimation of animal numbers based on those distances seem to go back only to the 1930s. Gates (1979a) credited R. T. King as being the first to recognize that not all animals were seen in the presumably sampled strip. To devise a way to adjust for animals missed, King apparently thought in terms of trying to estimate the effective half-width of the transect. Thus, the estimated density would be $\hat{D} = n/\hat{A}$, where \hat{A} = the estimated area “effectively” sampled. Because the line length is known, one can, without loss of generality, write the “area” as $\hat{A} = 2L(\hat{ESW})$ where \hat{ESW} is the estimated effective strip half-width (referred to here as just the effective strip width).

The effective strip width can be defined in many ways. Gates (1979a) provided a convenient definition: ESW is that distance from the line such that the number of unseen animals located closer to the line than the ESW equals the number of animals seen at distances greater than the ESW. It is surmised that it was apparent to King that an estimate of ESW would have to be constructed from the observed distances. In any case, King used the average sighting distance, \bar{r} , as the estimate of ESW.

Kelker (unpublished dissertation) introduced another basic approach to estimation of density from line transects. Instead of trying to retain the total sample of n distances and estimate the “area”

sampled, Kelker reversed the process and determined a strip width Δ about the transect center line, within which all animals were seen. Then, one defines n_1 as the number of animals seen in the area $2L\Delta$ about the line and a density estimator (computing equation) is

$$\hat{D} = \frac{n_1}{2L\Delta}.$$

No attempts to create a firm conceptual and mathematical foundation for line transects were published until Hayne's pioneering paper in 1949. All estimators then in use were ad hoc and generally based on either the ESW concept or the related idea of determining a strip width in which no animals were missed. Variants of King's and Kelker's approaches are still being used (and sometimes “re-discovered”) today, even though better methods have existed for some time.

Hayne (1949) provided the first estimator that has a rigorous justification in statistical theory. That is, a set of explicit assumptions (a model) can be made from which the Hayne estimator is justified. Moreover, at least Hayne's key assumption that $\sin(\theta)$ is a uniform random variable on the interval [0,1] can be tested (cf. Eberhardt 1978a). The Hayne estimator, which is based on sighting distances only, is

$$\hat{D} = \frac{n}{2L} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{r_i} \right).$$

In early line transect sampling, it apparently was common to record only sighting distance, although the existence of other measurements was recognized (Webb 1942). In fact, the validity of the Hayne estimator cannot be tested unless sighting angles are recorded.

The Hayne estimator need not be developed from the idea of an effective strip width. It can be based on concepts from probability and statistical sampling, although Hayne's original derivation was heuristic rather than formal. The Hayne estimator is based on the concept that detection is due to the animal “flushing”

in response to the observer. For an explanation of the derivation of the Hayne estimator see *Estimation Based on Sighting Distances and Angles under MATHEMATICAL BACKGROUND* (cf. Overton and Davis 1969).

After Hayne's paper in 1949, almost no significant theoretical advances appeared in the literature until 1968. During that 20-year period, line transect sampling was used frequently, and on a variety of species. The assumptions behind the method were sharpened in the wildlife literature and some evaluations of the method were presented (e.g., Robinette et al. 1956). Yapp (1956) presented a theoretical paper, but his work found no practical application because it made unrealistic assumptions and required information that cannot be gathered by line transect sampling as it is commonly conducted.

In 1968, 2 important papers were published in which some of the basic ideas and conceptual approaches to line transect sampling finally appeared (Eberhardt 1968, Gates et al. 1968). Gates et al. (1968) published the first truly rigorous statistical development of a line transect estimator applicable (only) to *untruncated, ungrouped* perpendicular distance data. They proposed that the detection curve $g(x)$ be taken as a negative exponential form, $g(x) = e^{-ax}$, where a is an unknown parameter to be estimated. Under that model, the probability density function (pdf) of perpendicular distances is the negative exponential, $f(x) = ae^{-ax}$ and $f(0) = a$. Gates et al. (1968) developed the optimal estimator of \bar{N} based on the sample of perpendicular distances, and then constructed the estimator of numbers, \hat{N} . For the first time, rigorous consideration was given to questions such as optimal estimation under the model and estimation of sampling variances, confidence intervals, and testing assumptions. The one weakness was that because the assumed detection model is a very restrictive special case, that may easily fail to be true, the resultant estimate of abundance can be severely biased.

By way of contrast, Eberhardt (1968) conceptualized a fairly general model in which the probabilities of detection decreased with increasing perpendicular distance. Eberhardt reflected on the shape of the detection curve, $g(x)$, and suggested both that there was a lack of information on the appropriate shape and that the shape of $g(x)$ might change from study to study. Consequently, he suggested that the appropriate approach would be to adopt a family of curves to model $g(x)$, a family incorporating a variety of shapes. Eberhardt suggested 2 such families, a power series, and a modified logistic family, both of which are flexible parametric models for the detection function. His exploration of statistical inference based on those models was limited; no estimators based on the logistic model have yet appeared in the literature (cf. Eberhardt 1978a).

As evidenced by those 2 papers, only since 1968 has rigorous attention been given to statistical inference from line transects as they are normally used in ecological studies. From 1968 to 1976, development of rigorous models and estimators was limited mostly to the case of untruncated, ungrouped, perpendicular distance data. Moreover, parametric approaches were predominant, with the notable exception of Anderson and Posapahala (1970) who introduced (inadvertently) some of the basic ideas that underlie a nonparametric approach to the analysis of line transect data.

A general model structure for line transects based on perpendicular distances was presented by Seber (1973:28–30). For an arbitrary detection function $g(x)$, Seber gave the probability distribution of the distances x_1, \dots, x_n and the general form of the estimator of animal density (D). Those equations were left at a conceptual stage and not pursued to the final step of a workable general approach for constructing line transect estimators, and the approach was still based on the ESW concept.

Few developments since Hayne (1949) have dealt with estimation based on sighting distances and angles. Gates

(1969) considered an extension of the negative exponential model to sighting distances and angles. Overton and Davis (1969) presented a conceptual generalization of Hayne's flushing radius model, but did not develop any explicit estimators. In fact, there was, and remains, a tendency to think of approaches based on perpendicular distance as appropriate for inanimate (or nonresponsive) objects, whereas methods for flushing animals were to be based on sighting distances and angles. That artificial distinction tended to prevent the development of a unified theory of line transects. By the mid-1970s, line transects remained a relatively unexplored methodology for estimation of animal abundance.

Burnham and Anderson (1976) pursued the general formulation of line transect sampling and showed a basis for the general construction of line transect estimators. They developed the result $\hat{D} = \hat{n}f(0)/2L$, wherein the parameter $f(0)$ is a statistically well-defined function of the data. The key problem of line transect data analysis was seen to be the modeling of $f(x)$ and the subsequent estimation of $f(0)$. The nature of the data (perpendicular distance vs. sighting distances and angles, truncated vs. untruncated, group vs. ungrouped) is almost irrelevant to the basic problem. Consequently, their formulation is applicable for developing parametric or nonparametric estimators for any type of line transect sampling. Burnham and Anderson's (1976) framework is the basis for the methods presented here.

In the past few years, interest and progress in line transect methodology have expanded considerably. Pollock (1978) has proposed a rigorous, generalized parametric procedure, thus merging in a sense, the research lines of Gates et al. (1968) and Eberhardt (1968). Work on nonparametric procedures has been pursued; for example, Crain et al. (1978) introduced the Fourier series method which forms the backbone of this monograph. The importance of associated problems, such as the treatment of data when animals are grouped, has been recognized. Specific generalizations of

Hayne's estimator have appeared (e.g., Burnham and Anderson 1976, Burnham 1979) for the first time.

Finally, there have been 2 recent developments that we believe should have occurred long ago. Papers are now appearing in the literature that offer guidance on the design and field aspects of line transects (Anderson et al. 1979), and research on the analysis of line transect studies is now being guided by criteria for selecting models and producing estimation procedures (Burnham et al. 1979).

Summary

The preceding material is intended to present a history of concepts and many of the factors that enter into real line transect studies. It is important that the reader be aware of them and of the fact that the various real world aspects of line transects will be treated throughout this monograph.

Any given study is likely to include a variety of those features. For example, a 1977 aerial porpoise study (Holt and Powers In press) dealt with a population that was grouped, schools of porpoise, and the schools were of variable species composition and size. The objective of the study was to estimate numbers by species. Also in that study, only perpendicular distance data were taken and they were taken grouped, but there was no truncation, (i.e., w was infinite). The study extended over 5 months, had numerous replicate lines, and porpoise density varied over the ocean study area. Finally, schools of porpoise were mobile, but movement was not a problem because the airplane was fast relative to the porpoise and not perceived by them.

A quite different study was reported on by Anderson and Pospahala (1970). The target population was duck nests, which are, of course, immobile. The study extended over 2 summers and replicate transects were used. Data were taken truncated, but were reported and analyzed grouped. Despite the differences between those studies, the same underlying theory applies to both and we rec-



FIG. 9. Line transect sampling schemes can be useful in estimating population density of slow moving reptiles like the gila monster. (Photo by Ralph Foote)

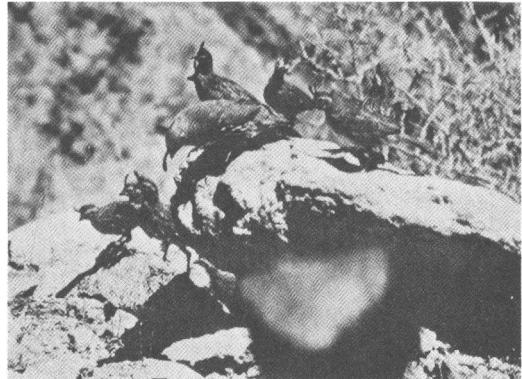


FIG. 10. Pheasants, quails, porpoises, and many plant populations aggregate into coveys, schools, or clumps and are examples of clustered populations. In line transect surveys of such populations, each cluster is treated as a point and the size of each cluster must also be recorded in the field.

ommend use of the same estimator for both (i.e., the Fourier series method, presented in PART 2).

FIELD SAMPLING PROCEDURES AND STUDY DESIGN

That few biologists fully understand the many important, and sometimes subtle, factors involved in planning and conducting a survey using line transect sampling is not surprising, because there is essentially no literature on practical field procedures except for Anderson et al. (1979). Our objective in this section is to provide broad guidelines for the design and conduct of field surveys to estimate the size or density of biological populations by means of line transects. The analysis methods outlined in PARTS 2 and 3 depend on proper planning of the survey and data collection. In general, a competent statistician or a scientist with training in quantitative methods should be involved in planning a survey.

We hope to illustrate the practical advantages of line transect studies in estimating a wide variety of animal populations. Immobile populations (e.g., nests, dead deer, yucca and saguaro plants), "flushing" populations (e.g., pheasants, sage grouse), and slow-moving popula-

tions (e.g., gila monsters, desert tortoise) seem particularly suited to line transect sampling methods (Fig. 9). Bobwhites (flush in coveys) and rabbits may also represent populations whose size can be likewise estimated (Fig. 10). At the other extreme are populations such as whales and moles (probability of being seen on the center line of the transect is less than 1) and chukars (assemble in loose groups and run rather than flush), for which line transects are simply not appropriate, at least without additional information (such as the probability of being on the surface at the time of survey).

Measurements and Methods of Searching

The major advantage of the line transect sampling scheme is the relative ease of its implementation in the field once a proper sample of lines has been chosen. The problems encountered in the field are threefold:

- (1) defining the *straight* line of travel,
- (2) obtaining accurate measurements of distances and angles, and
- (3) ensuring that all objects on, and very near, the transect center line are seen with certainty (i.e., obtaining $g(0) = 1$).



FIG. 11. Walkie-talkies and a transit are used to assure the observer (background in photo) follows a straight line. As objects are found, measurements are radioed back to the base station (foreground).

Although the field procedures described are simple, an adequate understanding of them is essential to obtaining good estimates of density.

The entire procedure of line transect sampling is based on being able to follow a *straight* transect line or series of *straight* line segments. This is almost always compromised to some degree in the field because of physical obstacles and the difficulty of traveling a straight line and searching for objects simultaneously. However, unless the line of travel is defined in some manner, accurate measurements of perpendicular distance and sighting angle cannot be obtained. If the observer tends to walk toward the object when it is sighted, the perpendicular distance and sighting angle will be negatively biased and the density estimate positively biased.

A simple approach to defining the line is to follow a compass line by sighting on landmarks on the line of travel to a point on the horizon. Such an approach does not necessitate a fixed line, therefore it is not recommended because precise line placement is essential. Precision can be increased by marking the line with flagged or painted stakes at convenient intervals. The interval between markers on such a permanent line will depend on

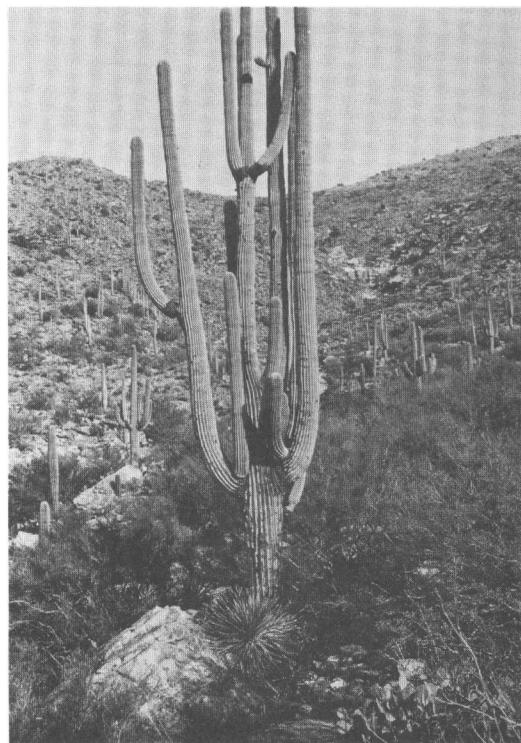


FIG. 12. Many plant populations such as the saguaro can be sampled by line transect. Here, the probability of seeing a plant on the center line is most assuredly 1. Careful measurements are possible with immobile objects.

their visibility, but it can be as much as 0.8 km (0.5 miles) and still provide for accurate measurements. A more elaborate and sophisticated technique involves the use of a transit and walkie-talkies to keep the observer on the line (Fig. 11). That technique allows the observer to spend more time searching for objects than concentrating on following a line. It is an ideal method but is limited to open flat areas and requires 2 people. The use of any one of those methods will be determined by the terrain, the available manpower, and the precision required.

In searching for objects, the observer scans a semicircle of reasonable radius within which most objects will be sighted, while *being especially careful not to miss any objects on the center line*. Once

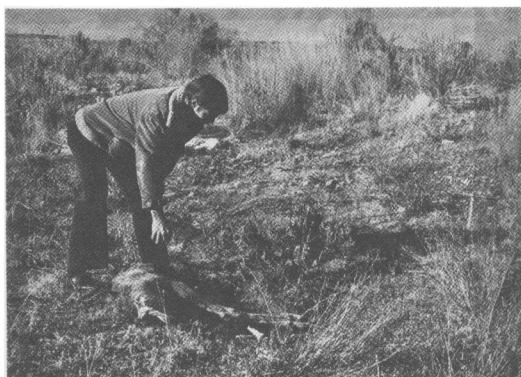


FIG. 13. Accurate measurements are made best with a steel tape. Notice the metal stake used at the center line to facilitate careful measurement.

an object is sighted, the necessary distance and angle data are taken. As a *minimum*, the observer can measure either

- (1) perpendicular distances, x_i , or
- (2) sighting distances r_i and sighting angles θ_i .

All 3 measurements should be taken, when possible, as each object is found. The measurements can then be checked for gross errors because each can be computed from the other 2.

The use of estimators based on perpendicular distance data is recommended. If perpendicular distances are not taken, both sighting distance and sighting angles must be taken. We do not believe that reliable analytical methods can be developed to use information only on sighting distances because very strong (and unreasonable) assumptions would have to be made for such analyses. It is essential that straight transect lines be used and the observer must

- (1) conduct the survey in such a way as to maximize the probability that *all objects on the center line will be found* (Fig. 12), and
- (2) be able to determine the center line of the transect accurately once an object is found.

Distance measurements are preferably made with a steel tape measure (Fig. 13).

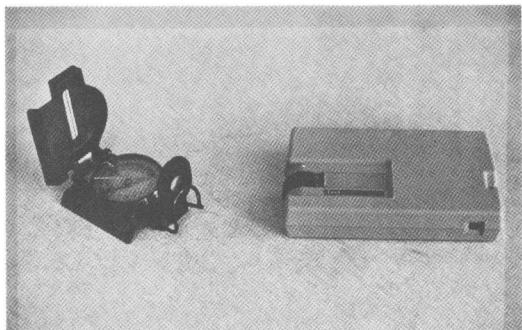


FIG. 14. A liquid-filled lensatic compass (left) is adequate for measuring sighting angles in the field. It is important that those angles be measured accurately. A range finder (right) can be used to estimate distance (y_i and r_i) if the terrain is extremely rough (e.g., surveys of nene geese on lava flows in Hawaii). Where possible, a steel tape is recommended for accurate measurements.

In many instances, long distances and hilly terrain require pacing or using a range finder (Fig. 14). The use of both of those methods is discouraged, and their occasional use can be supported only when it is not practical actually to measure the distances. We found that a liquid filled lensatic compass (Fig. 14) gave quick and accurate readings of θ_i because of its simple sighting mechanism. Many observers use a stick as a "beater" to aid in locating objects, probing into dense cover. If most objects are found fairly close to the center line, a heavy yardstick (or meter stick) may serve the dual purpose of a beater and measuring device.

It may sometimes only be possible to assign objects seen to one of several distance classes (e.g., 0–3 m, 3–6 m, 6–9 m, etc.). Although that practice is discouraged, it is realized it may be necessary under certain field conditions. It is recommended that accurate measurements of the perpendicular distances be recorded. Then, if grouping is to be done before analysis and estimation, one is free to select the number of intervals and their widths. It is important that the various measurements not be rounded (unintentionally or intentionally) to convenient values (e.g., rounded to 5, 10, 15, . . .).

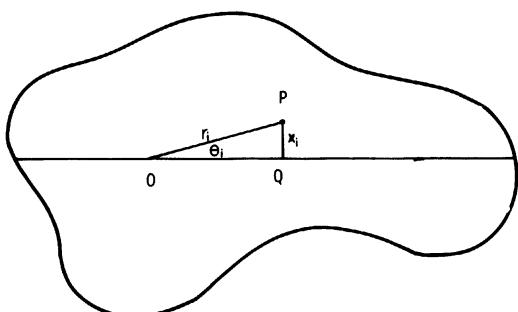


FIG. 15. O is the position of the observer when an object is detected at point P. Q is the point on the line perpendicular to the object. The sighting distance r_i , sighting angle θ_i , and the perpendicular distance x_i are shown also.

Emlen (1971) suggested a combination field technique and analytical method. Our analytical approaches (PARTS 2 and 3) can be used with grouped data of unequal widths. It is suggested that the individual distances be taken; various analyses can then be attempted. It is important to recognize that *grouping the distance data in the field is not recommended*. In any case, the analysis must rest on a sample of objects that is reasonably large ($n > 25$ or 30). The analysis of, say, 7 objects will be essentially meaningless whatever the analytical method selected.

We have developed an approach that enables all 3 measurements to be obtained in the field. For immobile objects (Fig. 15) the procedure is to:

- (1) measure sighting angle and place a marker at point O,
- (2) find point Q with a compass reading in the direction of point P, at a right angle to the line of travel,
- (3) measure perpendicular distance from point P to Q, and
- (4) measure sighting distance from point O to P.

The procedure changes for mobile objects because it is seldom possible to determine the exact position of the animal before it moved. A marker is placed at point O and sighting distance is measured first and a marker is placed at point



FIG. 16. Offroad vehicles can sometimes be used in line transect surveys. Here, an army mule is shown. Visibility, maneuverability, and portability are good.

P. The sighting angle is then measured, and finally, the perpendicular distance is measured (Fig. 15). The transit procedure, described previously, simplifies the matter considerably: the observer measures sighting angle and sighting distance to point P and then, using a compass, walks at a right angle back to the line of travel. The person operating the transit can tell the observer when he has reached the line. All the measurements can be recorded by the person at the transit, so the observer need not carry a clipboard, field forms, or other gear.

The general nature of line transect sampling permits searching by many methods other than on foot. Travel on horseback can be advantageous in upland situations. Use of a horse can improve visibility and reduce observer fatigue but may cause problems because the horse must be secured while the observer takes the measurements when an object is found. Well-trained dogs can be used to flush game birds and to help in detecting other animals; the dogs should work a fairly tight pattern around the line of travel so that all birds near the line are seen. Offroad vehicles (ORVs) can be used effectively under some conditions (Fig. 16). This will often reduce observer fatigue and increase the line length covered per day. However, if the use of an offroad vehicle, aircraft, or other vehicle might tend to lessen the chance that objects on the line would be observed, such

use cannot be recommended. We suspect the line transect survey data reported by Leatherwood et al. (1978) may have been biased because observers failed to search the center line beneath the airplane carefully.

Applications of line transect sampling for whales or porpoise require the use of a boat or airplane. Basically, any convenience method of locomotion that will not violate the 4 fundamental assumptions can be used.

Finally, only competent, interested, and trained personnel should be employed to conduct line transect surveys. This point is crucial and cannot be overemphasized.

Applications

Immobile Populations

Populations whose individuals are fixed and sedentary or move very slowly in relation to the observer are the easiest to survey. The assumptions are most easily met with immobile objects. Examples of such populations include surveys of dead deer, land snails, conspicuous plants, and bird nests. Tortoises probably can be considered as immobile for purposes of line transect surveys.

Our 4 basic assumptions (see *Assumptions*) are extremely important. The degree to which mobile populations can be surveyed depends on the degree to which those assumptions can be closely approximated.

Mobile Populations

Many problems associated with line transect sampling are encountered with mobile populations. The probability of detecting an animal may depend on the response of the animal to the observer. If an animal hides, or runs and then flushes, or is driven in front of the observer, the assumptions of the model will not be met. Also, the measurements will be less accurate because the exact position of the animal before flushing usually cannot be determined.

Consider a study of live deer. Deer in the survey area will not stand and be counted. They may hide motionless in thick cover or perhaps be evasive and leave the area before they can be spotted by the observer. In other words, they may make either a passive or an active effort to avoid detection that causes difficulties in the proper statistical analyses of the data (see ANALYSIS AND INFERENCE PROCEDURES FOR MOBILE POPULATIONS).

To determine the applicability of line transect sampling to a certain species, the behavior and biology of the animal must be examined with respect to the assumptions of the model. Those assumptions are described below in terms of field applications to mobile populations; an example that violates one of the assumptions is given for each assumption.

(1) *Animals on the line are always seen.* This assumption presents problems that deal with burrowing animals such as ground squirrels and cottontails, or aquatic animals such as whales. An animal could be on the line but hidden by the respective medium.

(2) *Animals do not move before being seen or flushed.* The response of an animal may be to move away from the observer's line of travel without being noticed. If it flushes then, the observed perpendicular distance will be exaggerated. Chukars often respond by running before they flush, making a good survey of their population size quite difficult. (Movement is not critical if it is independent of the observer and "slow" relative to the observer's speed along the line.)

(3) *Animals are not counted more than once.* In response to the observer, some animals move ahead of the observer near the line of travel. Unless this can be observed and negated, an animal may later be recounted. Many songbirds behave in this manner.

(4) *The sighting of animals is independent.* The flushing of one animal does not cause another to flush. This assumption will not be applicable to animals that gather in loose flocks, such as pinyon jays. Failure of this assumption does not

affect the point estimate of density (i.e., \hat{D}); only the sampling variance of \hat{D} will be affected.

Clustered Populations

Animals that aggregate in coveys, schools, and other tight groups are referred to as clustered populations (Fig. 10). Clustered populations can be either mobile or immobile. For line transect sampling, each cluster (e.g., covey) is treated as a point, and cluster detections are assumed to be independent events. Distance and angle measurements relate to the geometric center of the cluster as the point of observation, which enables estimation of cluster density. To estimate population density of the individual objects, it is essential to determine the size of each cluster accurately and to obtain the measurements of r_i , x_i , and θ_i .

Hayne (1949) considered estimation of the size of clustered populations. More recently, Pollock (pers. comm.) has attempted to examine the estimation theory of clustered populations. No published field methods for such populations are yet available, however it is believed that the line transect sampling method can be applied only if the geometric center of the cluster can be determined adequately. Because quail and gray partridge often form tight coveys, the method would be applicable. Animals in loose groups, however, cannot be analyzed appropriately as a clustered population because they usually do not flush simultaneously and the geometric center is nearly impossible to establish accurately. Chukars, pinyon jays, and some songbirds fall into that category. Examples of immobile objects that occur in clusters would include plants with poor dispersal mechanisms.

Survey Design

As with any study, it is important to consider carefully the statistical design of line transect studies. By study design, we mean the determination of the desired level of sampling effort and its allocation

over both the study area and the duration of the study.

Factors that affect the level of effort include the number of (replicate) transect lines, the length of each line, the frequency with which each line is surveyed (e.g., only once, or once a week for several months), and the number of observers or teams of observers who conduct the study. In addition, the effort allocated over the area to be studied and over the time (e.g., week, month, year) of interest must be determined. Considerations here include, for example, placement of transect lines in the study area, the number of times each line is to be run, and the schedule of observers for specific times and transects.

One line, randomly placed in the area with 1 observer conducting the survey, is too simplistic a design to be useful; most line transect studies employ replicate lines in the area, extend over some period of time, and involve several observers. If the period is short (a few days, perhaps), the density estimate may effectively be related to one point in time. If the study is to extend over many months or years, however, it is necessary to partition the data by time periods and consider carefully to which period the estimates relate. Questions of immigration, emigration, survival, and recruitment are also important in such longer-term studies.

The placement of the transects within the study area is critical. Transects can be placed systematically, randomly, or in a stratified design. Whatever of the type of placement, the lines should not be too close together nor should they overlap.

Stratifying a study area by some feature such as habitat type is common and may be useful in line transect sampling; however, a stratified design presents a potential pitfall. If each stratum contains a sufficient length of lines to provide a sample large enough to enable analysis of the data by stratum, no problems arise, but if the effort per stratum is insufficient to analyze the distance data separately by stratum, the data will have to be pooled over strata before analysis. Such a pooling will lead to a biased estimator of av-

verage density unless the total line length is allocated to strata in proportion to the area of each stratum. This technical point is explained further in APPENDIX D (*Pooling Robustness and Area Stratification*). Here, it is simply stated that spatially stratified designs should have proportional line length allocation per stratum.

The timing sequence must be considered in conducting a survey, including such major considerations as the time of year for the survey and the assignment of observers to the replicate lines. It also includes lesser considerations such as the starting and stopping time of observation during a single day and even the direction of travel (e.g., going north vs. south) along given lines. The main objective in these considerations is to avoid confounding any changes in density that may occur over time with the sequence in which the lines are run.

The principles of statistical design of studies (or experiments) are well documented and it is beyond our objectives to cover the fundamentals here. The reader is referred to Cochran and Cox (1957), Eberhardt (1978b), Cochran (1963), and Cox (1958). It is recommended that an investigator either be familiar with those principles of sampling and experimental design or consult an appropriately trained person. Some additional sources of specific information on planning line transect studies are Seber (1973), Hayes (unpublished thesis), and Eberhardt (1978a, 1978b).

The design of each individual study requires advance information on (1) the biological aspects of the population to be studied, and (2) a delineation of the area to be surveyed by shape, size, and type of habitat. Such information will enable optimal design of the survey in terms of smaller variances, smaller biases, shorter confidence intervals and greater reliability and overall efficiency.

Biological considerations are, of course, important in line transect sampling work. For example, if a type of plant is being sampled, counts along one transect line



FIG. 17. Permanent transect markers such as that pictured are ideal if the survey is to be replicated; 30×40-cm plywood is used and attached to a 2.5-m pole or pipe.

should not appreciably affect counts along neighboring transect lines, because plants are stationary. On the other hand, if deer are being sampled, it would be important to place transect lines some distance apart, because taking counts along one transect line might influence counts along nearby transects. Statistically speaking, one would like counts along different transect lines to be "independent" random variables.

The placement of transect lines may be either temporary or permanent. When a survey is a one-time occurrence, temporary lines are appropriate. Permanent transect lines, delineated by poles or other markers, should be considered if the transects are to be surveyed periodically,

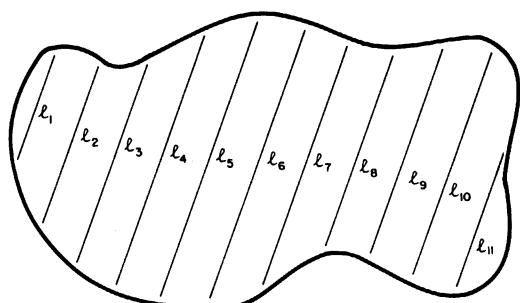


FIG. 18. Transect lines chosen in a systematic or nonrandom fashion. The initial transect location usually is selected as a random distance from the boundary of the area to be surveyed.

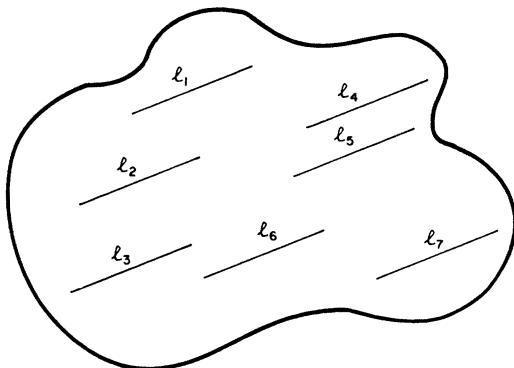


FIG. 19. Transect lines selected with random starting points and common (but random) direction. Lines may be of equal or unequal length. Note,
 $L = \sum \ell_i$.

as might be the case in a long-term study (Fig. 17). Use of permanent transects enables “pairing” of the data for the analysis of differences in density over time and thereby increases the power of such analyses.

When the regions to be surveyed have been selected, the layout of transects must be determined. That layout will vary according to statistical design requirements and ecological factors, as well as considerations of logistics, supply, and access. Some of the possible designs are shown in Figs. 18–22, but they by no means exhaust the possible designs. Individual line segments (ℓ_i) have been numbered arbitrarily in those figures. Data (x_i , r_i , and θ_i) should be recorded separately for each line segment.

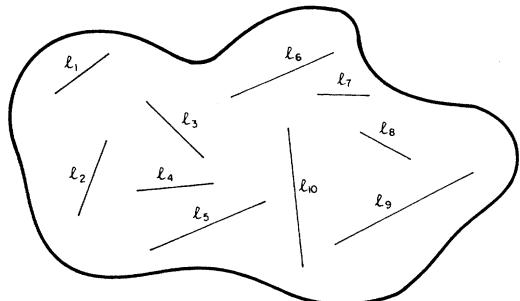


FIG. 21. Transect lines selected with random starting points, lengths, and directions. The lines should not intersect or cross. Such a design is not easy to implement, hence, it is not very practical.

The use of a field form (Fig. 23) is recommended as an aid in keeping careful records of data; however, forms may need to be tailored for each individual study.

Selection of a survey design is not a simple or automatic process. By considering the relevant factors carefully, one usually can devise an efficient and practical design that will yield informative and meaningful data. Two important principles that apply are (1) to obtain enough valid data to enable the conduct of proper statistical analyses, and (2) to avoid systematic errors or biases in the data, such as highly correlated or statistically dependent transect lines.

Sample Size Guidelines

Another aspect of survey design is the problem of how much data to collect (see

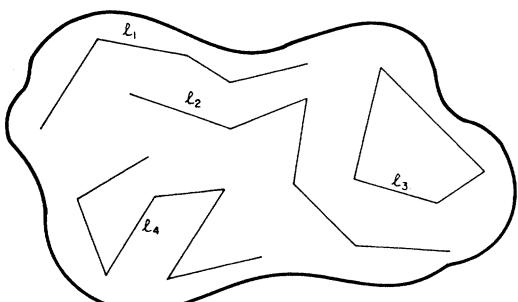


FIG. 20. Piecewise linear transects. Such a design may be necessary in rugged or nonhomogeneous terrain.

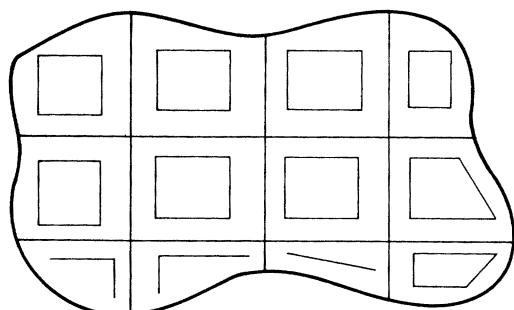


FIG. 22. The area is divided by sections, and each section is sampled separately. A variety of designs may be used within each section.

SURVEY FORM

Observer _____		Date _____	Survey Number _____	Area _____
Transect lengths (m): $\ell_1 =$ _____ $\ell_2 =$ _____ $\ell_3 =$ _____ $\ell_4 =$ _____ $\ell_5 =$ _____ $\ell_6 =$ _____ $\ell_7 =$ _____ $\ell_8 =$ _____ $\ell_9 =$ _____ $\ell_{10} =$ _____ $\ell_{11} =$ _____ $\ell_{12} =$ _____				
Description of Survey: _____				
Object number <i>i</i>	Transect number ℓ_i	Perpendicular distance x_i	Sighting distance r_i	Sighting angle θ_i

FIG. 23. Example of a field form.

Eberhardt 1978b), i.e., the classical problem of sample size. Certain minimal amounts of data must be obtained to enable the various statistical analyses to be done and valid inferences to be drawn. Too much data can be a problem, because cost is often nearly proportional to sampling effort (primarily to line length, L). Certainly, those 2 considerations need to be balanced as an integral part of the survey design. As a practical minimum, studies should be designed to assure that at least 40 total objects ($n \geq 40$) are detected; it might be preferable, if the total length (L) of the survey were sufficient, to allow the location of at least 60–80 objects.

The width, w, of the area searched on each side of the transect will influence sample size. It is recommended that w be large with respect to the expected average perpendicular distance. In fact, w need not be finite and there are theoretical advantages (mathematically speaking) to having $w = \infty$. The only danger one encounters with large w is the recording of "outliers," extreme values that do not belong in the data set; however, outlier data can be deleted during data analysis.

The selection of a sample size requires advance information on the variability of \hat{D} as a function of the line length. Some guidelines are provided below. This material is somewhat out of sequence in the sense that the sampling variance of \hat{D} will not be discussed until *Standard Errors and Confidence Intervals* under STATISTICAL INFERENCE BACKGROUND. The 2 equations needed are

$$E(n) = 2LD/\hat{f}(0)$$

and

$$(cv(\hat{D}))^2 = \frac{\text{var}(n)}{(E(n))^2} + \frac{\text{var}(\hat{f}(0))}{(\hat{f}(0))^2},$$

where $cv(\hat{D})$ is the coefficient of variation of \hat{D} and $\text{var}(\cdot)$ denotes a sampling variance. To a first approximation, $\text{var}(n) = a_1 n$ and $\text{var}(\hat{f}(0)) = (\hat{f}(0))^2 a_2/n$ can be written. Those relations are justified in many

models and have been confirmed by experience. The constants a_1 and a_2 are unknown parameters, and it is important to note that in a given study they are independent (or almost so) of n, L and $f(0)$. Thus, replacing $E(n)$ by just n,

$$(cv(\hat{D}))^2 = \frac{1}{n}[a_1 + a_2] = \frac{b}{n},$$

where b is an unknown parameter.

The main consideration in sample size determination is achieving a specified coefficient of variation on \hat{D} , such as 0.2, 0.1, or 0.05 (20, 10, and 5%, respectively). But unless b is known, n cannot be solved for. Finally, even when n is known, the needed line length as $L = nf(0)/2D$ must be solved for and requires knowledge of $f(0)$ and D, or at least knowledge of their ratio. Unfortunately, this latter information is likely to be totally unavailable without a good pilot study or access to the results of a similar, earlier study.

Assume that a pilot (or previous) study has been done. A survey of length L_1 was conducted, n_1 objects were seen and $cv(\hat{D}) = se(\hat{D})/\hat{D}$ was computed. That information can be used to estimate b and obtain the ratio $2L_1/n_1 = \hat{f}(0)/\hat{D}$. Then, given the desired $cv(\hat{D})$ for the study being planned, it is easy to solve for n and then solve for L.

For example, assume that a pilot study had $L_1 = 4$ km and yielded $n_1 = 33$ and $cv(\hat{D}) = 0.3$ (or 30%). Then,

$$b = n_1(cv(\hat{D}))^2 = 33(0.3)^2 = 3.$$

Assume that the goal is to estimate D with a coefficient of variation of 10 percent ($cv(\hat{D}) = 0.1$). Then,

$$n = b/(cv(\hat{D}))^2 = 3/(0.1)^2 = 300$$

and solving for line length by equating $(L/n) = (L_1/n_1)$ in this example gives $(L/300) = (4/33)$, and thus, $L = (300 \times 4)/33 = 36.4$ km.

The above manipulations can be combined into 1 useful equation:

$$L = \frac{b}{(cv(\hat{D}))^2} \left(\frac{L_1}{n_1} \right).$$

The quantities b and (L_1/n_1) must come from a pilot study or through an educated guess, if absolutely necessary. The desired coefficient of variation of \hat{D} must be supplied by the planner. Thus, there are only 2 unknowns, b and the ratio (L_1/n_1) . The latter ratio can vary greatly by survey for different species or objects. For example, one might see 10 duck nests per kilometer but observe only 0.002 porpoise schools per kilometer.

A pilot study to estimate L_1/n_1 can be quite simple. For one thing, no actual distances are needed, and one can obtain useful results with n_1 as small as 5 or 10. Thus, one could traverse randomly placed transects until sighting 10 objects and use the recorded value of L_1 in estimating L_1/n_1 . Such a pilot study should use the same value of w as is intended for the actual line transect study.

Determination of b from a pilot study is much less feasible: it would require a sample size of 40 or more and all the rigor of an actual survey. Fortunately, the value of b seems quite stable. Eberhardt (1978b) provided evidence that b may typically be between 2 and 4. Our own experience tends to corroborate that (see also Seber 1973:31). However, $b = 4$ arises with less efficient estimators of $f(0)$ than recommended here. A useful range of b is found to be 1.5 to 3, and to be conservative (more likely to overestimate the required sample size n), use $b = 3$. A value of $b = 2.5$ is tenable; however, use of $b = 1.5$ risks underestimating the needed line length.

As another hypothetical example, assume that a small pilot study is conducted with n_1 fixed at 5 and that $L_1 = 1$ km is traversed to achieve sighting 5 objects. This is insufficient to estimate D or $\text{var}(\hat{D})$. However, if we use $b = 3$ as a conservative value for planning, the needed line length is computed as

$$\begin{aligned} L &= \frac{b}{(\text{cv}(\hat{D}))^2} \left[\frac{L_1}{n_1} \right] = \frac{3}{(\text{cv}(\hat{D}))^2} \left[\frac{1}{5} \right] \\ &= \frac{0.6}{(\text{cv}(\hat{D}))^2}. \end{aligned}$$

To achieve a coefficient of variation of 0.15, it is thus expected that a line length of about 27 km will be required. This would yield an expected sample size of approximately $n = 135$.

When a given line transect study may be limited by available money or manpower, one might be able to specify in advance the maximum line length that could be sampled. Then, it is recommended to work the equations backwards to compute the expected coefficient of variation. If it is too large, the survey may be virtually useless (and hence, not worth doing). For example, if in the above example, no more than 10 km could be run, the expected coefficient of variation of \hat{D} would be

$$\text{cv}(\hat{D}) = \left(\frac{0.6}{10} \right)^{1/2} = 0.24,$$

or 24 percent. For some purposes this might be an unacceptably large $\text{cv}(\hat{D})$.

Those equations for line length determination are completely general. It is possible to develop equations under specific assumptions, such as a negative exponential model for $f(x)$ and a random distribution of objects (Gates et al. 1968, Seber 1973). Except as a lower bound on the required line length, however, we caution against such specific formulae (cf. Hayes unpublished thesis:85).

Finally, note that the line length (sample size) considerations given do not include cost considerations or specific aspects of the survey design. These considerations can be included in equations for allocation of sampling effort, but the resultant calculations become complex. If such added sophistication is necessary, an expert should be consulted.

Summary

Proper estimation of density by using line transect sampling schemes involves a mix of basic statistical sampling theory and knowledge of the biology of the population under study. Ten important points to consider in designing a survey are:

(1) The center line of the transect must be straight and well marked. The observer must be able to determine the position of the line at all times. In some cases, a series of straight line segments will suffice.

(2) Care must be taken to assure that objects on the center line of the transect are seen with probability 1. In practice, that requirement often can be met if the observer walks carefully along the center of the line transect at all times.

(3) The width of the transect should be taken as quite large, or effectively unbounded. Outlier data can be deleted, if necessary, during the analysis.

(4) All measurements of distances and angles must be accurate. A steel tape or other appropriate device should be used to assure a high degree of accuracy. Careless measurements and rounding errors lead to poor estimates of density and sampling variances.

(5) The 3 basic measurements should all be taken: perpendicular distance, sighting (flushing) distance, and sighting (flushing) angle.

(6) The measurements must be recorded separately for each segment or replicate line (say ℓ_j , where $L = \sum \ell_j$) of the total transect length.

(7) A target goal should be established for the precision of \hat{D} in terms of the coefficient of variation; then, the required line length should be found. As a practical minimum, studies should be designed to assure that at least 40 objects are seen ($n > 40$); it might be preferable if the length (L) were sufficient to allow the location of at least 60–80 objects ($n > 60–80$).

(8) A pilot survey should be made as an aid in planning the survey design. Often, a simple visit to the area to be surveyed, and basic biological information about the animal and its habits and habitat, will be sufficient to design a survey adequate to estimate density.

(9) The survey should be designed to assure that the population to be surveyed is not correlated with the sample line transects (e.g., avoid transects running

along roads, ridgetops, stream bottoms).

(10) The survey should be conducted by competent, interested, and trained personnel. This point is particularly relevant to points (1), (2), and (4) (above).

MATHEMATICAL BACKGROUND

Introduction

This section is a presentation of a basic mathematical model and some derived results for line transects. This material is oriented toward the mathematical reader; its comprehension is desirable, but not necessary to application of the results of PART 2. Most of this material is not new (see e.g., Seber 1973, Burnham and Anderson 1976), consequently, proofs are either much abbreviated or omitted.

The approach to the modeling of line transects and analysis of the data has been entirely in terms of the philosophy of the “frequentist” school of statistics. Moreover, the philosophy has been adopted that the observable data (n, x, r, θ) are random variables from a conceptually infinite “response” population and are thus characterized by some unknown probability distribution. This is in distinct contrast to the finite population approach sometimes used in wildlife sampling, in which results are developed on the basis of finite population sampling theory (such as survey sampling, see e.g., Cochran 1963). That philosophical approach requires finite N (population size), and associates a *fixed* sampling probability with each population member. To maintain the main thread of development, the distinction between those approaches is not elaborated here, but relegated to APPENDIX D (*Finite versus Infinite Population Sampling Theory Approaches*). Note, however, that we do not believe finite population sampling theory is appropriate for the general development of line transect theory.

In addition to those 2 conceptual statistical inference approaches to line transects, other possibilities include Bayesian, or decision theoretic. We have

chosen not to base our theory on those philosophies because they require additional inputs which the biologists or statisticians may be unwilling or unable to supply.

A General Model for Perpendicular Distances

The general assumptions were given under CONCEPTUAL BACKGROUND. Briefly, there is a density, D , of objects in the area to be sampled; objects need not be uniformly or independently distributed. There is a detection function $g(x)$ that represents the probability of detecting an object at perpendicular distance x . It is assumed $g(0) = 1$ and $g(x)$ is nonincreasing in x . Also needed is some notion of smoothness near $x = 0$ for the detection function (see Hayes unpublished thesis). Continuity is a desirable and natural property to assume, but not a sufficient one. In fact it has not been possible to conceive of a satisfactory mathematical criterion that will ensure $g(x)$ is well behaved so that anomalies cannot arise near $x = 0$.

Assume for the moment that w is finite (i.e., searching for objects is restricted to a strip of half width w about the line). Then, the unconditional probability of detecting a point in the strip is

$$P_w = \frac{1}{w} \int_0^w g(x)dx.$$

That result is based on random line placement. The area searched for objects is $A_w = 2Lw$. Consequently, the expected number of objects detected is

$$E(n) = DA_w P_w = D2LwP_w,$$

because DA_w is the expected number of objects in the strip and P_w is the unconditional (average) detection probability.

Define the quantity

$$\mu_w = \int_0^w g(x)dx.$$

Note that it is permissible to have $w = \infty$, i.e., the case of an unbounded area

about the line being scanned for objects. Now $wP_w = \mu_w$, and

$$E(n) = D2L\mu_w. \quad (1.4)$$

In Eq. (1.4) there are only 2 parameters: density D , and μ_w .

Corresponding to the n detected objects there are perpendicular distances x_1, \dots, x_n . The marginal pdf of any distance x is

$$f(x) = \frac{g(x)}{\mu_w}. \quad (1.5)$$

This is a key result proved by Seber (1973) (see also Burnham and Anderson 1976).

The assumption that points on the line are seen with certainty ($g(0) = 1$) implies $f(0) = 1/\mu_w$. Combining Eqs. (1.4) and (1.5) leads to $E(n) = D2L/f(0)$ which can be rewritten as

$$D = \frac{E(n)f(0)}{2L}. \quad (1.6)$$

The key problem in the analysis of line transect data is to estimate $f(0)$ based on the distances x_1, \dots, x_n . The resultant estimator $\hat{f}(0)$ is conditional on n .

If the estimator $\hat{f}(0)$ is consistent, then so is \hat{D} . Moreover, if $\hat{f}(0)$ is an unbiased estimator of $f(0)$, \hat{D} is an unbiased estimator of D . Assuming that $\hat{f}(0)$ is asymptotically unbiased, an equation for the asymptotic sampling variance of \hat{D} can be derived:

$$\begin{aligned} \text{var}(\hat{D}) &= E(\hat{D} - D)^2 \\ &= \frac{E[n\hat{f}(0) - E(n)f(0)]^2}{(2L)^2} \\ &= \frac{f^2(0)\text{var}(n) + E_n(n^2\text{var}(\hat{f}(0)|n))}{(2L)^2}. \end{aligned} \quad (1.7)$$

Without making further assumptions, Eq. (1.7) cannot be simplified. It can, however, be rewritten as

$$\begin{aligned} \text{var}(\hat{D}) &= (D)^2 \left[\frac{\text{var}(n)}{(E(n))^2} \right. \\ &\quad \left. + \frac{E_n(n^2\text{var}(\hat{f}(0)|n))}{(E(n))^2(f'(0))^2} \right], \end{aligned}$$

from which it is evident that an estimator for $\text{var}(\hat{D})$ is

$$\text{var}(\hat{D}) = (\hat{D})^2 \{(\hat{c}\text{v}(n))^2 + [\hat{c}\text{v}(\hat{f}(0))]^2\}. \quad (1.8)$$

In that equation, $\text{cv}(\cdot)$ stands for coefficient of variation; for an arbitrary random variable y , $(\text{cv}(y))^2 = \text{var}(y)/(\text{E}(y))^2$.

Under some circumstances (e.g., parametric estimators of $f(0)$), it is reasonable to assume that the asymptotic variance of $\hat{f}(0)$ equals b/n for some constant b . The $\text{var}(\hat{f}(0))$ will then usually be estimable from the procedure used to get $\hat{f}(0)$. It is the sampling variance of n that is not known. Under the assumption of a random (Poisson) distribution of objects, $\text{var}(n) = \text{E}(n)$, thus, the ratio $\text{var}(n)/\text{E}(n)$ is 1. Typically, objects will be somewhat aggregated and it is expected that $\text{var}(n) > \text{E}(n)$, but it may often be assumed that $\text{var}(n) = a\text{E}(n)$, for some constant $a > 1$ (cf. Eberhardt 1978a). Thus, the following result will often be applicable:

$$\text{var}(\hat{D}) = \frac{1}{L} \left[\frac{\text{D}\hat{f}(0)}{2} \right] \left[a + \frac{b}{(\hat{f}(0))^2} \right].$$

This shows that $\text{var}(\hat{D})$ may be expected to be proportional to $1/L$ (equivalently, it is proportional to $1/\text{E}(n)$). That observation forms the basis of our recommendation that replicate estimates of D be weighted by line length.

The results given above are completely general, even though they started with $w < \infty$. The quantity w does not enter any of the equations except through $f(0) = 1/\mu_w$ and $\mu_w = \int_0^w g(x)dx$. Assuming that that integral is finite as $w \rightarrow \infty$, we may let w be finite or infinite and all the above equations hold. Also, nothing was said about the manner of estimating $f(0)$ on the basis of perpendicular data. Thus, the data may be ungrouped or grouped into frequency classes.

Let $0 < c_1 < c_2 < \dots < c_k$ define k perpendicular distance classes (we refer to the c_i as cut points). Let $c_0 \equiv 0$ and $c_k \equiv w$. Let n_i be the number of observed distances falling in interval i (i.e., in $[c_{i-1}, c_i]$). Given n , the frequencies n_1, \dots, n_k have

a multinomial distribution with cell probabilities

$$p_i = \int_{c_{i-1}}^{c_i} f(x)dx.$$

Estimation of $f(0)$ can be based on this multinomial distribution any time $f(x)$ is modeled as a function of unknown parameters (some estimators can be used without such a parametric model).

The point is made that the above results on \hat{D} and $\text{var}(\hat{D})$ are of substantial generality. Not only do they apply if estimation is based on perpendicular data, they are equally valid if $f(0)$ is estimated from sighting distances and angles.

Estimation Based on Sighting Distances and Angles

The general approach for perpendicular distances given above has an analog for sighting distances and angles. Much of this material is from Burnham and Anderson (1976). Given that an observation was made, it was made at some sighting distance $r \geq x$. Therefore, there exists a density function $h(r|x)$ that represents this conditional probability distribution of r given x . There can now be a generalization of $g(x)$ to a function $g(x,r) = g(x)h(r|x)$. The joint pdf of x and r is given by

$$f(x,r) = f(x)h(r|x) = \frac{g(x,r)}{\mu_w}. \quad (1.9)$$

Rearranging Eq. (1.9) it can be conceptualized as an extension of the detection function to incorporate both x and r :

$$g(x,r) = f(x,r)\mu_w = f(x)h(r|x)\mu_w.$$

Hayes (unpublished thesis) has considered that conceptual extension of $g(x)$. If one allows the convention that $-r$ denotes a sighting occurring after one has passed the point perpendicular to the object (hence $\theta > 90^\circ$), it shows that $g(x)$ can be expanded conceptually to incorporate r regardless of where the observer is at the time of detection. Hence, the model recognizes that detection depends pri-

marily (if not solely) on the well-defined value of x , but that the physical detection can occur anywhere along the line. The relationship of $g(x)$ to $g(x,r)$ is

$$g(x) = \int_x^{\infty} g(x,r)dr + \int_{-\infty}^{-x} g(x,r)dr.$$

Because most sightings occur before one is abreast of the object (let alone beyond it), $g(x,r)$ is very asymmetric in r ; the term $\int_{-\infty}^{-x} g(x,r)dr$ would be a very small component of $g(x)$.

Such a conceptualization shows that the detection function can be made a function of both x and r without loss of generality. The process by which objects are detected is irrelevant in the model of perpendicular distances, as are the associated data on r and θ . However, to model the sighting distance and angle data, the detection process itself must be considered. Unfortunately, not much is known about the process, and it can vary drastically for different applications. Some attempts to model the sighting process are presented, but the models are based neither on substantive information nor on understanding of the underlying processes.

The joint density function of x and r may also be written as $f(x,r) = f(x|r)h(r)$ where $h(r)$ is the unconditional pdf of sighting distance and $f(x|r)$ is the conditional pdf of x given r , $0 < x < r$. Now $f(x)$ can be expressed as

$$f(x) = \frac{g(x)}{\mu_w} = \int_x^{\infty} h(r)f(x|r)dr.$$

By evaluating the above equation at $x = 0$ the key result is obtained for estimation of $f(0)$ and hence D , from sighting distances r_1, \dots, r_n :

$$f(0) = \int_0^{\infty} h(r)f(0|r)dr = E_h(f(0|r)). \quad (1.10)$$

If the transformation $f(0|r)$ were known, an estimator of $f(0)$ would be

$$\hat{f}(0) = \frac{1}{n} \sum_{i=1}^n f(0|r_i). \quad (1.11)$$

In practice $f(x|r)$ is not known, hence $f(0|r)$ is not known. Thus, estimation of D based on sighting distances requires the assumption of a form for $f(0|r)$.

The best way to determine $f(0|r)$ is by theoretical or empirical specification of the entire distribution $f(x|r)$. Theoretical consideration of $f(x|r)$ is much simpler if $w = \infty$. Also, because it is impractical to specify a finite w if only sighting distances and angles are being recorded, we take $w = \infty$ in the remainder of this section.

Under the assumption that sightings are independent events, each r_i generates an independent estimator $\hat{f}_i(0) = f(0|r_i)$. Their average is $\hat{f}(0)$, Eq. (1.11), and the general, empirical estimator of $\text{var}(\hat{f}(0)|n)$ is

$$\text{var}(\hat{f}(0)|n) = \frac{\sum (\hat{f}_i(0) - \hat{f}(0))^2}{(n-1)n}. \quad (1.12)$$

The Hayne estimator (Hayne 1949) is derived easily if the model assumed is one of a fixed circular flushing radius for each animal (e.g., Overton and Davis 1969). This implies that $f(x|r) = 1/r, x \in [0,r]$ and from Eq. (1.10) $f(0) = E(1/r)$, hence

$$\hat{D}_H = \frac{n}{2L} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right],$$

which is the Hayne estimator. Note that the derivation does not rely on finite population sampling theory (cf. Eberhardt 1978a).

Hayne's estimator originally was based on the mathematical assumption that $\sin(\theta) = x/r$ had a uniform distribution on the interval $(0,1)$. This can be generalized by assuming that the distribution of x/r is independent of r but otherwise not specified. Therefore, a canonical distribution, $c(x/r)$, exists such that

$$f(x|r) = \frac{1}{r} c(x/r), \quad x/r \in [0,1].$$

Now, $f(0|r) = c(0)/r$, and \hat{D} is a simple modification of the Hayne estimator:

$$\hat{D} = \hat{c}(0) \left[\frac{n}{2L} \frac{1}{n} \sum \frac{1}{r_i} \right]. \quad (1.13)$$

In terms of the equation $\hat{D} = nf(0)/2L$, the estimator of $f(0)$ in Eq. (1.13) is

$$\hat{f}(0) = \hat{c}(0) \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right].$$

To take advantage of Eq. (1.7) for the variance of \hat{D} it is necessary to find a general expression for the variance of $\hat{f}(0)$ above. The estimator of $\hat{c}(0)$ is based on the sines of the sighting angles, which by assumption are independent of sighting distances. Therefore, a variance equation (conditional on n) for $\hat{f}(0)$ is derived easily ($c \equiv C(0)$ is used):

$$\begin{aligned} \text{var}(\hat{f}(0)) &= (c)^2 \frac{1}{n} \text{var}\left(\frac{1}{r}\right) \\ &\quad + \left(\frac{\hat{f}(0)}{c}\right)^2 \text{var}(\hat{c}(0)). \end{aligned} \quad (1.14)$$

If $c(0)$ is a known constant, $\text{var}(\hat{c}(0)) \equiv 0$. Let $\bar{r}_H = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i}$, then Eq. (1.14) can be written as

$$\text{var}(\hat{f}(0)) = (\hat{f}(0))^2 [(\text{cv}(\bar{r}_H))^2 + (\text{cv}(\hat{c}(0)))^2]. \quad (1.15)$$

Estimators of D have been derived based on this model of independence of r and θ (see *Estimation Based on Sighting Distances and Angles* in PART 1). However, the approach can be developed in other ways. The basic problem is to model the joint distribution of either (r, x) or (r, θ) . Very little has been published along those lines; the notable exception is Gates (1969). No unique approach exists; one can model $f(x, r) = f(x|r)h(r)$ or $f(x, r) = h(r|x)f(x)$. The latter case hardly makes sense because if a model for $f(x)$ is available, one can develop an estimator of $f(0)$ based strictly on perpendicular distances. In either case, it is evident that more assumptions are required to model and analyze sighting distance and angle data than to work only with the perpendicular distance data. Bearing in mind that r and θ can always be transformed into x , little advantage accrues from working with sighting distances and angles.

Perpendicular distance data can be dealt with readily whether they are grouped or ungrouped. No methods have been presented in the literature for the analysis of sighting distances and angles when one or both measurements are grouped. However, such methods can be derived on the basis of an assumed parametric model for $f(x, r)$. For ease of discussion, assume that θ and r are independent. Then $f(x, r) = \frac{1}{r} c(\sin(\theta))h(r)$ where both $c(\sin(\theta))$ and $h(r)$ are parametric functions. The parameter $f(0)$ is

$$f(0) = c(0) \int_0^\infty \frac{1}{r} h(r) dr,$$

assuming $h(r)$ is such that the integral exists. Now, $c(0)$ is a parametric function of the parameters of $c(\sin(\theta))$, and as such can be estimated from grouped or ungrouped data. Similarly,

$$\alpha(h) = \int_0^\infty \frac{1}{r} h(r) dr$$

is a parametric function of the parameters of $h(r)$, and can thus be estimated from grouped or ungrouped sighting distances.

Such an approach would allow rigorous analysis of grouped sighting distance and angle data, but its use requires that formidable assumptions be made about the distribution $f(x, r)$. If those data were grouped, they could not be transformed reliably to perpendicular distances. Until more realistic models are developed for that situation, the solution is to take all sighting distance and angle data carefully (not by intervals) and then transform to perpendicular distance.

STATISTICAL INFERENCE BACKGROUND

Introduction

A general form of an estimator for density from line transects is $D = nf(0)/(2L)$. It is maintained that the central problem for analysis of line transect data is to obtain the estimator of $f(0)$ from the distance data. Many features of that problem can

be discussed in the abstract, especially the problem of estimating the sampling variance of \hat{D} given an estimator of $f(0)$. Also, criteria must be established for the modeling of $f(x)$ and the subsequent selection of an estimator. Other general subjects include the role of assumptions, the definition of some terms, and the role of testing the goodness of fit of the model in line transect data analysis. Those subjects are all discussed in this section, and because this material is regarded as essential background to much of the rest of this monograph, especially PART 2, it is intended for the biologist as well as the statistician.

The Central Problem

The Role of Assumptions

Although assumptions are a necessary part of any modeling or statistical inference problem, unjustified assumptions should not be made. Assumptions are justified only when they are based on prior knowledge, are so general as to be believable, or have been shown to be acceptable after rigorous testing procedures. One must avoid specific assumptions that may be false and that, if false, invalidate the model or the desired inferences.

At its simplest, line transect sampling involves few assumptions. The assumption that there are objects distributed in the study area is innocuous because one does not have to assume anything about the spatial distribution of objects. In particular, a random distribution is not necessary for any developments (cf. Quinn unpublished thesis, Hayes unpublished thesis, Eberhardt 1978a). The only real assumptions made were presented under CONCEPTUAL BACKGROUND and are repeated here, in order of importance:

- (1) Objects directly on the line will never be missed, thus $g(0) = 1$.
- (2) Objects are fixed at the initial sighting position (they do not move before being sighted) and none are counted twice.

- (3) No measurement errors and no rounding errors occur.
- (4) Sightings are independent events.

Primarily on the basis of assumptions 1 and 2, the probability density function (pdf) of perpendicular distance, $f(x)$, is $g(x)/\mu$, where $\mu = \int_0^w g(x)dx$. That model (set of assumptions) provides the necessary link between the observed data (n and x_1, \dots, x_n) and the parameter of interest D , because $\hat{D} = nf(0)/2L$ and $f(x)$ is totally determined by the detection function $g(x)$. The key problem is to derive an estimator of $f(0)$. That may be done in many ways, although most of them require making some further specific assumptions about either $g(x)$ or $f(x)$. For example, the assumption that $f(x)$ is a particular function type (e.g., negative exponential or half-normal) enables one to derive an estimator of $f(0)$ and hence of \hat{D} . But if the assumption is wrong, the derived estimator is likely to be biased (wrong), and perhaps badly so.

The point is that the specific assumptions made about $g(x)$ or $f(x)$ are crucial. The weaker those assumptions, the better in terms of deriving unbiased estimators of D regardless of the true (unknown) detection function. Unfortunately, the weaker the assumptions about $g(x)$, the harder it is to derive a precise statistical estimator of $f(0)$. In statistical terminology, one is faced with a trade-off of precision (sampling variance) versus accuracy (bias).

Assumptions are equated with the model. The model defines the relation between the parameters of interest and the sampled data, and thus guides the selection of parameter estimators. An estimator is a method of computing, from the data, a number presumed to be close to the true unknown parameter. Here, "closeness" is measured by such statistical properties as precision and bias. Without a well-defined model, there is no basis for an estimator. When there is no model, an estimator is defined as *ad hoc*. The wildlife literature abounds with ad hoc estimators, the validity of which are typically unknown.

It is virtually impossible to justify selecting a unique model (hence estimator) for a given set of data based strictly on analysis of that set of data. Primarily, this is because with any set of real data, many different models will fit, as judged by, say, the usual chi-square goodness of fit test (see *Chi-Square Goodness of Fit Tests*). Moreover, failure to reject a model (i.e., an “acceptable” fit is found) is not valid evidence that it is the “true” model for the data. There are lots of data sets where several models fit, yet their point estimates and assumptions are quite different (see *Example 9*, also see APPENDIX B). A strategy for model selection must be as independent of the data as possible. That is, the choice of a model should depend largely upon prior considerations, not just the analysis of the one set of data in hand.

Definition of Some Terms

The main concept to be presented here is the idea of a *robust* estimator. Roughly speaking, an estimator of D is robust if it has small bias relative to its standard error regardless of the true detection function. Thus, robust estimators do not depend on any specific assumptions about the form of $f(x)$. It is recommended that robust estimators of $f(0)$, hence of D , be used.

Because of their importance, an attempt is made to clarify the 3 terms parametric estimation, nonparametric estimation, and robustness in the context of estimation of $f(0)$. Let $f(x)$ be the unknown pdf of perpendicular distances. A parametric approach would involve assuming that $f(x)$ was a member of a family of proper pdfs of known functional form but dependent on an unknown parameter θ (possibly vector valued). Given that approach to specifying the problem, any of a variety of approaches to estimating θ will lead to $f(0) = f(0; \hat{\theta})$. Thus, a parametric approach is not defined as one that simply involves estimation of parameters.

In a nonparametric approach, no spe-

cific distribution assumptions are made about the random variables under study. Conceptually, there exists a pdf $f(x)$, but one takes the approach of assuming an approximate representation for that pdf (see Tapia and Thompson 1978 for an introduction to nonparametric density function estimation). That representation will often be based on the mathematical theory of approximations by means of various types of representations, e.g., polynomials, Fourier series, or splines (kernels are yet another nonparametric approach, but they are not based on mathematical approximation theory). Using that approach, an approximation is obtained to $f(x)$ as a function of known form and unknown parameters. However, the number of terms of the representation may not be defined a priori. Instead, the number of terms must be determined for any sample. Also, there is no well-defined parameter space and the chosen representation may not be a true pdf, either because it is not constrained to the strictly positive, or because it may not be constrained to integrate to 1.

In the parametric approach, the dimensionality of θ is fixed for all sample sizes. In some of the nonparametric approaches (e.g., Fourier series), the number of unknown parameters in the model will increase as n increases. Thus, the approximation to the true $f(x)$ improves as sample size increases.

The method used to estimate the parameters that enter the representation of $f(x)$ is not particularly relevant. Consequently, one can call the Fourier series a nonparametric approach and proceed to talk about estimation of parameters in the FS representation of $f(x)$. Maximum likelihood (ML) estimation could even be attempted, but the asymptotic properties of ML estimates could not be deduced from standard theory.

No clear-cut distinction exists between nonparametric and robust estimators (nor between those and generalized parametric estimators). However, robustness best describes the property that line transect estimates should have. Unfortunately, the concept of robustness has never been (and

cannot be) defined precisely; for a discussion of this, see Bradley (1968:25–26).

Robustness of an estimator relates to its insensitivity to specific, usually distributional, assumptions. But it goes further than that: robust estimators must have an “acceptable” value of the ratio bias/standard error, and that ratio should be relatively insensitive to specific distributional assumptions. An arbitrary estimator selected before sampling could be used, and although it would be robust in the sense of not depending upon any distributional assumptions, it would also be unacceptable. Robustness implies some concept of small bias. We are unable to quantify that element, but we judge that an estimator would qualify as robust if the ratio of bias to standard error were less than 0.5 for a wide range of possible true distributions.

Nonparametric estimators tend to be robust. However, one can conceive of a parametric approach based on a functional form sufficiently general to “cover” whatever the true pdf is likely to be. A mixture of an exponential and half normal distribution certainly is more robust than either simple model would be by itself. Among other possible estimators, the exponential polynomial, exponential power series, reversed logistic, and incomplete gamma should also be robust (see COMMENTS ON OTHER ESTIMATORS, PART 4). Given that a variety of robust estimators exist, one needs a way of choosing among them. A small standard error is one consideration. Conversely, it is considered that ease of computation is not a major consideration because of the existence of computers in general, and of program TRANSECT in particular.

Robust approaches exist whether or not the data are grouped, and whether or not the data were taken truncated. What matters most is the general approach one takes to modeling $f(x)$, not the actual mechanics of deriving an estimator (the estimators may not be numerically simple, but they can be computed). Our philosophy is that estimators of $f(0)$, and hence D , should be chosen deliberately to be

robust because specific assumptions about $f(x)$ are not justified.

Criteria for Robust Estimation of $f(0)$

Introduction

There are 2 types of criteria for robust estimation of $f(0)$: (1) criteria that relate to the properties of the assumed model for $f(x)$, and (2) criteria that relate to properties of the data. Burnham et al. (1979) gave 6 criteria for robust estimation of $f(0)$, in order of importance, they are:

- (1) *Model robustness,*
- (2) *Pooling robustness,*
- (3) *Shape criterion,*
- (4) *Estimator efficiency,*
- (5) *Data truncation, and*
- (6) *Data grouping.*

The first 4 criteria are properties of models or estimators (independent of the data being analyzed), whereas the last 2 relate to properties of the data (independent of the model or estimator being used).

Model Robustness

One can model $f(x)$ directly, or by first assuming a model for the detection curve $g(x)$. Knowledge of line transect sampling is insufficient to specify accurately the true detection curve, $g(x)$. Moreover, the shape of the curve probably varies with such variables as the species and habitat sampled, the observer, and weather conditions among others. In the face of that ignorance about the detection probability curve and what can affect it, it is recommended that a general, flexible model be assumed for $f(x)$. An estimator based on such a general model will be called *model robust*. That is, it will be flexible enough to fit closely a wide variety of true $f(x)$ shapes. Such a general model can therefore be used to devise an estimator of animal density that does not depend on an unjustified assumption about the specific form of $f(x)$.

Specific parametric estimators such as

the negative exponential or half-normal, are not *model robust*. If the true probability density function $f(x)$ differs from those assumed forms, the computed density estimate will be biased and perhaps badly so.

Pooling Robustness

Model robustness does not significantly reduce the number of possible candidate models. It is proposed that the next most important model criterion for $f(x)$ is that the estimator should be robust to variations in detection probability. The probability of detecting an object at any given perpendicular distance will vary according to numerous factors, both external and internal to the population being sampled. Examples of external factors are replicate lines, weather, light conditions, habitat, observer, and method of observation. Examples of internal factors are the species (if several species are being sampled), age, sex and size of animals, alertness of individuals or their varying conspicuousness, and the size of groups (when individuals are found in groups).

Most line transect studies will extend over a period of time and probably include replicate lines. There may be more than 1 observer, and conditions are likely to vary during the sampling of a given line as well as between lines. For those reasons, and those given above, the detection curve will vary during the survey. Thus, one can think of the total data as arising from pooling subsets of the data that correspond to differing detection curves. The estimator of density should be robust to such pooling, i.e., it should not be affected by such variations in detection probability.

As discussed above, line transect data are inexorably pooled over many factors that affect $g(x)$. Although some of the factors might be recognized in the field while sampling, many would not be, and stratifying the data to resolve the problem completely is therefore not possible. The criterion that the model for $f(x)$ be

robust to variations in detection probability (for fixed x) will be called *pooling robustness*.

To characterize pooling robustness and more rigorously define it, let there be r strata (subsets) of the data, with sample sizes n_1, \dots, n_r . Then, total sample size $n = n_1 + \dots + n_r$. Let one general model for $f(x)$ be assumed for all subsets and let the same estimator, $\hat{f}(0)$, be used for all subsets as well as for the overall pooled data. Two cases must be considered: variations in detection probability are due to either internal or external reasons (as discussed above). In the case of internal variations, the population being sampled is conceptualized as having r classes (strata). For each class (which could be based on factors such as species, age, sex, group size), there is a detection curve and a corresponding probability density function $f_j(x)$. For each class, let the true density be D_j ; overall density is $D = D_1 + \dots + D_r$. The corresponding estimator for each class is $\hat{D}_j = n_j \hat{f}_j(0)/2L$, $j = 1, \dots, r$. The overall stratified estimator of density, \hat{D}_s , is the sum

$$\hat{D}_s = \hat{D}_1 + \hat{D}_2 + \dots + \hat{D}_r.$$

Let $\hat{D}_p = n\hat{f}(0)/2L$ be the pooled estimator of density. That is, \hat{D}_p is derived from Eq. (1.3) by computing $\hat{f}(0)$ from the pooled data. The estimation method based on $\hat{f}(0)$ is defined to be *pooling robust* if $\hat{D}_s \equiv \hat{D}_p$, i.e., if the overall estimator of density is identical whether data are analyzed pooled, or are analyzed stratified, and the results are then properly combined. Note that both \hat{D}_s and \hat{D}_p are estimating exactly the same parameter, D = total density of sampled objects.

As described above, $\hat{D}_s \equiv \hat{D}_p$ is equivalent to the condition

$$\begin{aligned} n\hat{f}(0) &= n_1\hat{f}_1(0) + \dots + n_r\hat{f}_r(0) \\ &= \sum_{j=1}^r n_j\hat{f}_j(0). \end{aligned} \quad (1.16)$$

Consequently, Eq. (1.16) is now defined as being the criterion of *pooling robustness* for any estimator $\hat{f}(0)$. The proof that

this also applies in the case of external variation in the detection curve is given below.

Let sighting conditions change during the survey so that the total length of line L can be partitioned into segments (which might actually be separate lines) ℓ_1, \dots, ℓ_r , such that $L = \ell_1 + \dots + \ell_r$. Corresponding to those line lengths, the data are stratified into subsets (strata) of sizes n_1, \dots, n_r . Estimates of density for each stratum are

$$\hat{D}_j = \frac{n_j \hat{f}_j(0)}{2\ell_j}.$$

To compute one combined estimate of density for all strata, one should take a weighted average of $\hat{D}_1, \dots, \hat{D}_r$, weighting by ℓ_j (Burnham and Anderson 1976: 329):

$$\hat{D}_s = \frac{\sum_{j=1}^r \ell_j \hat{D}_j}{\sum_{j=1}^r \ell_j} = \frac{\sum_{j=1}^r n_j \hat{f}_j(0)}{2L}.$$

The criterion of pooling robustness ($\hat{D}_s \equiv \hat{D}_p$) applied to the above result is equivalent to Eq. (1.16).

Equation (1.16) defines the condition under which an estimator $\hat{f}(0)$ is *pooling robust*. That is, if an estimator satisfies that simple condition, then the same point estimate of animal density will be obtained whether data are analyzed pooled or whether they are stratified, analyzed by strata, and the results properly combined into one overall estimate. As defined by Eq. (1.16), *pooling robustness* can be verified or disproved for any proposed estimator.

If it were always possible to stratify the data and analyze each stratum separately, *pooling robustness* would not be needed. However, the factors that affect the sighting probabilities often cannot all be identified or measured and therefore the appropriate stratifications often cannot be done. Even when a relevant variable can be measured (e.g., group size, weather conditions, habitat type), many of the resultant strata will have sample sizes so

small that it is not possible to compute an estimator of density for them. The practical use of stratification is thus quite limited in most line transect surveys.

We maintain that all data, even after any feasible stratification, must be considered as pooled over many factors that affect the detection curve. It is concluded that estimators of density for line transect data should be both *model robust* and *pooling robust*.

One caution regarding *pooling robustness* is necessary. Assume an area A is to be sampled by using a stratified design. Let the subareas (strata) have sizes A_1, \dots, A_r . Further assume that density differs by area and that one goal of the study is to estimate average density over the entire area of size A. The pooling robustness property will not produce an unbiased estimator unless the total line length is allocated in proportion to subareas. That is, the design should have $\ell_i = LA_i/A$. (That constraint was also mentioned under FIELD SAMPLING PROCEDURES AND STUDY DESIGN, and is discussed in APPENDIX D, *Pooling Robustness and Area Stratification*.)

The arguments that Eq. (1.16) defines *pooling robustness* have been presented here in greatly simplified form. Those arguments can be expanded to allow more realism, for example, with the simultaneous occurrence of internal and external variation. Also, we have dealt only with estimation based on perpendicular data; those same considerations, arguments, and criteria are applicable to estimators based on sighting distance and angles, for example, the well-known Hayne estimator (Hayne 1949). As Eberhardt (1978a:21) stated, the Hayne estimator is robust. However, we find that that robustness is guaranteed only if the sine of the sighting angle has a uniform probability distribution on the interval [0,1]. Given that one critical assumption, the Hayne estimator can be shown mathematically to be *model robust* and *pooling robust*; but without that assumption it is not at all robust.

There is another way the Hayne esti-

mator (and its various modifications) fails to be robust. If even 1 recorded sighting distance, r , is *very* close to zero (relative to the rest of the sighting distance data), it has an enormous effect on the Hayne estimator. This is, of course, because one averages the *reciprocals* of the distances r_i . In the extreme, even one sighting distance of zero would lead to an arbitrarily large estimate of density. In practice, zero sighting distances (which implies a perpendicular distance of zero) must be discarded in applying Hayne's estimator (or its modifications). Zero perpendicular distances, however, cause no problem at all for any methods based on perpendicular distance data.

Shape Criterion

Although the robustness criteria given above serve to restrict the class of estimators for $f(0)$, the resulting class of candidate models for $f(x)$ is still very large. It is proposed that an additional criterion of model selection is to consider, at least in general terms, the probable shape of the true detection curve $g(x)$. We believe that $g(x)$ ought to have a "shoulder" near $x = 0$. It is believed that in some, perhaps very small, region near the line $g(x)$ will in fact be 1. This implies that $g'(0) = 0$, i.e., the derivative of $g(x)$ is 0 at $x = 0$.

The shoulder condition on the shape of $f(x)$ has been suggested by others (e.g., Eberhardt 1978a:11). It is a logical condition suggested by the physical sighting process itself, and it will be true if and only if $f'(0) = 0$ is true. Although this shape criterion is somewhat vague, the condition $f'(0) = 0$ is very specific and easy to check. We recommend imposing it on all line transect models unless there are theoretical reasons not to do so.

An example of a model for $f(x)$ which is model robust, pooling robust (with appropriate linear estimators of the parameters) and satisfies $f'(0) = 0$ is the polynomial without the linear term:

$$f(x) = a_0 + a_2x^2 + a_3x^3 + \dots + a_nx^n.$$

Estimator Efficiency

Another criterion that should be used in selecting an estimator is statistical efficiency, i.e., the estimator $f(0)$ should have as small a sampling variance as possible. Although that criterion should be imposed only after the 3 preceding criteria have been met, it should be included because it will eliminate inefficient estimators.

Data Truncation

The 4 aspects of estimation described above are features of models and estimators that apply without reference to the data. Some practical statistical tools of data manipulation also can enhance robustness, regardless of the model or estimator. Two techniques for data analysis that are useful in line transect analysis to achieve more robustness are data truncation and grouping of data. If one could collect "perfect" data, there would be no need for such techniques. However, data often exhibit 2 problems: outliers and "heaping." Outliers are discussed here and heaping is discussed under Data Grouping (below).

Note that robust estimation is a subject of much concern in statistics and that one of the primary tools used to create robust estimators (in many contexts) is truncation or "trimming" of extreme data values (cf. Andrews et al. 1972). In line transect sampling, extreme values are those few, very large perpendicular distances that are often obtained. In the context of parametric analysis, those extreme observations would be called outliers and one would attempt to identify and eliminate them.

As an hypothetical example, assume that one obtains a sample of 53 perpendicular distances; 51 of them are less than 30 but the other 2 are at 91 and 131 m. Those 2 are sharply out of line with the bulk of the data. Including them in the estimation of $f(0)$ will tend to bias $\hat{f}(0)$ under most types of models rather than improve it. We recommend routine ex-

TABLE 1.—PERPENDICULAR DISTANCE DATA, ILLUSTRATING SOME OF THE DATA FEATURES THAT COMMONLY MOTIVATE TRUNCATION AND GROUPING. SAMPLE SIZE IS 197; x DENOTES PERPENDICULAR DISTANCE, n_x DENOTES THE FREQUENCY OF OCCURRENCE OF THAT DISTANCE. DISTANCES WERE RECORDED TO THE NEAREST FOOT (1 FOOT = 0.3048 m)

x	n_x	x	n_x
0	33	24	2
3	3	25	8
4	1	27	30
5	4	30	22
6	3	32	2
7	2	33	2
8	3	35	5
9	2	36	4
10	17	38	1
11	1	40	5
12	8	42	2
14	2	45	1
15	14	47	2
16	1	48	2
17	1	50	3
18	10	52	1
20	11	54	2
21	4	66	2
22	3	90	1
23	1	100	1

amination of data for such extreme values and their elimination by truncation. This means that one specifies a distance w^* and discards all data at distances greater than w^* . Normally, only 1–3 percent of the data should be truncated. If the data were taken truncated (a designated value of w), it may not be necessary to truncate. However, we often recommend taking data untruncated and truncating after the fact if necessary.

A real set of perpendicular distances that exhibits both outliers and heaping is given in Table 1. Those data are from a 1977 terrestrial application of line transects to sage grouse *Centrocercus urophasianus*. The study was conducted carefully by interested, trained observers. The total sample size of 197 is good. The values of 90 and 100 are extreme, and we would recommend truncating them before analysis. The 2 values at $x = 66$ are also a little out of line with the bulk of the data; in practice, we would also eliminate them. Thus we would

truncate the data at some $w^* < 66$, say $w^* = 56$. That truncation discards only 4 of 197 values (2%) and does, in fact, improve the results of analyses of these data. For the Fourier series estimator (see PART 2, *Fourier Series Estimator*) the exact choice of w^* is not critical.

Data Grouping

Grouping is done when one constructs a frequency histogram of data; grouping may be thought of as a smoothing device. The problems with field data that motivate such smoothing are certain systematic inaccuracies in recording distances. Those inaccuracies show up as a “heaping” of values at certain convenient distances. For example, 33 perpendicular distances are recorded as zero in Table 1, yet only 3 values are at 3 feet, and only 1 value is at 4 feet. Furthermore, 17 values are recorded at 10 feet, 14 values at 15 feet, 11 values at 20 feet, and 22 values at 30 feet (1 foot = 0.3048 m). Those simply cannot be accurate recordings of the true distances of the animals from the transect center line; recording inaccuracies must have occurred.

For objects very near the line, the perpendicular distance is commonly recorded as zero when, in fact, it should often be a small positive value. At greater distances, there is a strong tendency to record distances as certain convenient values, especially 5, 10, 15, and 20. That tendency is very common when people are asked to judge exact quantitative values rather than to make precise measurements. Social scientists have labeled that phenomenon “heaping” (see Hobson 1976). The more subjective the distance measurement, the more heaping will occur, and the more likely it is that there will be inaccuracies at very small perpendicular distances.

Grouping, with appropriate intervals, reduces the impact of such data problems, and can lead to more robust estimation of $f(0)$, depending on the model and estimator used. To achieve smoothing, the data should be grouped in no

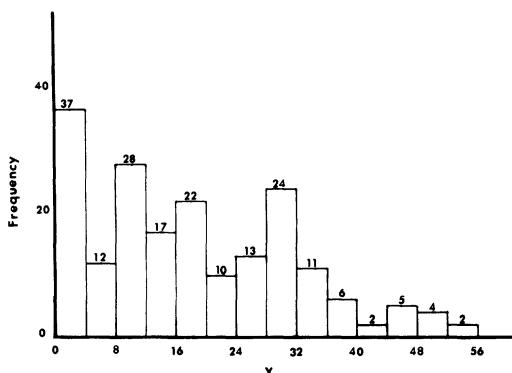


FIG. 24. Data of Table 1 grouped into 14 equal length intervals.

more than about 10 intervals. Figs. 24 and 25 show the frequency histogram and actual data values for the data of Table 1 after truncation at $w^* = 56$, for 14 and 7 intervals of equal lengths 4 and 8 feet, respectively. It is noted that grouping the data based on 14 intervals is still too fine to smooth the data adequately. With 7 groups, however, the histogram is smooth enough to exhibit the basic shape of the apparent underlying probability density function $f(x)$.

Heaping at convenient values can occur in the most rigorous of studies. For example, in Anderson and Pospahala (1970), perpendicular distances were measured with a 4-foot (1.2-m) stick, marked off in 1-foot (0.3-m) intervals and distances were interpolated to the near-

est inch. For $w^* = 8$ feet (2.4 m), there are 97 possible values of perpendicular distance $\{0, 1, \dots, 96\}$ inches (2.4 m), but when the ungrouped data were examined in detail, it was surprising to find obvious heaping at multiples of 12 inches (0.3 m). In fact, 24 percent of all recorded distances were recorded as exactly on some multiple of 12 inches. Yet, there are only 9 such values $\{0, 12, \dots, 96\}$. Hence, less than 10 percent of the possible distances occurred 24 percent of the time. Such heaping at 1-foot intervals caused no problems in analysis, especially after a suitable smoothing of the data.

Another type of problem with field data occurs when w is fixed (finite) and quite a few observations are made at distances near w . Then, there may be a tendency to record any distance near w units as being exactly w . For example, a recent line transect survey of nongame birds used $w = 100$ feet (30.48 m). Many species were visible well beyond 100 feet, and as a result there were many potential observations near and beyond that distance. In the sample of size 651 (data pooled over the summer's sampling), 71 values were recorded as exactly 100. From the histogram of the data, it was evident that it represented heaping and many of the observations probably corresponded to distances between 85 and 100 feet (25.9–30.48 m) (and probably many of the true distances were really more than 100 feet).

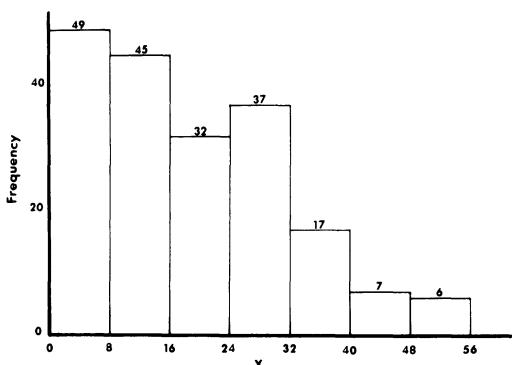


FIG. 25. Data of Table 1 grouped into 7 equal length intervals.

Chi-square Goodness of Fit Tests

A recommended procedure is to test the assumptions one makes. The model used for perpendicular distances, $f(x)$, represents an assumption, and as such it is logical to want to test the goodness of fit of the model (after estimation of any parameters). The only omnibus test available is the chi-square goodness of fit test based on grouping the data. The purpose of this section is to discuss the role of such a test and its limitations for line transect data. Indeed, the chi-square test

has little value when the Fourier series estimator of $f(0)$ is used.

The proper role of the goodness of fit test is to determine when a model fails to give an adequate fit to the data. Then, one should feel justified in rejecting that model, provided the rejection is not a result of anomalies in recording data. However, the data must be grouped before the chi-square test can be made. If the data are initially ungrouped, the choice of the number of groups and the exact cut points that define the intervals is somewhat arbitrary. Depending upon the selections made, one can often get the test to either fail or not fail (i.e., indicate the model does not fit, or does fit). Such arbitrariness renders the chi-square goodness of fit test in the case of line transects less useful for model selection than it might have seemed initially.

A second problem is that of heaping and inaccuracies, especially near $x = 0$, the critical area for model fit. Both phenomena tend to cause erroneous (misleading) rejection of model fit. For example, a 1-term Fourier series model was fit to the data of Fig. 24. Here, there are 14 cells, which is too fine a grouping to smooth out the anomalies arising due to field problems. The nature of this model is explained in PART 2 (*Fourier Series Estimator*); for the time being all that is relevant is that it is a tenable model for analysis of line transect data. In this example, the model has 1 parameter denoted by (a) and is

$$f(x) = \frac{1}{w^*} + a \cos\left(\frac{\pi x}{w^*}\right), \quad 0 < x < w^*.$$

After fitting this 1 parameter model to the data of Fig. 24, $\hat{a} = 0.0148 \pm 0.0012$ (i.e., the standard error of \hat{a} is 0.0012) and $\hat{f}(0) = 0.0327 \pm 0.0012$. The data, expected values, and chi-square contribution for each cell are shown in Table 2. The chi-square value for each cell is $(\text{Observed} - \text{Expected})^2/\text{Expected}$. The value of the overall chi-square test is the sum of all of those. The degrees of freedom for the example are 12. In general,

TABLE 2.—RESULTS OF THE CHI-SQUARE GOODNESS OF FIT TEST FOR THE ONE-TERM FOURIER SERIES MODEL FIT TO THE DATA GIVEN IN FIG. 24 (i.e., $w = 56$, $k = 14$, CUT POINTS EVENLY SPACED)

Cell	Cell boundaries	Observed values, n_i	Expected values $E(n_i)$	Chi-square value
i = 1	0–4	37	25.2	5.583
2	4–8	12	24.6	6.439
3	8–12	28	23.5	0.875
4	12–16	17	21.9	1.086
5	16–20	22	19.9	0.228
6	20–24	10	17.6	3.257
7	24–28	13	15.1	0.283
8	28–32	24	12.5	10.566
9	32–36	11	10.0	0.098
10	36–40	6	7.7	0.376
11	40–44	2	5.7	2.401
12	44–48	5	4.1	0.197
13	48–52	4	3.0	0.341
14	52–56	2	2.4	0.073
Total chi-square value =				31.801

they are $k - 1 - m$, where k = number of cells, and m = number of parameters estimated.

In the example, the overall chi-square value is 31.801 and therefore, we reject the model as providing an adequate fit ($P < 0.01$). Moreover, by examination of each individual value, the fit would be judged to be bad for cells 1, 2, 6, and 8. All of this must be judged after recognizing that the data are subject to severe problems of heaping, especially at $x = 0$, 10, 20, 30. When they occur, problems of heaping and inaccuracies near $x = 0$ (in particular), render the goodness of fit test useless for testing model fit.

Another aspect of the problems with the goodness of fit test is the arbitrariness of the data grouping. We recommend no more than 10 cells, but choice of the exact cut points can influence the results. From the standpoint of ease of analysis, it is best to have cut points evenly spaced. From the same data set as in the above example but using only 7 cells, the overall result (Table 3, Fig. 25) is a chi-square value of 4.922 (5 degrees of freedom) and the model is judged to fit the data. In that case, $\hat{a} = 0.150 \pm 0.00121$ and $\hat{f}(0) = 0.0328 \pm 0.0012$, neither of which is dif-

TABLE 3.—RESULTS OF THE CHI-SQUARE GOODNESS OF FIT TEST FOR THE 1-TERM FOURIER SERIES MODEL FIT TO THE DATA IN FIG. 25 (i.e., $w = 56$, $k = 7$, CUT POINTS EVENLY SPACED)

Cell	Cell boundaries	Observed values, n_i	Expected values, $E(n_i)$	Chi-square value
1	0–8	49	49.9	0.016
2	8–16	45	45.5	0.005
3	16–24	32	37.5	0.809
4	24–32	37	27.6	3.224
5	32–40	17	17.6	0.023
6	40–48	7	9.7	0.734
7	48–56	6	5.2	0.110
Total chi-square value = 4.922				

ferent from the estimates obtained with data based on 14 cells.

The difference between the example with 7 cells versus 14 cells is that grouping smoothed the data enough to get a better judgment about the adequacy of the 1-term Fourier series model. Unfortunately, the power of such a goodness of fit test based on 7 cells is very low and many other models would also fit the grouped data (Fig. 25). We have seen real data with realistic sample sizes of 40–60 for which almost any of the models we examined during our research would fit, as judged by a chi-square goodness of fit test based on 5–8 cells (which is about what one must use to smooth out the data problems of heaping and rounding near $x = 0$). Consequently, it is concluded that the chi-square goodness of fit test is of little value in model selection regardless of what model (estimator) is used. Instead, other criteria must be used for model selection, such as *model robustness*, *pooling robustness*, the *shape criterion*, and *estimator efficiency*. We recommend computing the model goodness of fit test, but being cautious in rejecting models on the basis of the results, and we suggest extreme caution in recommending models just because they happen to fit the data as judged by that test.

Those considerations apply regardless of the model assumed for $f(x)$. An additional consideration is worth mentioning when the Fourier series model is used.

The Fourier series estimation method involves selecting the number of terms (m) to be used to estimate $f(0)$. This selection is achieved using a stopping rule (section entitled Stopping Rule for the Selection of m , PART 2) which itself tends to achieve as good a fit to the data as possible. Thus, given adequately smoothed (grouped) data, the Fourier series estimator can be expected to fit the data quite well.

Standard Errors and Confidence Intervals

Introduction

In the general estimation formula $\hat{D} = \hat{n}(0)/2L$, both n and $\hat{f}(0)$ are subject to sampling variation; therefore, the properties of \hat{D} depend upon the sampling properties of both n and $\hat{f}(0)$. Within limits, ways to estimate the sampling variance of \hat{D} , and to construct confidence intervals on D without reference to the nature of the specific estimator of $f(0)$ can be discussed. The basic equation for estimating $\text{var}(\hat{D})$ derives from the theoretical results in the MATHEMATICAL BACKGROUND section. Specifically, based on Eq. (1.7) the general result is

$$\text{var}(\hat{D}) = (\hat{D})^2[(\hat{cv}(n))^2 + (\hat{cv}(\hat{f}(0)))^2], \quad (1.17)$$

where cv stands for a coefficient of variation:

$$(\hat{cv}(n))^2 = \frac{\hat{\text{var}}(n)}{(n)^2}, \quad (1.18)$$

$$(\hat{cv}(\hat{f}(0)))^2 = \frac{\hat{\text{var}}(\hat{f}(0))}{(\hat{f}(0))^2}, \quad (1.19)$$

$$\begin{aligned} (\hat{cv}(\hat{D}))^2 &= \frac{\hat{\text{var}}(\hat{D})}{(\hat{D})^2} \\ &= [(\hat{cv}(n))^2 + (\hat{cv}(\hat{f}(0)))^2]. \end{aligned} \quad (1.20)$$

Finally, note that the standard errors of \hat{D} , n , and $\hat{f}(0)$ are just the square roots of their respective sampling variances, e.g., $\hat{s}_e(\hat{D}) = (\text{var}(\hat{D}))^{1/2}$.

Estimation of $\text{var}(\hat{D})$ using Eq. (1.17) requires that the sampling variance of

both $\hat{f}(0)$ and n also be estimated. Because $\hat{f}(0)$ derives from the sample of perpendicular distance (or sighting distance and angle) data, its sampling variance is conditional on n . However, because of the probabilistic nature of the detection process, n itself is a random variable and thus has a sampling variance that depends in part on the unknown spatial distribution of objects in the study area. The contribution of those 2 sources of variation to $\hat{\text{var}}(\hat{D})$ are embodied in Eqs. (1.18) and (1.19) which comprise the 2 components of $\hat{\text{cv}}(\hat{D})$ in Eq. (1.20). All methods recommended for obtaining $\hat{f}(0)$ include equations (or methods) for obtaining $\hat{\text{var}}(\hat{f}(0))$. Estimating $\hat{\text{var}}(n)$ is more difficult; in fact, the heart of the problem of obtaining $\hat{\text{var}}(\hat{D})$ from Eq. (1.17) is obtaining an estimate for $\text{var}(n)$.

Alternatively, one can bypass Eq. (1.17) and construct the sampling variance of \hat{D} directly from the replicate lines. That and other approaches are given below.

In our experience, the 2 separate coefficients of variation are often of about equal magnitude, and thus neither dominates the sampling variance of \hat{D} . In particular, we note that it is $\hat{\text{var}}(n)$, hence $\hat{\text{cv}}(n)$, that cannot be found as easily as $\hat{\text{var}}(\hat{f}(0))$; therefore, if $\hat{\text{cv}}(n)$ were of a smaller order of magnitude than $\hat{\text{cv}}(\hat{f}(0))$, obtaining $\hat{\text{var}}(\hat{D})$ would be simplified. If such a situation occurred, one could ignore the contribution due to $\hat{\text{cv}}(n)$.

Constructing confidence intervals on D is recommended based either on the approximation of a normal sampling distribution for \hat{D} or on the t -distribution, depending on the manner in which $\hat{\text{var}}(\hat{D})$ is obtained. For the normal distribution, $\hat{D} \pm 1.96 \hat{s}\text{e}(\hat{D})$ is used to obtain an approximate 95 percent confidence interval on D (for a 90% interval use $\hat{D} \pm 1.645\hat{s}\text{e}(\hat{D})$, however, we recommend 95% confidence intervals).

Under some direct methods of estimating $\hat{\text{var}}(\hat{D})$ from replicate lines, it is more appropriate to assume \hat{D} has a Student's t -distribution (any elementary statistics text deals with that distribution, see e.g.,

Steele and Torrie 1960:43). In that case, the 95 percent confidence intervals are $\hat{D} \pm t_{R-1} \hat{s}\text{e}(\hat{D})$, where t_{R-1} is the tabulated value of the t -distribution, with $R-1$ degrees of freedom, needed to produce a valid 95 percent confidence interval (e.g., for 10 degrees of freedom $t_{10} = 2.228$, while for greater than 30 degrees of freedom the use of 1.96 for t is acceptable). This confidence interval computation is illustrated under various circumstances under ILLUSTRATIVE EXAMPLES.

Direct Estimation of $\text{var}(\hat{D})$

Let there be R replicate lines, with corresponding lengths, ℓ_i and sample sizes n_i . If the individual sample sizes are large enough, an independent estimate of density can be computed for each line:

$$\hat{D}_i = \frac{n_i \hat{f}_i(0)}{2\ell_i}.$$

An overall, weighted estimate of density should then be computed as

$$\hat{D} = \frac{\sum_{i=1}^R \ell_i \hat{D}_i}{\sum_{i=1}^R \ell_i} = \frac{\sum_{i=1}^R \ell_i \hat{D}_i}{L}. \quad (1.21)$$

(Note, that the equations are not necessarily appropriate for those types of stratified sampling schemes, in which strata are subareas of the total study area. In such designs the equations given here should be applied separately to each stratum.)

The corresponding empirical estimator of $\text{var}(\hat{D})$, based on replicate samples is

$$\hat{\text{var}}(\hat{D}) = \frac{\sum_{i=1}^R \ell_i (\hat{D}_i - \hat{D})^2}{L(R-1)}. \quad (1.22)$$

Confidence intervals in that instance are constructed based on the t -distribution. Thus the 95 percent confidence interval on true density D is $\hat{D} \pm t_{R-1} \hat{s}\text{e}(\hat{D})$.

That approach to estimating $\text{var}(\hat{D})$ is recommended; however, it has a serious limitation. If some or all of the sample sizes n_i are too small (say less than 25),

TABLE 4.—HYPOTHETICAL EXAMPLE TO ILLUSTRATE THE JACKKNIFE METHOD OF DIRECT ESTIMATION OF $\text{var}(\hat{D})$; $R = 12$, $L = 28$, $n = 135$, $\hat{D} = 100$

i	n_i	ℓ_i	$n - i$	$L - \ell_i$	$\hat{D}_{(i)}$	$\hat{D}^{(i)}$
1	5	1	130	27	99	127.0
2	9	2	126	26	102	74.0
3	17	3	118	25	103	75.0
4	3	1	132	27	98	154.0
5	23	5	112	23	96	118.4
6	20	4	115	24	102	88.0
7	12	2	123	26	100	100.0
8	6	1	129	27	101	73.0
9	11	3	124	25	95	141.7
10	12	2	123	26	97	139.0
11	2	1	133	27	101	73.0
12	15	3	120	25	100	100.0

then reliable estimators of $f_i(0)$ may not exist. In fact, they may be virtually uncomputable for very small n_i . When numerous short lines are used, it is to be expected that some n_i will even be zero. Although it is defensible to use $\hat{D}_i = 0$ if $n_i = 0$, it is not very satisfying. Moreover, if $n_i = 1$ should occur, it is not at all clear what to use for \hat{D}_i . Direct estimation of $\text{var}(\hat{D})$ (and weighted estimation of \hat{D}) in such cases is still possible, however, by using the “jackknife” technique.

Jackknife Estimation of $\text{var}(\hat{D})$

To illustrate the material of this section a hypothetical line transect study with 12 randomly located lines is presented first. The 12 sample sizes and the corresponding line lengths in columns 2 and 3, respectively, are presented in Table 4. The additional columns are explained, and referred to below. Some of the sample sizes are too small (e.g., $n_{11} = 2$) to allow any line transect estimator to be applied. Even sample sizes of 10 or 20 are not adequate for most procedures. The total set of distance data ($n = 135$) could be pooled to estimate $f(0)$, but then one could not take advantage of the replicate lines to compute replicate estimators \hat{D}_i and directly get an estimator of $\text{var}(\hat{D})$. The jackknife method provides a means of pooling data to get large enough sam-

ple sizes and still take direct advantage of the replicate lines. (For detailed information on the jackknife method see Gray and Schucany 1972, or Miller 1974.)

The mechanics of the jackknife are illustrated with the example in Table 4. One must first compute 12 estimates of density (and hence of $f(0)$), but not from the 12 separate lines. Instead, one must sequentially pool all data *except* 1 line, and compute the estimate of density from those 11 pooled lines. Deleting line 1 and pooling all the distance data from lines 2 through 12 gives a sample of size 130 ($135 - 5$). From those data, one computes an estimator of density denoted $\hat{D}_{(1)}$. Then returning to the entire sample, drop the data from just line 2 and compute $\hat{D}_{(2)}$ based on 126 ($135 - 9$) distances remaining. Note that the corresponding line lengths for $\hat{D}_{(1)}$ and $\hat{D}_{(2)}$ are 27 (= 28 - 1) and 26 (= 28 - 2), respectively. In general, one drops the i^{th} line and computes $f(0)$ from the distances based on a sample of size $n - n_i$ with a total line length of $L - \ell_i$; $\hat{D}_{(i)}$ denotes the resultant estimator. It is also necessary to compute the overall estimate \hat{D} on the basis of all the pooled data. Hence, sample size n and length L correspond to \hat{D} (no data are deleted).

Given \hat{D} , $\hat{D}_{(1)}, \dots, \hat{D}_{(R)}$ we next compute what are called “pseudovalue” in the jackknife method. For R replicate lines there will be R pseudovalue; the i^{th} one is computed as

$$\hat{D}^{(i)} = \frac{L\hat{D} - (L - \ell_i)\hat{D}_{(i)}}{\ell_i}, \quad i = 1, \dots, R.$$

For example,

$$\hat{D}^{(1)} = \frac{L\hat{D} - (L - \ell_1)\hat{D}_{(1)}}{\ell_1}.$$

The pseudovalue $\hat{D}^{(1)}, \dots, \hat{D}^{(R)}$ are now treated like R replicate estimators of density, and Eqs. (1.21) and (1.22) are applied to compute \hat{D}_j and $\text{var}(\hat{D}_j)$, respectively:

$$\hat{D}_j = \frac{\sum_{i=1}^R \ell_i \hat{D}^{(i)}}{L}$$

$$\hat{\text{var}}(\hat{D}_j) = \frac{\sum_{i=1}^R \ell_i (\hat{D}^{(i)} - \hat{D}_j)^2}{L(R-1)}.$$

This procedure works especially well when the underlying estimator of $f(0)$ is pooling robust (i.e., based on a linear model such as the Fourier series). In particular, it is noted that \hat{D} and \hat{D}_j will differ very little; the procedure is not meant to improve upon the estimate of D , but rather to provide an estimator of $\text{var}(\hat{D})$. When the jackknife procedure is used, we do recommend taking the estimate of D as \hat{D}_j and, of course, its sampling variance is $\hat{\text{var}}(\hat{D}_j)$.

The (hypothetical) results of the jackknife method applied to the 12 replicate lines are shown in Table 4. Applying Eq. (1.21) to the $\hat{D}^{(i)}$ of Table 4 yields $\hat{D}_j = 105.3$, which is not much different from $\hat{D} = 100$. The advantage of the method is that Eq. (1.22) can be applied to the $\hat{D}^{(i)}$ to yield $\hat{\text{var}}(\hat{D}_j) = 58.63$, and this $\hat{s}e(\hat{D}_j) = 7.66$. The approximate 95 percent confidence interval computed based on the Student's t -distribution with 11 degrees of freedom, is $105.3 \pm (2.201)7.66$, or 88.4 to 122.2.

Indirect Estimation of $\text{var}(\hat{D})$

The empirical estimator of Eq. (1.22) is the most assumption free estimator of $\text{var}(\hat{D})$. However, it is often impractical to use Eq. (1.22) directly and it may not always be convenient to use the jackknife approach because it requires a lot of computation. Here, the alternative is presented of obtaining $\hat{\text{var}}(\hat{D})$ indirectly by estimating separately the sampling variances of $f(0)$ and n . Only $\hat{\text{var}}(n)$ presents problems; it depends on the unknown spatial distribution of objects in the study area. One must select one of the several ways to estimate $\text{var}(n)$ and can then use Eq. (1.17) to compute $\hat{\text{var}}(\hat{D})$.

First, an empirical estimate of $\text{var}(n)$ can be computed if one has replicate lines. The appropriate estimator must account for varying line lengths:

$$\hat{\text{var}}(n) = L \frac{\sum_{i=1}^R \ell_i \left[\frac{n_i}{\ell_i} - \frac{n}{L} \right]^2}{R-1}. \quad (1.23)$$

For reliable results one should have at least $R = 5$; and $R \geq 10$ is much better. In the special case of equal replicate line lengths $\ell_i \equiv \ell$, (1.23) becomes

$$\hat{\text{var}}(n) = R \frac{\sum_{i=1}^R (n_i - \bar{n})^2}{R-1}, \quad (1.24)$$

where $\bar{n} = (\sum n_i)/R$.

As an example, if Eq. (1.23) is applied to the data of Table 4, $R = 12$, $n = 135$, and

$$\hat{\text{var}}(n) = 28 \frac{25.0738}{11} = 63.8242.$$

It then follows that $\hat{s}e(n) = (63.8242)^{1/2} = 7.9890$ and $\hat{cv}(n) = 7.9890/135 = 0.05917$. If from the pooled data, one had $f(0) = 41.5$ and $\hat{s}e(f(0)) = 2.49$, the density estimate would be $\hat{D} = (135)(41.5)/2 \times 28 = 100$ and from Eq. (1.17)

$$\begin{aligned} \hat{\text{var}}(\hat{D}) &= (100)^2[(0.05917)^2 + (0.06)^2] \\ &= (100)^2[0.0035 + 0.0036] \\ &= 71.0. \end{aligned}$$

The standard error of \hat{D} is $\hat{s}e(\hat{D}) = 8.42$ and the coefficient of variation of \hat{D} is $\hat{cv}(\hat{D}) = 0.0842$ or 8.42 percent. The approximate 95 percent confidence interval in this example is computed as $\hat{D} \pm 2.201\hat{s}e(\hat{D})$, where 2.201 is from the t -distribution with 11 degrees of freedom.

The empirical approach given above is the best method for estimating $\text{var}(n)$ because no assumptions are necessary. However, sometimes, replication is insufficient for such an approach (for example, 1 line run only once). Then one way of proceeding is to assume that $\hat{\text{var}}(n) = a$ for some constant a . Typically, one may find $1 < a < 3$ (cf. Eberhardt 1978b). The value of a will depend upon the unknown spatial distribution of objects in the study area. If objects have a random (Poisson) distribution, then $a = 1$, i.e., $\hat{\text{var}}(n) = n$, and $(\hat{cv}(n))^2 = 1/n$.

Generally, one can expect some degree

of aggregation in the distribution of objects, which usually implies that a is greater than 1. If a compromise value is needed, $a = 2.0$ may be useful. Clearly, it is not recommended to guess at a (i.e., guessing at $\hat{\text{var}}(n)$), but those guidelines can be used if no other approach is possible.

Transect Segments as Replicates

Equations (1.22) and (1.23) are valid only when the ℓ_i derive from true replicate lines; i.e., there are either physically separate lines or the lines are run at different times. The assumption behind this type of replication is that the data from each line are statistically independent samples. A type of replication that violates that independence assumption is the use of different segments of the same line as the ℓ_1, \dots, ℓ_R . It can be useful to divide the lines into segments and record data separately by segment. Proper use of that type of replication, however, must recognize that the sample sizes from adjacent segments are likely to be correlated.

If data are recorded separately by tran-

sect segment, it is very doubtful that sample sizes would be large enough to estimate D_i for each segment. Rather, that information would be useful only for estimating $\text{var}(n)$, but not by using Eq. (1.23). Let there be 1 line with R equal length segments ($R > 3$). An equation adapted from Seber (1973:6–7) to estimate $\text{var}(n)$ is

$$\hat{\text{var}}(n) = \frac{3(S_1)^2 - (S_2)^2}{R - 3},$$

where

$$(S_1)^2 = \sum_{i=1}^R (n_i - \bar{n})^2$$

$$(S_2)^2 = \sum_{i=1}^{R-1} (n_{i+1} - n_i)^2.$$

Corresponding equations can be developed if the line segments are of unequal length. Also there may be several different lines, each one supplying considerable information about $\text{var}(n)$ through within line segments. Although no equations are presented for dealing with such situations, we note that such equations can be developed.

PART 2

ROBUST ESTIMATION METHODS

Material in this part concerns the application of the recommended estimation procedures for perpendicular distance data (grouped and ungrouped) and for sighting distance and angle data. PART 2 is directed to the biologist and presents many figures and examples to illustrate the concepts and calculations. In addition, sections cover estimation for mobile populations and grouped data. A final section discusses the use of a comprehensive computer program TRANSECT that should be helpful in the analysis of the line transect data.

The methods described appear to have good properties in comparison with other methods in the literature or with other methods that we developed and examined (cf. Part 4). In PART 2, we stress the application of methods that we believe have good properties. The mathematical theory underlying these estimation methods and computer simulation results on their small sample properties are covered in PART 3.

ANALYSIS AND INFERENCE PROCEDURES FOR PERPENDICULAR DISTANCES, UNGROUPED DATA

Fourier Series Estimator

The estimation procedure described in this section is based on a Fourier series (FS) expansion of the probability density function (pdf), $f(x)$. A Fourier (pronounced Fouréay) series is an expression of sines and cosines and was derived in the 18th century by the French mathematical physicist, Jean Baptiste Joseph Fourier (1768–1830). The Fourier series is one of a number of mathematical expansions of a function. Such series expansions are exact for infinitely many terms. By choosing one such expansion and limiting the number of terms, one arrives at a general model for $f(x)$. An objective rule must then be derived to se-

lect the appropriate number of terms for a given set of data. The interest is in an approximation with only m terms, where m is a small value, say 6 or less. In general,

$$f(x) \doteq \frac{1}{w^*} + \sum_{k=1}^m \left[a_k \cos\left(\frac{k\pi x}{w^*}\right) + b_k \sin\left(\frac{k\pi x}{w^*}\right) \right].$$

Because one is concerned with the estimation of $f(0)$ and $\sin(0) = 0$, the sine terms and the coefficients b_k are not used in the context of line transect analysis.

The Fourier series (FS) is, of course, periodic and we are concerned with expansions from 0 to w^* , where $w^* \leq w$. In this section, only the analysis of ungrouped perpendicular distance data x_i are treated. (Refer to APPENDIX A for a list of symbols and definitions.)

The FS estimator is a nonparametric procedure that is *model robust* and *pooling robust* and meets the *shape criterion* $f(0) = 0$. The method is easy to compute and its estimation efficiency for small samples is quite good.

Estimation Equations

The estimation of density from ungrouped data on perpendicular distances is straightforward if the FS approach is used. The procedure employs 7 simple equations. The estimator of density follows from the general theory discussed in PART 1

$$\hat{D} = \frac{n\hat{f}(0)}{2L}. \quad (2.1)$$

The estimator of the pdf at 0 distance for the FS method is

$$\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^m \hat{a}_k. \quad (2.2)$$

Finally, an expression for an estimator of the coefficients a_k is required

$$\hat{a}_k = \frac{2}{nw^*} \left[\sum_{i=1}^n \cos\left(\frac{k\pi x_i}{w^*}\right) \right] \quad k = 1, 2, 3, \dots \quad (2.3)$$

It is often convenient to denote $2/w^*$ as a_0 .

Four additional expressions are required to estimate sampling variances and covariances. The estimator of the sampling variance of the FS coefficients is

$$\hat{\text{var}}(\hat{a}_k) = \frac{1}{(n-1)} \left[\frac{1}{w^*} \left(\hat{a}_{2k} + \frac{2}{w^*} \right) - \hat{a}_k^2 \right] \quad k \geq 1. \quad (2.4)$$

The sampling covariances of those coefficients are estimated from the following expression

$$\hat{\text{cov}}(\hat{a}_k, \hat{a}_j) = \frac{1}{(n-1)} \left[\frac{1}{w^*} (\hat{a}_{k+j} + \hat{a}_{k-j}) - (\hat{a}_k \hat{a}_j) \right] \quad k > j > 1. \quad (2.5)$$

It can be seen from equation (2.2) that $\hat{f}(0)$ is merely the sum of the \hat{a}_k coefficients (m of them) plus a constant $1/w^*$. Therefore, one would expect the sampling variance of $\hat{f}(0)$ to be the sum of all the sampling variances and covariances of the coefficients \hat{a}_k (i.e., equations 2.4 and 2.5). Realizing that $\hat{\text{var}}(\hat{a}_k)$ can be written as $\hat{\text{cov}}(\hat{a}_k, \hat{a}_k)$, an expression for the sampling variance of $\hat{f}(0)$ is

$$\hat{\text{var}}(\hat{f}(0)) = \sum_{j=1}^m \sum_{k=1}^m \hat{\text{cov}}(\hat{a}_j, \hat{a}_k). \quad (2.6)$$

That expression may at first appear formidable, but it is merely the sum of all the elements in the m by m matrix of sampling variances and covariances (from Eq. 2.4 and 2.5). For example, letting $m = 3$, the following 3 by 3 matrix would be needed:

$$\begin{bmatrix} \hat{\text{var}}(\hat{a}_1) & \hat{\text{cov}}(\hat{a}_1, \hat{a}_2) & \hat{\text{cov}}(\hat{a}_1, \hat{a}_3) \\ \hat{\text{cov}}(\hat{a}_2, \hat{a}_1) & \hat{\text{var}}(\hat{a}_2) & \hat{\text{cov}}(\hat{a}_2, \hat{a}_3) \\ \hat{\text{cov}}(\hat{a}_3, \hat{a}_1) & \hat{\text{cov}}(\hat{a}_3, \hat{a}_2) & \hat{\text{var}}(\hat{a}_3) \end{bmatrix}$$

Again, note that $\hat{\text{var}}(\hat{a}_k)$ could be expressed as $\hat{\text{cov}}(\hat{a}_k, \hat{a}_k)$. The $\hat{\text{var}}(\hat{f}(0))$ is merely the sum of all 9 elements in the

variance-covariance matrix. Approaches to estimation of $\text{var}(\hat{D})$ and $\text{var}(n)$ are given in PART 1. The use of those equations is illustrated in the examples below; however, first one requires an objective way to choose m , the number of cosine terms in the Fourier series.

Stopping Rule for the Selection of m

The bias in the estimators $\hat{f}(0)$ and \hat{D} can be made arbitrarily small by increasing m , the number of terms in the Fourier series. However, the sampling variance of $\hat{f}(0)$ and \hat{D} increases with increasing values of m . This suggests a trade-off between small bias and large variance. The following stopping rule achieves such a trade-off. Choose the first value of m such that

$$\frac{1}{w^*} \left(\frac{2}{n+1} \right)^{1/2} \geq |\hat{a}_{m+1}| \quad (2.7)$$

where

$|\hat{a}_{m+1}|$ is the absolute value of \hat{a}_{m+1} .

The theory for this stopping rule is given in PART 3. That rule appears to be fairly good, and our studies (both theoretical and simulation) provide evidence to suggest that m should not be chosen larger than about 6.

Examples

The use of the FS approach to estimating density is illustrated below with 3 examples. The first 2 are “hand worked” to show the calculations. They were also chosen because the true density is known and, therefore, the bias can be examined. The final example is illustrated by output from the computer program TRANSECT.

Simulated Data

A set of data from our computer simulation studies was chosen to illustrate the use of the Fourier series. Perpendicular distance data x_i were simulated from

TABLE 5.—EXAMPLE OF SIMULATED LINE TRANSECT DATA, AND OTHER STATISTICS, FROM AN UNDERLYING HALF-NORMAL DISTRIBUTION: $n = 40$, $w = 65$ m, $\sigma = 33.334$, $f(0) = 0.0252$, $L = 1,000$, AND $D = 0.000504$. THE PERPENDICULAR DISTANCES x_i ARE IN METERS

i	x_i	k = 1		k = 2		k = 3	
		$k\pi x_i/w^1$	$\cos(k\pi x_i/w)$	$k\pi x_i/w^1$	$\cos(k\pi x_i/w)$	$k\pi x_i/w^1$	$\cos(k\pi x_i/w)$
1	24.10	1.165	0.395	2.330	-0.688	3.495	-0.938
2	32.92	1.591	-0.020	3.182	-0.999	4.773	0.061
3	8.53	0.412	0.916	0.824	0.679	1.236	0.329
4	36.76	1.777	-0.204	3.554	-0.916	5.330	0.579
5	2.14	0.103	0.995	0.206	0.979	0.309	0.953
6	37.61	1.818	-0.244	3.636	0.880	5.454	0.675
7	0.00	0.000	0.000	0.000	0.000	0.000	1.000
8	43.16	2.086	-0.493	4.172	-0.514	6.258	1.000
9	10.98	0.531	0.862	1.062	0.487	1.593	-0.022
10	43.20	2.088	-0.494	4.176	-0.511	6.264	1.000
11	27.26	1.317	0.251	2.634	-0.874	3.951	-0.690
12	14.27	0.690	0.771	1.380	0.190	2.070	-0.479
13	61.88	2.991	-0.989	5.982	0.955	8.973	-0.900
14	36.33	1.756	-0.184	3.512	-0.932	5.268	0.527
15	5.99	0.290	0.958	0.580	0.836	0.870	0.645
16	16.28	0.787	0.706	1.574	-0.003	2.361	-0.710
17	8.30	0.401	0.921	0.802	0.695	1.203	0.360
18	27.61	1.334	0.234	2.668	-0.890	4.002	-0.652
19	25.25	1.220	0.343	2.440	-0.764	3.660	-0.869
20	16.25	0.785	0.707	1.570	0.001	2.355	-0.706
21	59.69	2.885	-0.967	5.770	0.871	8.655	-0.718
22	25.31	1.223	0.341	2.446	-0.768	3.669	-0.864
23	3.05	0.147	0.989	0.294	0.957	0.441	0.904
24	9.06	0.438	0.906	0.876	0.640	1.314	0.254
25	48.56	2.347	-0.701	4.694	-0.018	7.041	0.726
26	32.68	1.579	-0.009	3.158	-1.000	4.737	0.025
27	23.58	1.140	0.418	2.280	-0.651	3.420	-0.961
28	24.78	1.120	0.365	2.240	-0.620	3.360	-0.976
29	44.42	2.147	-0.545	4.294	-0.406	6.441	0.988
30	6.94	0.335	0.944	0.670	0.784	1.065	0.484
31	13.18	0.637	0.804	1.274	0.292	1.911	-0.334
32	41.01	1.982	-0.400	3.964	-0.680	5.946	0.944
33	48.05	2.322	-0.683	4.644	-0.068	6.966	0.776
34	17.48	0.845	0.664	1.690	-0.119	2.535	-0.822
35	14.99	0.724	0.749	1.448	0.122	2.172	-0.566
36	37.84	1.829	-0.255	3.658	-0.870	5.487	0.699
37	5.32	0.257	0.967	0.514	0.871	0.771	0.717
38	27.73	1.340	0.229	2.680	-0.895	4.020	-0.638
39	23.71	1.146	0.412	2.292	-0.660	3.438	-0.956
n = 40	7.95	0.384	0.927	0.768	0.719	1.152	0.407
			10.586		-2.888		1.332

¹ Values in radians, not degrees.

an underlying truncated half-normal density function

$$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)}{\sqrt{2\pi} \sigma \left(\phi\left(\frac{w}{\sigma}\right) - \frac{1}{2}\right)}, \quad 0 \leq x \leq w, \quad 0 < \sigma$$

where $\phi(\cdot)$ is the standard normal cumulative distribution function (cdf).

All the perpendicular distance data x_i were assumed to represent distances in meters. We took the transect width to be 65 m (=w) and generated 40 distances using $\sigma = 33.334$. The true density is $D = 0.000504$ (5.04 objects/ha) and $f(0) =$

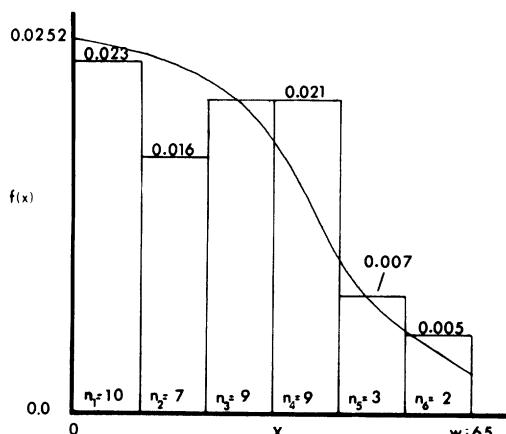


FIG. 26. Truncated half-normal pdf used to illustrate the Fourier series estimator, $\sigma = 33.334$, $w = 65$ m. Forty perpendicular distances were generated from this distribution. A histogram of the observed frequencies and relative frequencies ($n_i/(\Delta x)$) of the data is also shown. Data are grouped into class intervals of 10.83 m.

0.0252. The simulated data are shown in Table 5, with the simple statistics (for $k = 1, 2, 3$) needed to compute estimates of the FS coefficients \hat{a}_k . A sample histogram and the underlying pdf are shown in Fig. 26, and Fig. 27 shows the underlying and the estimated pdf.

Thus, $n = 40$ and $w = w^* = 65$ m (because the data were collected [i.e., simulated] with a fixed width transect, it is

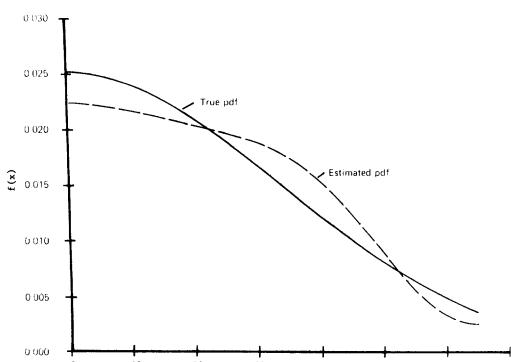


FIG. 27. Underlying half-normal density ($\sigma = 33.334$, $w = 65$, and $f(0) = 0.0252$) with the Fourier series estimator of $f(x)$ for a sample size $n = 40$.

convenient to let $w^* = w$) and we assumed $L = 1,000$ m.

The coefficients a_k are computed by using Eq. 2.3

$$\hat{a}_k = \frac{2}{nw^*} \left[\sum_{i=1}^n \cos\left(\frac{k\pi x_i}{w^*}\right) \right] \quad k = 1, 2, \dots$$

The terms $\sum_{i=1}^{40} \cos\left(\frac{k\pi x_i}{w^*}\right)$ for $k = 1, 2$, and 3 are 10.586, -2.888, and 1.332, respectively (Table 5). Then,

$$\hat{a}_1 = \frac{2}{40 \times 65} [10.586] = 0.008,$$

$$\hat{a}_2 = \frac{2}{40 \times 65} [-2.888] = -0.002, \text{ and}$$

$$\hat{a}_3 = \frac{2}{40 \times 65} [1.332] = 0.001.$$

The example above illustrates the basic calculations; substantial rounding error is introduced by doing the computations by hand (or on a small desk calculator) and in recording only a few significant digits (i.e., Table 5). The values we use for the first six \hat{a}_k 's were calculated on a computer, using program TRANSECT, and are $\hat{a}_1 = 0.008912$, $\hat{a}_2 = -0.002896$, $\hat{a}_3 = 0.001063$, $\hat{a}_4 = 0.001046$, $\hat{a}_5 = -0.001446$, and $\hat{a}_6 = 0.001480$.

Because the stopping rule requires knowledge of only w^* , n , and the \hat{a}_k 's, one can determine the number of terms required to estimate $f(0)$, and, subsequently, D. From Eq. (2.7)

$$\frac{1}{w^*} \left(\frac{2}{n+1} \right)^{1/2} \geq |\hat{a}_{m+1}|.$$

For the example, we want to choose a value of m such that

$$\frac{1}{65} \left(\frac{2}{41} \right)^{1/2} = 0.003398 \geq |\hat{a}_{m+1}|.$$

We see that this inequality is satisfied for $m = 1$ because $0.003398 > |\hat{a}_{1+1}| = |\hat{a}_2| = 0.002896$. The stopping rule suggests that only one term in the Fourier series is needed to estimate $f(0)$ in this example.

The estimate of $\hat{f}(0)$ is made by using Eq. (2.2),

$$\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^m \hat{a}_k,$$

and for this example,

$$\begin{aligned}\hat{f}(0) &= \frac{1}{65} + \hat{a}_1 \\ &= 0.015385 + 0.008912 \\ &= 0.0242966 \text{ or } 0.0243.\end{aligned}$$

Density can now be estimated by using Eq. (2.1)

$$\begin{aligned}\hat{D} &= \frac{n\hat{f}(0)}{2L} \\ &= \frac{40 \times 0.0243}{2 \times 1000} \\ &= 0.000486 \text{ or } 4.86 \text{ objects/ha.}\end{aligned}$$

This estimate compares fairly well with the true density of 5.04 objects/ha.

In this simple example, one needs only Eq. (2.4) to estimate the sampling variance of \hat{a}_k ,

$$\begin{aligned}\hat{\text{var}}(\hat{a}_k) &= \frac{1}{(n-1)} \left[\frac{1}{w^*} \left(\hat{a}_{2k} + \frac{2}{w^*} \right) \right. \\ &\quad \left. - \hat{a}_k^2 \right].\end{aligned}$$

In the example, only the first coefficient is used to estimate $f(0)$ and D , therefore, only the variance of \hat{a}_1 is required (note, however, that \hat{a}_2 will be needed to compute this variance). In this case one needs only

$$\begin{aligned}\hat{\text{var}}(\hat{a}_1) &= \frac{1}{39} \left[\frac{1}{65} \left(-0.002896 + \frac{2}{65} \right) \right. \\ &\quad \left. - 0.008912^2 \right] \\ &= 0.000008959.\end{aligned}$$

There are no covariances, therefore, Eq. (2.5) is not needed, because the variance-covariance matrix contains a single element, $\hat{\text{var}}(\hat{a}_1)$. Likewise, Eq. (2.6) is not needed, because the double sum of the variance-covariance matrix is merely $\hat{\text{var}}(\hat{a}_1)$ in this special case. Thus, from Eq. (2.6), the $\hat{\text{var}}(\hat{f}(0)) = \hat{\text{var}}(\hat{a}_1) = 0.000008959$. No actual area is involved

because the data were simulated on a digital computer. It will be assumed, for illustration, that n is Poisson (as if the 40 objects were distributed randomly); therefore, $\hat{\text{var}}(n) = n$. Then,

$$\begin{aligned}\hat{\text{var}}(\hat{D}) &= (\hat{D})^2 \left[\frac{\hat{\text{var}}(n)}{n^2} + \frac{\hat{\text{var}}(\hat{f}(0))}{\hat{f}(0)^2} \right] \\ &= (\hat{D})^2 \left[\frac{40}{1600} + \frac{0.000008959}{(0.0243)^2} \right] \\ &= (\hat{D})^2 [0.025 + 0.015157] \\ &= (0.000486)^2 [0.040157] \\ &= (0.00009739)^2.\end{aligned}$$

The $\hat{s}e(\hat{D}) = 0.00009739$ and $\hat{cv}(\hat{D}) = 0.20$ or a 20 percent coefficient of variation. The approximate 95 percent confidence interval based on $\hat{D} \pm 1.96 \hat{s}e(\hat{D})$ is (0.000295 to 0.000677).

To summarize, the FS estimate of density is 0.000486 ± 0.000191 . Converting to objects per ha, $\hat{D} = 4.86 \pm 1.91$. In that example, the method performed fairly well; the estimated percentage relative bias (PRB) = $100 \times (\hat{D} - D)/D = -3.6$.

As an extended example, now consider the case where $m = 3$. That case is included to illustrate better the estimation of the sampling variance of $\hat{f}(0)$ and \hat{D} . The estimation of the a_k 's (e.g., Table 5) has already been shown (i.e., $\hat{a}_1 = 0.008912$, $\hat{a}_2 = -0.002896$, $\hat{a}_3 = 0.001063$, $\hat{a}_4 = 0.001046$, $\hat{a}_5 = -0.001446$, and $\hat{a}_6 = 0.001480$). Given that $m = 3$, the estimate of $f(0)$ is computed from Eq. (2.2),

$$\begin{aligned}\hat{f}(0) &= \frac{1}{w^*} + \sum_{k=1}^m \hat{a}_k \\ &= \frac{1}{65} + (\hat{a}_1 + \hat{a}_2 + \hat{a}_3) \\ &= 0.022463.\end{aligned}$$

Density is estimated easily from Eq. (2.1),

$$\begin{aligned}\hat{D} &= \frac{n\hat{f}(0)}{2L} \\ &= \frac{40 \times 0.022463}{2 \times 1000} \\ &= 0.000449 \text{ or } 4.49 \text{ objects/ha.}\end{aligned}$$

This estimate is inferior to the estimates

of 4.86 obtained by using the FS approach and the stopping rule (which suggested $m = 1$), and it has a larger sampling variance because 3 cosine terms are used.

The elements of the variance-covariance matrix of the coefficients \hat{a}_k are computed by using Eq. (2.4) and (2.5); the matrix appears below (note that the matrix is symmetric about the diagonal):

$$\begin{bmatrix} 0.000008959 & 0.000004597 & -0.000000973 \\ 0.000004597 & 0.000012335 & 0.000003024 \\ -0.000000973 & 0.000003024 & 0.00001693 \end{bmatrix}.$$

Although $m = 3$, one needs $2m = 6$ coefficients to estimate the variance-covariance matrix.

The sampling variance of $\hat{f}(0)$ is computed by merely summing all the elements in the variance-covariance matrix. Summing the 9 elements, $\hat{\text{var}}(\hat{f}(0)) = 0.000047296$.

Finally, the sampling variance of the density estimate can be computed. We illustrate the computation by using the Poisson assumption (because no actual area is involved). Using that assumption as an illustration, $\hat{\text{var}}(n) = n$. Then,

$$\begin{aligned} \hat{\text{var}}(\hat{D}) &= \hat{D}^2 \left[\frac{\hat{\text{var}}(n)}{n^2} + \frac{\hat{\text{var}}(\hat{f}(0))}{(\hat{f}(0))^2} \right] \\ &= \hat{D}^2 \left[\frac{1}{n} + \frac{\hat{\text{var}}(\hat{f}(0))}{(\hat{f}(0))^2} \right] \\ &= (0.000449)^2 \left[\frac{1}{40} + \frac{0.0000473}{(0.022463)^2} \right] \\ &= 0.0000000239 \\ \hat{s}\text{e}(\hat{D}) &= (\hat{\text{var}}(\hat{D}))^{1/2} = 0.0001547. \end{aligned}$$

An approximate 95 percent confidence interval can be computed as $\hat{D} \pm 1.96 \hat{s}\text{e}(\hat{D})$, which is 1.46 to 7.52 objects/ha in the example. One might expect the actual confidence interval coverage to be less than 95 percent because the estimator of $f(0)$ is based on a trade-off between variance and bias. The trade-off is accomplished by the stopping rule based on the mean integrated squared error, where $MISE = \text{variance} + \text{bias}^2$. The actual

confidence interval coverage may be less than 95 percent because the bias need not be zero and the “best” FS model is not always selected by the stopping rule. The extent that those intervals are approximate is not known and is the subject of needed research. Some theoretical discussion of this subject is given in PART 3.

The estimated PRB for this example is -10.9 (Fig. 27). The average PRB for 25 replications of this experiment (with m fixed at 3) was 4.0. This seems quite good considering the small sample size of $n = 40$. Further properties of the FS approach are discussed in PART 3.

Stake Data

Laake (1979, unpublished master's thesis, Utah State University, Logan, Utah) conducted an investigation of line transect sampling by placing a known number of wooden stakes in a sagebrush meadow east of Logan, Utah. Lines were walked by graduate students in the Department of Wildlife Science at Utah State University. The sample data on perpendicular distances from one such survey are used as a second example. The observer found 68 stakes within a fixed width of 20 m ($w = 20$ m) on a single line transect 1,000 m long ($L = 1,000$). The actual density was known to be 0.00375, or 37.5 stakes/ha. The true form of $f(x)$, the pdf, was unknown and varied from observer to observer. The perpendicular distances x_i for the 68 stakes found are shown in Table 6. Note that the greatest distance was 18.60 m (that is, the largest order statistic $X_{(68)} = 18.60$), and that value was used for w^* (i.e., $w^* = 18.60 < w = 20.0$). In many situations, one may wish to select a value for w^* such that 1–3 percent of the outliers are excluded from the analysis (see *Example 6 Data Truncation* and Fig. 44). However, for the purpose of this example, $w^* = 18.60$ will suffice.

Estimates of \hat{a}_k for $k = 1, 2, \dots$ are computed from Eq. (2.3),

TABLE 6.—EXAMPLE OF LINE TRANSECT DATA FROM A SAMPLE OF WOODEN STAKES OF KNOWN DENSITY; $n = 68$, $w^* = 18.60$ m, $w = 20$ m, $L = 1,000$ m, AND $D = 0.00375$ STAKES PER METER. THE DATA SHOWN ARE THE PERPENDICULAR DISTANCES IN METERS

1.	2.02	35.	3.79
2.	0.45	36.	15.24
3.	10.40	37.	3.47
4.	3.61	38.	3.05
5.	0.92	39.	7.93
6.	1.00	40.	18.15
7.	3.40	41.	10.05
8.	2.90	42.	4.41
9.	8.16	43.	1.27
10.	6.47	44.	13.72
11.	5.66	45.	6.25
12.	2.95	46.	3.59
13.	3.96	47.	9.04
14.	0.09	48.	7.68
15.	11.82	49.	4.89
16.	14.23	50.	9.10
17.	2.44	51.	3.25
18.	1.61	52.	8.49
19.	31.31	53.	6.08
20.	6.50	54.	0.40
21.	8.27	55.	9.33
22.	4.85	56.	0.53
23.	1.47	57.	1.23
24.	18.60	58.	1.67
25.	0.41	59.	4.53
26.	0.40	60.	3.12
27.	0.20	61.	3.05
28.	11.59	62.	6.60
29.	3.17	63.	4.40
30.	7.10	64.	4.97
31.	10.71	65.	3.17
32.	3.86	66.	7.67
33.	6.05	67.	18.16
34.	6.42	68.	4.08

$$\hat{a}_k = \frac{2}{nw^*} \left[\sum_{i=1}^n \cos\left(\frac{k\pi x_i}{w^*}\right) \right], \quad k = 1, 2, \dots$$

In addition, $a_0 \equiv 2/w^*$. After careful computation, we have $a_0 = 0.107527$, $\hat{a}_1 = 0.050130$, $\hat{a}_2 = 0.010910$, $\hat{a}_3 = -0.004979$, and $\hat{a}_4 = -0.000325$.

Use of the stopping rule (Eq. (2.7)) enables an objective choice of m , the number of FS terms in the estimator of $f(0)$. The stopping rule indicates that 2 terms are appropriate;

$$\frac{1}{w^*} \left(\frac{2}{n+1} \right)^{1/2} \geq |a_{m+1}|$$

or

$$\frac{1}{18.60} \left(\frac{2}{69} \right)^{1/2} = 0.00915 \geq |a_{2+1}| \\ = 0.00498.$$

Therefore, the estimate of $f(0)$ is computed by using Eq. (2.2),

$$\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^m \hat{a}_k,$$

and for that example,

$$f(0) = \frac{1}{18.60} + \sum_{k=1}^2 \hat{a}_k \\ = 0.053763 \\ + (0.050130 + 0.010910) \\ = 0.114803.$$

The estimate of density is made by using Eq. (2.1)

$$\hat{D} = \frac{n\hat{f}(0)}{2L} \\ = 0.0039033 \text{ or } 39.0 \text{ stakes per ha.}$$

In that example, the estimated PRB is 4.

The sampling variance of \hat{D} requires the use of Eq. (2.4)–(2.6). The variance of \hat{a}_1 and \hat{a}_2 is computed as

$$\hat{\text{var}}(\hat{a}_1) = \frac{1}{67} \left[\frac{1}{18.6} (\hat{a}_2 + a_0) - (\hat{a}_1)^2 \right] \\ = 0.0000575$$

$$\hat{\text{var}}(\hat{a}_2) = \frac{1}{67} \left[\frac{1}{18.6} \left(\hat{a}_4 + \frac{2}{18.6} \right) - (\hat{a}_2)^2 \right] \\ = .0000842$$

and the covariance between \hat{a}_1 and \hat{a}_2 is

$$\hat{\text{cov}}(\hat{a}_1, \hat{a}_2) = \frac{1}{(n-1)} \left[\frac{1}{w^*} (\hat{a}_{1+2} + \hat{a}_{2-1}) \right. \\ \left. - (\hat{a}_1 \hat{a}_2) \right] \\ = \frac{1}{67} \left[\frac{1}{18.6} (\hat{a}_3 + \hat{a}_1) - (\hat{a}_1 \hat{a}_2) \right] \\ = 0.0000281.$$

Because the stopping rule indicated $m = 2$, a 2 by 2 variance-covariance matrix with elements is

$$\begin{bmatrix} 0.0000575 & 0.0000281 \\ 0.0000281 & 0.0000842 \end{bmatrix}.$$

Summation of the 4 elements in the variance-covariance matrix (Eq. 2.6) gives $\hat{\text{var}}(\hat{f}(0)) = 0.0001979$.

Several approaches can be taken to compute the estimate of the sampling variance of \hat{D} . The stakes were placed randomly, therefore, n is distributed as a binomial with parameters N and p , where N is known ($N = 150$ in the data). The estimator for p is $\hat{p} = n/N$. If p is small, the Poisson approximation to the binomial is good. However, $\hat{p} = 68/150 = 0.4533$ is large, so the Poisson approximation can be expected to be poor. Here, for illustration, we explore both approaches.

If the Poisson assumption holds, the $\hat{\text{var}}(n)$ is simply

$$\hat{\text{var}}(n) = n.$$

Assuming n is binomial, the $\hat{\text{var}}(n)$ is

$$\hat{\text{var}}(n) = N\hat{p}(1 - \hat{p})$$

and because $\hat{p} = n/N$,

$$\begin{aligned} \hat{\text{var}}(n) &= N\left(\frac{n}{N}\right)\left(1 - \frac{n}{N}\right) \\ &= n\left(1 - \frac{n}{N}\right). \end{aligned}$$

The general estimator for the sampling variance of \hat{D} is appropriate under either assumption.

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

Concerns about Poisson or binomial assumptions for n involve only the first term in the brackets. Note, however, under the binomial assumption

$$\frac{\hat{\text{var}}(n)}{(n)^2} = \frac{1}{n} \left(1 - \frac{n}{N}\right),$$

which shows that as N gets large the Poisson assumption holds, i.e.,

$$\frac{\hat{\text{var}}(n)}{(n)^2} = \frac{1}{n}.$$

One computes the estimate of $\text{var}(\hat{D})$, using the Poisson assumption

$$\begin{aligned} \hat{\text{var}}(\hat{D}) &= (\hat{D})^2 \left[\frac{n}{n^2} + \frac{\hat{\text{var}}(\hat{f}(0))}{(\hat{f}(0))^2} \right] \\ &= (\hat{D})^2 \left[\frac{1}{68} + \frac{0.0001979}{(0.114803)^2} \right] \\ &= 0.0000004528 \\ \hat{s}e(\hat{D}) &= 0.000673. \end{aligned}$$

The estimate of $\text{var}(\hat{D})$ under the binomial assumption is slightly smaller, even though the Poisson assumption is poor because $\hat{p} = 0.45$

$$\begin{aligned} \hat{\text{var}}(\hat{D}) &= (\hat{D})^2 \left[\frac{1}{n} \left(1 - \frac{n}{N}\right) + \frac{\hat{\text{var}}(\hat{f}(0))}{(\hat{f}(0))^2} \right] \\ &= (\hat{D})^2 [0.00804 + 0.0150155] \\ &= 0.0000003513 \\ \hat{s}e(\hat{D}) &= 0.000593. \end{aligned}$$

Hemingway's Data

P. Hemingway (pers. comm.) studied various ungulates in Africa using line transect sampling. Here, some of his 1971 data are used to illustrate use of the FS procedure on ungrouped perpendicular distance data. The emphasis of this example is on the interpretation of the computer output from program TRANSECT. During the survey, 73 animals were detected on a 60-km transect. Only the sighting distance (SD) and the measured sighting angle (MESANG) were recorded in the field; program TRANSECT computes the perpendicular distance (PD) from them. A partial listing of the data and various summary statistics are given in Fig. 28a. (Program TRANSECT lists all the data; space limitations prevent displaying the entire list here.) For example, the average perpendicular distance was 99.25 m, with a standard deviation of 82.37. Finally, a matrix of correlation coefficients and significance levels is printed. The perpendicular distance data are printed, in increasing order (Fig. 28b). Such ordered data are called "order statistics" and are denoted as $X_{(i)}$ for $i = 1$,

DATA TYPE		UNGROUPED DATA			
SAMPLE SIZE	73	LINE LENGTH	60.00	KILOMETERS	
WIDTH		TRUNCATED DATA, WIDTH =		400.0	METERS
PD	SD	MESANG	COMANG	SINE	
1	71.93	100.0	45.00	.719	
2	26.05	150.0	19.00	.174	
3	54.47	200.0	17.00	.292	
4	92.35	150.0	39.00	.616	
5	163.8	200.0	55.00	.819	
6	44.52	200.0	25.00	.423	
7	163.8	200.0	55.00	.819	
8	157.3	250.0	39.00	.629	
9	22.27	160.0	8.00	.139	
10	72.11	140.0	31.00	.515	
11	86.99	130.0	42.00	.669	
12	50.80	130.0	23.00	.391	
13	0.	100.0	0.00	0.000	
14	73.14	100.0	47.00	.731	

SUMMARY STATISTICS

PD	SD	MESANG	COMANG	SINE
99.25	191.0	33.51		.505 MEANS
92.37	81.22	23.25		.306 STANDARD DEVIATIONS
9.641	9.506	7.72		.036 STANDARD ERRORS OF THE MEANS

CORRELATIONS (LOWER OFF DIAGONAL), AND THEIR TWO TAILED SIGNIFICANCE LEVELS (UPPER OFF DIAGONAL)

PD	SD	MESANG	COMANG	SINE
*****	.0000	0.0000		0.0000
.6076	*****	.1998		.3450
.8168	.1522	*****		0.0000
				MESANG
.7880	.1121	.9806		*****
				SINE

FIG. 28a. Partial listing of Hemingway's (1971) data on African ungulates. A variety of summary statistics is shown. The mnemonic codes are Perpendicular Distance (PD), Sighting Distance (SD), MEasured ANGLE (MESANG), COMputed ANGLE (COMANG), and COMputed SINE (COMSINE). In this example, the PD was computed from SD and MESANG, therefore, COMANG is blank.

..., n. For example, $X_{(1)}$ through $X_{(8)}$ are zero, while $X_{(9)} = 8.72$, and so on. A sample histogram of the 73 perpendicular distances is shown in Fig. 28c. Program TRANSECT computes 3 such histograms automatically, using different groupings (class intervals). In that example, the printed histogram was based on an interval width of 66.67 m. The user can specify additional histograms using any specified interval width; such an option also

allows the use of unequal intervals. Program TRANSECT prints similar sample histograms for the data on sighting distance, sighting angle, and the sine of the angle.

The FS estimation procedure produced by TRANSECT is shown in Fig. 28d. The width was chosen to be 400 m, the maximum observed distance. The stopping rule indicated that 2 terms were required for the data. The program then

0.	0.	0.	0.	0.	0.	0.	0.	8.72	10.5
22.3	26.0	26.0	30.5	30.5	31.7	34.2	35.1	38.0	41.0
42.1	50.8	55.1	58.5	63.6	64.3	65.0	68.8	71.1	71.8
71.9	72.1	73.1	76.6	77.6	78.1	84.5	84.5	86.0	86.0
87.0	90.0	92.3	94.0	96.4	96.4	106.	115.	123.	123.
129.	129.	143.	143.	150.	151.	153.	157.	161.	164.
164.	164.	166.	175.	188.	193.	200.	200.	246.	260.
272.	378.	400.							

FIG. 28b. The order statistics for the perpendicular distances computed from Hemingway's (1971) data.

printed the estimates of the 2 Fourier coefficients, a_1 and a_2 , as 0.003106 and 0.000909, respectively, followed by estimates of $f(0)$, 0.0065 and D , 0.0396. The estimated standard errors, coefficients of variation, and approximate 95 percent confidence intervals are given, and the sampling correlation matrix of the parameter estimates is printed. In that example, the density of animals is estimated to be approximately 0.04 animals per hectare with a standard error of 0.0006.

Several figures such as that shown in

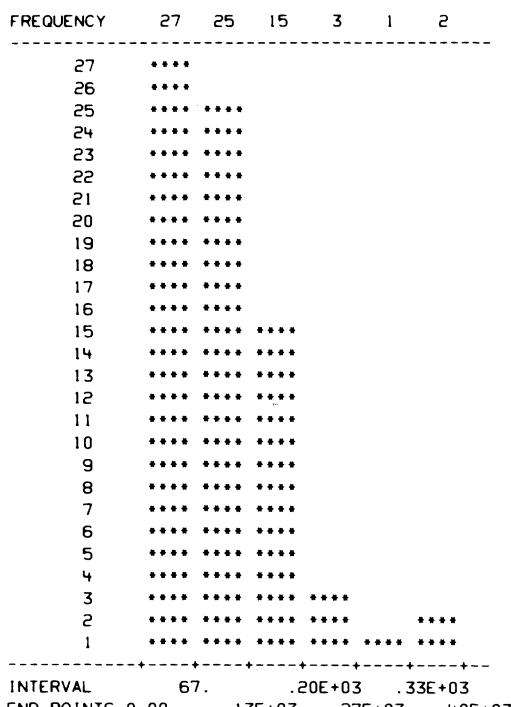


FIG. 28c. Sample histogram of the perpendicular distance data collected by Hemingway (1971). A variety of such histograms is prepared by program TRANSECT. Interval width is 66.67 m.

Fig. 28e are printed by program TRANSECT to allow the user to get an impression as to how the selected FS model "fits" the data. Here, TRANSECT has grouped the data into 9 groups of equal size and printed a histogram. Actually, $n_i/(\Delta n)$ is plotted; therefore, the first bar in the histogram is $19/(40 \times 73)$. The estimated Fourier series function, $\hat{f}(x)$, is printed on the sample histogram. From Fig. 28e, it is obvious that the model fits the data quite well. The question of "fit" is further quantified by several chi-square goodness of fit tests. One such test that is automatically prepared by TRANSECT is shown in Fig. 28f. In that example, the same grouping is used and the test statistics are printed. Overall, there is a test statistic of 4.03 with 6 degrees of freedom. The degrees of freedom for the test is the number of groups (9) minus the number of parameters (2 in this example, a_1 and a_2) minus 1 = 6. This is a test of the null hypothesis that the model fits the data. A test statistic of 4.03 is expected with probability 0.67 if the null hypothesis is true. Therefore, we have no reason to reject the 2-term FS model.

Program TRANSECT provides a multitude of other capabilities and options. Several of these are explored under ILLUSTRATIVE EXAMPLES. Use of TRANSECT for the analysis of ungrouped, perpendicular distances by means of the FS procedure is simple.

Summary

The use of nonparametric approaches is advantageous because they provide good estimates under much broader circumstances than do the simpler parametric models. Basically, the only requirement in applying the Fourier series is the

UNGROUPED DATA				
SAMPLE SIZE	73	KILOMETERS		
LINE LENGTH	60.00	TRUNCATED DATA, WIDTH =	600.0	METERS
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.3106E-02	.2567E-03	9.6	.2584E-02 .3629E-02
A(2)	.9087E-03	.4068E-03	44.8	.1115E-03 .1706E-02
F(0)	.6515E-02	.6004E-03	9.2	.5337E-02 .7693E-02
D	.3963E-01	.1906E-02	14.9	.2806E-01 .5121E-01

DENSITY(D) UNITS ARE NUMBERS/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2
1	1.000	.573
2	.573	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 73. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1370E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .8505E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 61.70

FIG. 28d. Summary of the Fourier series estimation procedure using Hemingway's (1971) data as an example. The estimates appear in exponential notation, e.g., 0.310633E-02 is 0.00310633, 0.908744E-03 is 0.000908744, etc.

continuity of $f(x)$ at the point $x = 0$. Under some distributions, the FS estimator actually is more efficient for the sample sizes likely to be encountered in practice than are the corresponding parametric estimators. In general, the efficiency of the FS estimator developed to cover a broad set of underlying distributions compares favorably with simple parametric estimators developed only for very specific distributions.

The FS estimator seems to perform well, often even with samples as small as 30–40. Consideration of the shape of the cosine function, results of our simulation studies, and our theoretical results on estimation efficiency indicate that the FS estimator will do quite well if the underlying distribution $f(x)$ or $g(x)$ is concave, nearly linear, or mildly convex.

ANALYSIS AND INFERENCE PROCEDURES FOR PERPENDICULAR DISTANCES, GROUPED DATA

The basic problem in the analysis of line transect data is to estimate $f(0)$ and it does not change just because the data are grouped. If x_1, \dots, x_n are a sample of

perpendicular distances, they may be grouped by specifying k intervals and counting the frequency of observed distances in each interval. The boundaries of such intervals are called *cut points* and denoted by c_j . Hence, $0 < c_1 < c_2 < \dots < c_{k-1} < c_k$ defines k intervals; specifically, the j^{th} interval is $(c_{j-1}, c_j]$, with $c_0 \equiv 0$ and $c_k \equiv w$ (or w^*). As for ungrouped data, w can be infinite; however, for application of the FS analysis method to grouped (or ungrouped) data, either w must be finite or a finite w^* must be specified. Of course, w^* may be chosen to be larger than the greatest perpendicular distance, thus no data need be excluded from the analysis if so desired.

The 2 basic reasons for considering analysis of grouped data are first, that line transect data are sometimes recorded grouped, and second, that grouping is sometimes necessary to achieve robustness (due to problems of data recording or inaccurate field measurements). In the first case, exact perpendicular distances are not even recorded; rather, at the time of detection each observation is classified into one of the k preestablished distance intervals. In the second case, severe inaccuracies in recorded distances, espe-

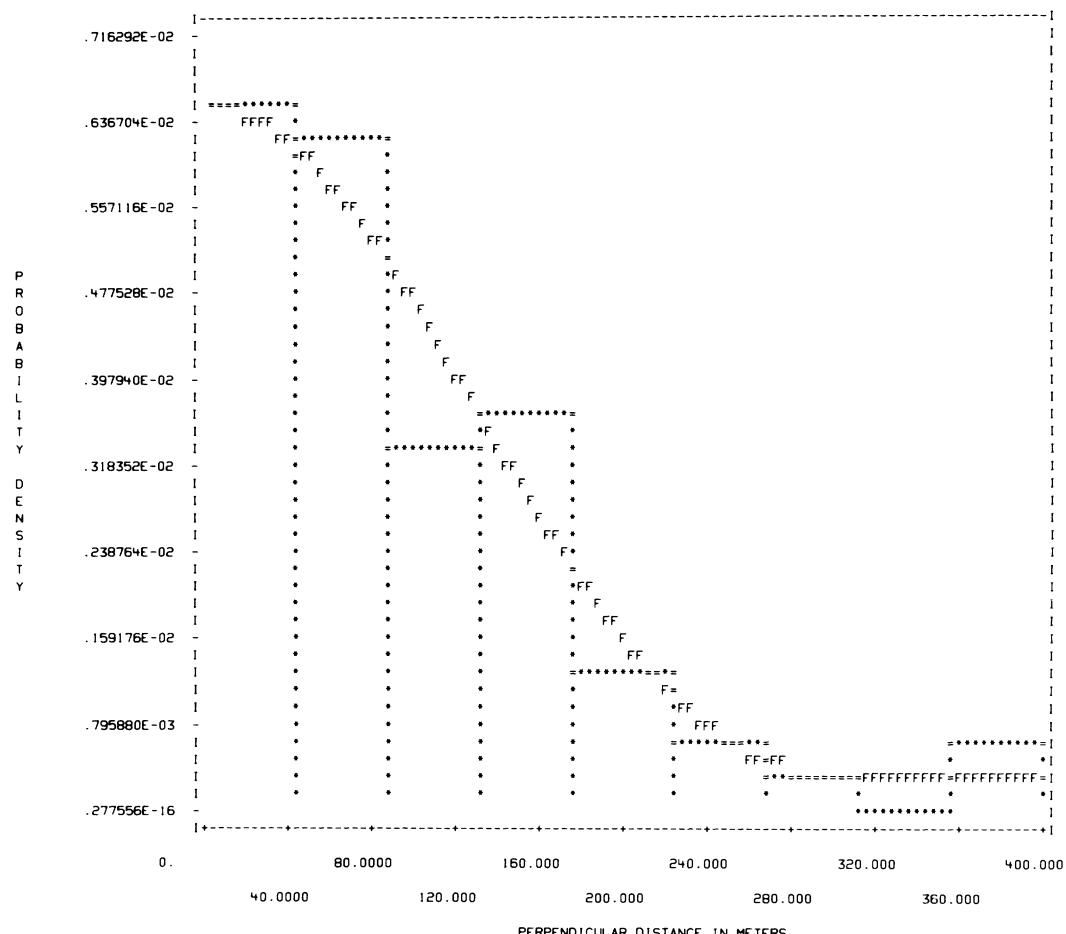


FIG 28e. A sample histogram of Hemingway's (1971) data showing the estimated Fourier series model $f(x)$. That model intercepts the Y axis at 0.0065 (i.e., $\hat{f}(0)=0.0065$).

cially near $x = 0$, preclude the analysis of the data as exact distances (see ILLUSTRATIVE EXAMPLE 5). It is not appropriate to analyze grouped data by methods derived only for ungrouped data.

The basic grouped data are the frequencies n_1, \dots, n_k , where total sample size is $n = n_1 + \dots + n_k$, and n_j is the number of perpendicular distances falling in the j^{th} interval. For example, in the preceding section, the first example used simulated data from the half-normal distribution (see Table 5) with sample size 40 and $w = 65$. Table 7 shows the cut points and frequencies when the data are grouped into 6 intervals of equal length

(see Fig. 26 for a histogram of the grouped data). From Table 7, we see that 10 objects were detected in the interval 0.0 to 10.8 m from the transect line, but only 2 sightings were between 54.2 and 65.0 m.

If $f(x)$ is any probability density function model for the ungrouped perpendicular distance data, an estimator of $f(0)$ can validly be based on this model using grouped data. As in the case of ungrouped perpendicular data, the model recommended is the Fourier series, with an estimation method appropriate for grouped data. Hence, the FS approach is the only method discussed below.

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	44.4	21	.433E-02
2	44.4	88.9	20	.170
3	88.9	133.	11	.664
4	133.	178.	12	.766
5	178.	222.	4	.286
6	222.	267.	2	.890E-01
7	267.	311.	1	.231E-01
8	311.	356.	0	.872
9	356.	400.	2	.95
TOTAL CHI-SQUARE VALUE =		4.031	DEGREES OF FREEDOM =	6
PROBABILITY OF A GREATER CHI-SQUARE VALUE = .67243638				

FIG. 28f. Goodness of fit summary of Hemingway's (1971) data. Here, a satisfactory fit is indicated ($\chi^2=4.03$, 6 df, $P=0.67$).

The Multinomial Distribution

The probability distribution of grouped frequency data is multinomial. This is a common and thoroughly understood statistical distribution (see, e.g., Johnson and Kotz 1969). For k classes, there are k cell probabilities p_1, \dots, p_k interpreted as follows. Given that an object is detected, p_j is the probability that its perpendicular distance will be in the j^{th} interval. Because it must be in one of the intervals, the probabilities sum to 1 ($p_1 + \dots + p_k = 1$).

To link the FS model for $f(x)$ to the grouped data properly requires that those cell probabilities are formally expressible as functions of the parameters a_1, \dots, a_m of the FS model. Recall that the FS model for $f(x)$ is

$$f(x) \doteq \frac{1}{w} + \sum_{i=1}^m a_i \cos\left(\frac{i\pi x}{w}\right).$$

TABLE 7.—EXAMPLE OF GROUPED LINE TRANSECT DATA GENERATED FROM AN UNDERLYING HALF-NORMAL DISTRIBUTION WITH $\sigma = 33.334$, $w = 65$, m , AND SAMPLE SIZE $n = 40$

Interval j	Interval limits		Frequency n_j
	Lower (c_{j-1})	Upper (c_j)	
1	0.0	10.8	10
2	10.8	21.7	7
3	21.7	32.5	9
4	32.5	43.3	9
5	43.3	54.2	3
6	54.2	65.0 = w	2

Under the FS model, each cell probability p_j has a closed form expression:

$$p_j = \frac{c_j - c_{j-1}}{w} + \sum_{i=1}^m (a_i) \frac{w}{i\pi} \cdot \left[\sin\left(\frac{i\pi c_j}{w}\right) - \sin\left(\frac{i\pi c_{j-1}}{w}\right) \right]. \quad (2.8)$$

Although formidable looking, Eq. (2.8) is not really complex; only the a_i are unknown parameters to be estimated. The expression for p_1 in the example of Table 7 is

$$p_1 = \frac{10.8 - 0}{65} + \sum_{i=1}^m (a_i) \left(\frac{65}{i\pi} \right) \cdot \left[\sin\left(\frac{(i\pi)(10.8)}{65}\right) - 0 \right] \\ = \frac{1}{6} + \sum_{i=1}^m a_i \left(\frac{20.60}{i} \right) \sin(i(0.52199)).$$

If the value of m is specified, this equation can be made even more specific. For $m = 1$, $p_1 = 0.1667 + a_1(10.345)$. If $m = 3$, $p_1 = 0.1667 + a_1(10.345) + a_2(8.959) + a_3(6.8967)$. For any given m , all the cell probabilities can be expressed as simple, linear functions of the unknown parameters a_1, \dots, a_m .

Under the multinomial probability distribution for grouped data, the probability of the observed frequency counts is given by

$\Pr\{n_1, \dots, n_k\} = A(p_1)^{n_1}(p_2)^{n_2} \cdots (p_k)^{n_k}$, where A is a function only of the observed data,

$$A = \frac{(n!)^k}{(n_1!) \cdots (n_k!)},$$

and the cell probabilities p_1, \dots, p_k are functions of the unknown parameters a_1, \dots, a_m .

For a given set of data, we define a special function called the log-likelihood function:

$$\ln \mathcal{L}(a_1, \dots, a_m) = \ln(A) + \sum_{i=1}^k n_i \ln(p_i). \quad (2.9)$$

Because the data are considered fixed here, Eq. (2.9) is a function of only the

unknown parameters. In particular, the term $\ln(A)$ is a constant, not affected by variations in the parameters a_1, \dots, a_m . Good estimates of those parameters can be found as those values \hat{a}_i that maximize the likelihood function (hence the term maximum likelihood (ML) estimators).

Estimation with the Fourier Series Model

Under the FS model,

$$\hat{f}(0) = \frac{1}{w} + \sum_{i=1}^m \hat{a}_i,$$

where m itself must be chosen by some procedure. For ungrouped perpendicular distance data, Eq. (2.3) is an explicit formula for the \hat{a}_i ; also there is a simple stopping rule for the choice of m . No such simple formulas exist, however, when the data are grouped.

With cell probabilities given by Eq. (2.8), it is possible to compute the ML estimators $\hat{a}_1, \dots, \hat{a}_m$ and their sampling variances and covariances for any given value of m . It is, however, necessary to use iterative numerical methods (i.e., use a computer). Program TRANSECT performs those computations. The basics of the underlying theory are sketched in PART 3.

Program TRANSECT computes the sampling covariances but prints only a related quantity, the sampling correlations of the estimators. The estimated sampling correlation between \hat{a}_i and \hat{a}_j is just

$$\hat{\text{cor}}(\hat{a}_i, \hat{a}_j) = \frac{\hat{\text{cov}}(\hat{a}_i, \hat{a}_j)}{\hat{s}\text{e}(\hat{a}_i)\hat{s}\text{e}(\hat{a}_j)}.$$

As for ungrouped data, $\hat{\text{cov}}(\hat{a}_i, \hat{a}_j)$ can be used to represent the sampling variance $\hat{\text{var}}(\hat{a}_i)$. With that notational convention, the sampling variance of $\hat{f}(0)$ is

$$\hat{\text{var}}(\hat{f}(0)) = \sum_{i=1}^m \sum_{j=1}^m \hat{\text{cov}}(\hat{a}_i, \hat{a}_j),$$

which is the same as Eq. (2.6) for ungrouped data.

For the example data presented in Table 7, the ML estimator of a_1 , when m is 1, is $\hat{a}_1 = 0.0097 \pm 0.0029$. Therefore, for $m = 1$,

$$\hat{f}(0) = \frac{1}{65} + 0.0097 = 0.0251,$$

with a standard error of 0.0029 (see Fig. 29a). As discussed in the previous section, these data were from a computer simulation study with $L = 1,000$ m, $f(0) = 0.0252$, and a true density of 0.000504 objects/m², or 5.04 objects/ha. Applying the fundamental formula, $\hat{D} = n\hat{f}(0)/2L$,

$$\hat{D} = \frac{40 \times 0.0251}{2 \times 1000} = 0.000502$$

or $\hat{D} = 5.02$ objects/ha.

As discussed in PART 1, the sampling variance of \hat{D} requires that an estimate of $\text{var}(n)$ be made. For illustration purposes only, it is assumed that n is a Poisson variable, hence $\hat{\text{var}}(n) = n$. From Eq. (1.17) for $\hat{\text{var}}(\hat{D})$,

$$\begin{aligned} \hat{\text{var}}(\hat{D}) &= (\hat{D})^2 \left[\frac{\hat{\text{var}}(n)}{(n)^2} + \frac{\hat{\text{var}}(\hat{f}(0))}{(\hat{f}(0))^2} \right] \\ &= (0.000502)^2 \left[\frac{1}{40} + \left(\frac{0.0029}{0.0251} \right)^2 \right] \\ &= (0.000502)^2 [0.025 + 0.01335] \\ &= (0.0000983)^2, \end{aligned}$$

or $\hat{s}\text{e}(\hat{D}) = 0.0000983$. Converting to hectares gives $\hat{D} = 5.02 \pm 0.98$; and an approximate 95 percent confidence interval for D is $\hat{D} \pm 1.96 \hat{s}\text{e}(\hat{D})$, or 3.10 to 6.94. These results should be compared to those for $m = 1$ when the FS model is used to analyze the corresponding ungrouped data: $\hat{D} = 4.86 \pm 0.97$. Complete results for this example are shown in Figs. 29a, 29b, and 29c (those figures also include material discussed below).

Selection of m

An objective way is needed to select the number of terms, m , in the FS model to be used. For ungrouped data, the rule is to compute \hat{a}_m sequentially until the first estimate \hat{a}_{m+1} is such that

SOME RESULTS DURING THE ITERATION FOR ML VALUES ARE GIVEN BELOW
 ITER = THE ITERATION COUNT.
 VLL = THE LOG-LIKELIHOOD VALUE AT THE PARAMETER VALUE(S) FOR THAT ITERATION.
 ERROR = ERROR CONDITIONS ENCOUNTERED DURING THE ITERATION. THE NUMBERS GIVEN
 CORRESPOND TO CONDITIONS EXPLAINED BELOW.
 PARAMETER VALUES ARE PRINTED IN DOTS, T.F., A(1),...,A(NPAR)

ITER	VLL	POINT0	PARAMETER VALUES
0	-48.44770		.7767405E-02
1	-68.23702		.9980497E-02
2	-68.22823		.9847571E-02
3	-68.22791		.9714479E-02
4	-68.22790		.9777786E-02
5	-68.22790		.7722564E-02

CONVERGENCE ACHIEVED, ITERATIONS = 5

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
				A(1)	F(0)
A(1)	.9730E-02	.2000E-02	20.8	.4045E-02	.1541E-01
F(0)	.2511E-01	.2000E-02	11.5	.1943E-01	.3080E-01
0	5.023	.0R14	10.6	.099	6.950

DENSITY(D) UNITS ARE NUMBERS/HECTARE

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 40. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .2500E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1334E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 65.21

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
 THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	10.4	10.69	.445E-01
2	10.9	21.7	7.62	.714
3	21.7	32.5	7.74	.204
4	32.5	43.1	5.50	2.09
5	43.3	54.2	3.72	.140
6	54.2	65.0	2.64	.155

TOTAL CHI-SQUARE VALUE = 3.243 DEGREES OF FREEDOM = 4

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .50216815

FIG. 29a. Results of maximum likelihood estimation of \hat{a}_1 in the 1-term Fourier series model for the grouped data of Table 7. See text for explanations and discussions.

$$\frac{1}{w} \left(\frac{2}{n+1} \right)^{1/2} \geq |\hat{a}_{m+1}|.$$

Although that stopping rule could be used here, it is not recommended because it does not necessarily lead to a unique choice of m. The problem is that the \hat{a}_i , $i = 1, \dots, m$, vary slightly with the value of m used (see Figs. 29a, 29b, 29c). Thus, for analysis of grouped data, \hat{a}_1 computed when $m = 1$ is not identical to \hat{a}_1 computed when $m = 2$. For ungrouped data using Eq. (2.3), \hat{a}_i is independent of the chosen value of m. The above rule was applied to analysis of grouped data and found it does not per-

form as well as the likelihood ratio method discussed below.

The basic idea in selecting m is to achieve a FS model that provides a good fit to the data. Two measures of fit are useful in making this judgment: first (and foremost) the log-likelihood value of Eq. (2.9) and, second, the usual chi-square goodness of fit test.

Likelihood Ratio Test

The maximum likelihood (ML) estimates $\hat{a}_1, \dots, \hat{a}_m$ are obtained by maximizing the log-likelihood function Eq. (2.9). Let $\ln \mathcal{L}_m$ represent the log-likelihood

SOME RESULTS DURING THE ITERATION FOR ML VALUES ARE GIVEN BELOW
ITER = THE ITERATION COUNT
XLL = THE LOG-LIKELIHOOD VALUE AT THE PARAMETER VALUE(S) FOR THAT ITERATION
ERROR = ERROR CONDITIONS ENCOUNTERED DURING THE ITERATION. THE NUMBERS GIVEN
CORRESPOND TO CONDITIONS EXPLAINED BELOW.
PARAMETER VALUES ARE PRINTED IN ORDER, I.E., A(1),...,A(NPAR)

ITER	XLL	ERROR	PARAMETER VALUES
0	-67.54568		.7257405E-02 -3839439E-02
1	-67.46013		.9349519E-02 -3979168E-02
2	-67.45598		.9331500E-02 -3938998E-02
3	-67.45598		.9330957E-02 -3945043E-02
4	-67.45598		.9331125E-02 -3944234E-02

CONVERGENCE ACHIEVED, ITERATIONS = 4

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.9331E-02	.2048E-02	35.4	.2554E-02 .1411E-01
A(2)	-.3944E-02	.3319E-02	84.1	-.1045E-01 .2561E-02
F(1)	.1977E-01	.5309E-02	27.3	.9196E-02 .3035E-01
D	3.954	1.247	31.5	1.510 6.399

DENSITY(D) UNITS ARE NUMBERS/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2
1	1.000	.481
2	.481	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 40. THIS ASSUMPTION IS CRITICAL SINCE THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .2500E-01

SQUARED COEFFICIENT OF VARIATION FOR F(1) = .7447E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 25.13

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	DESERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	10.9	9.70	.195
2	10.8	21.7	7.0	.524
3	21.7	32.5	9.00	.121E-06
4	32.5	43.3	7.15	.476
5	43.3	54.2	4.15	.317
6	54.2	65.0	2.00	.211E-01

TOTAL CHI-SQUARE VALUE = 1.533 DEGREES OF FREEDOM = 3

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .67469124

FIG. 29b. Results of maximum likelihood estimation of a_1 and a_2 in the 2-term Fourier series model for the grouped data of Table 7. See text for explanations and discussions.

hood function evaluated at the ML estimates. A test of the null hypothesis that $E(\hat{a}_{m+1}) = 0$ versus the alternative $E(\hat{a}_{m+1}) \neq 0$ can be based on the ratio $\mathcal{L}_m / \mathcal{L}_{m+1}$, hence the name likelihood ratio test. In practice, the test takes the form of $\chi^2 = -2(\ln \mathcal{L}_m - \ln \mathcal{L}_{m+1})$, where under the null hypothesis, χ^2 is distributed as a chi-square variable with 1 degree of freedom. Thus, if χ^2 exceeds 3.84, we reject the null hypothesis at the 5 percent level.

By rejecting that null hypothesis, it is

concluded that \hat{a}_{m+1} is a significant term in the FS model and, therefore, that it should be included in $\hat{f}(0)$. Conversely, if the hypothesis is not rejected, it is concluded that the m term FS model is an adequate model for the given data.

Program TRANSECT computes the ML estimates of a_1, \dots, a_m (for any allowable value of m) for the FS model with grouped data and prints all necessary statistics, such as \hat{a}_i , $\hat{s}e(\hat{a}_i)$, $\hat{f}(0)$, $\hat{s}e(\hat{f}(0))$, \hat{D} , $\hat{s}e(\hat{D})$ and the values of the log-

SOME RESULTS DURING THE ITERATION FOR MLE VALUES ARE GIVEN BELOW

ITER = THE ITERATION COUNT

XLL = THE LOG-LIKELIHOOD VALUE AT THE PARAMETER VALUE(S) FOR THAT ITERATION

ERROR = ERROR CONDITIONS ENCOUNTERED DURING THE ITERATION. THE NUMBERS GIVEN

CORRESPOND TO CONDITIONS EXPLAINED BELOW.

PARAMETER VALUES ARE PRINTED IN ORDER, I.E., A(1),...,A(NPAR)

ITER	XLL	FROM	PARAMETER VALUES		
0	-67.45990		.72874605E-02	-.3839439E-02	.1212505E-02
1	-67.42333		.1066304E-02	-.3628187E-02	.1190060E-02
2	-67.42250		.3986742E-02	-.3646034E-02	.1090697E-02
3	-67.42248		.1993823E-02	-.3647790E-02	.1046986E-02
4	-67.42247		.1084475E-02	-.3647504E-02	.1041548E-02
5	-67.42247		.1084489E-02	-.3647637E-02	.1041319E-02
6	-67.42247		.1084491E-02	-.3647621E-02	.1041224E-02

CONVERGENCE ACHIEVED, ITERATIONS = 6

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.8085E-02	.3915E-02	37.3	.2175E-02 .1399E-01
A(2)	-.3644E-02	.3927E-02	94.7	-.1056E-01 .3236E-02
A(3)	.1041E-02	.3810E-02	344.6	-.6073E-02 .8136E-02
F(0)	.2096E-01	.6940E-02	33.3	.7243E-02 .3448E-01
D	4.173	.1438	34.9	1.157 7.188

DENSITY(D) UNITS ARE NUMBERS/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2	3
1	1.000	.415	-.149
2	.416	1.000	.329
3	-.168	.328	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 40. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .2500E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1109

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 18.39

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	10.4	9.09	.114
2	10.8	71.7	9.83	.381
3	21.7	32.5	9.	.59
4	32.5	43.3	7.14	.364
5	43.3	54.2	4.51	.504
6	54.2	65.0	1.73	.436E-01

TOTAL CHI-SQUARE VALUE = 1.427 DEGREES OF FREEDOM = 2

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .49000337

FIG. 29c. Results of maximum likelihood estimation of a_1 , a_2 , and a_3 in the 3-term Fourier series model for the grouped data of Table 7. See text for explanations and discussions.

likelihoods, $\ln \mathcal{L}_m$ (labeled XLL on the computer output). The values of the log-likelihood, XLL, and the parameters at each step of the iteration are printed to provide information on the convergence of the iterative procedure. The last line printed gives $\ln \mathcal{L}_m$ at the maximum.

Figs. 29a, 29b, and 29c show the results of applying the FS model to the grouped data of Table 7 for $m = 1, 2$, and 3 terms, respectively. The results for the log-like-

lihoods ($\ln \mathcal{L}_m = \text{XLL}$) are -68.22790 , -67.45998 , and -67.42247 for the 1-, 2-, and 3-term models, respectively. The likelihood ratio test of $E(\hat{a}_2) = 0$ is computed as

$$\begin{aligned} \chi^2 &= -2(\ln \mathcal{L}_1 - \ln \mathcal{L}_2) \\ &= -2(-68.22790 + 67.45998) \\ &= -2(-0.76792) \\ &= 1.53584. \end{aligned}$$

From tables of the chi-square distribu-

tion with 1 degree of freedom, that value is not significant. Thus, it is concluded \hat{a}_2 is not needed in the FS model of these data; the 1-term FS model is adequate.

As a further example, the test of $E(\hat{a}_3) = 0$ is computed as

$$\begin{aligned}\chi^2 &= -2(-67.45998 + 67.42247) \\ &= 0.07502,\end{aligned}$$

which is clearly not significant. That result is to be expected, if the second term of the FS model is not significantly different from zero, one can expect \hat{a}_3 similarly to be not significantly different from zero.

The Chi-square Goodness of Fit Test

The likelihood ratio test is the critical procedure for selection of the number of terms in the FS model for grouped data. That procedure itself strongly tends to produce a model that fits the data. Nonetheless, it is still of interest to look at the chi-square goodness of fit test for each model (i.e., each value of m used). If the likelihood ratio test gives a significance level near 0.05, the goodness of fit test can be used to decide what value of m to use. We emphasize that problems are associated with this goodness of fit test as a basis for model selection (see PART 1, *Chi-square Goodness of Fit Tests*).

Figs. 29a, 29b, and 29c also give the goodness of fit tests for the data of Table 7 for $m = 1, 2$, and 3, respectively. Specifically, the results are:

m	Chi-square value	Degrees of freedom	P
1	3.3429	4	0.5022
2	1.5329	3	0.6747
3	1.4267	2	0.4900.

Here, P denotes the achieved significance level of the test. All 3 models fit the data. This is consistent with our earlier finding that the 1-term FS model is adequate for the data, and confirms our choice of $m = 1$.

To complete the example, one compares the results for the 1-, 2-, and 3-term

FS models as given in Figs. 29a, 29b, and 29c, respectively. For example, \hat{a}_1 is seen to vary slightly from model to model (it is 0.00973, 0.00833, and 0.00808 for 1, 2, and 3 terms, respectively). Of more importance is the behavior of $f(0)$ and \hat{D} . In particular, the best estimate of D comes from the 1-term FS model. The coefficients of variation of \hat{D} are 19.6, 31.5, and 36.9 percent for 1-, 2-, and 3-term FS models (based on using $\text{var}(n) = n$), respectively. Those results illustrate the loss in precision to be expected when unnecessary terms are included in the FS model (or, indeed, any model).

Examples

Stake Data

Table 6 presents the ungrouped perpendicular distance data from one repetition of a controlled line transect field study (Laake unpublished thesis). In that study, 150 stakes were placed at random in an area 1,000 m long by 40 m wide, hence $w = 20$ m. Observers walked the transect and searched for stakes visually. Distance and angle measurements were taken carefully for each stake detected. The data of Table 6 have already been used to illustrate the FS method for ungrouped data. That example is useful both because the true stake density is known (37.5 stakes/ha) and because the results for grouped analysis can be compared to those of the ungrouped analysis. (Note that this same example will also be used to illustrate methods based on sighting angles and distances.)

The example also illustrates that the intervals used to group the data do not have to be of equal length. We use 10 intervals ($k = 10$) with cut points at 1, 2, 3, 4, 5, 7, 9, 11, 15, and 20 m. The frequency counts corresponding to the intervals so determined are shown in Fig. 30a (specifically, the n_i are, in order, 8, 6, 4, 13, 7, 8, 7, 6, 5, and 4). Fig. 30a gives the analysis of those data for a 1-term FS model, while Fig. 30b gives results for the 2-term FS model.

STAE RESULTS DURING THE ITERATION FOR ML VALUES ARE GIVEN BELOW
 ITER = THE ITERATION COUNT
 XLL = THE LOG-LIKELIHOOD VALUE AT THE PARAMETER VALUE(S) FOR THAT ITERATION
 ERROR = ERROR CONDITIONS ENCOUNTERED DURING THE ITERATION. THE NUMBERS GIVEN
 CORRESPOND TO CONDITIONS EXPLAINED BELOW.
 PARAMETER VALUES ARE PRINTED IN ORDER, I.E., A(1),...,A(NPAP)

ITER	XLL	ERROR	PARAMETER VALUES
0	-156.1836		.49442195E-01
1	-157.5745		.4423826E-01
2	-157.5534		.4452984E-01
3	-157.5520		.4450934E-01
4	-157.5510		.4450944E-01
5	-157.5519		.4453605E-01
6	-157.5519		.44517105E-01

CONVERGENCE ACHIEVED. ITERATIONS = 6

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.4537E-01	.5284E-02	11.7	.3501E-01 .5573E-01
F(0)	.9537E-01	.4926E-02	5.5	.8501E-01 .1057
D	32.63	4.323	13.3	23.95 40.90

DENSITY(D) UNITS ARE NUMBERS/HECTARE

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 58. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1471E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .3072E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 82.72

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
 THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	1.00	5.47	.360
2	1.00	2.00	5.40	.246E-01
3	2.00	3.00	5.25	.808
4	3.00	4.00	5.03	8.06
5	4.00	5.00	4.74	.275
6	5.00	7.00	10.41	.559
7	7.00	9.00	4.70	.332
8	9.00	11.0	6.90	.941E-01
9	11.0	15.0	8.09	1.19
10	15.0	20.0	5.11	.253

TOTAL CHI-SQUARE VALUE = 11.95 DEGREES OF FREEDOM = 8

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .15341747

FIG. 30a. Analysis of stake data (see Table 6) grouped into 10 intervals using a 1-term Fourier series model. See text for discussion.

A decision must be made on the choice of m. Fig. 30a shows that the solution for the ML estimator of a_1 took 6 iterations. The value of the log-likelihood ($\ln \mathcal{L}_1$) at the maximum is given at the sixth step as -157.5519; the corresponding value of $\ln \mathcal{L}_2$ is -156.1836 (Fig. 30b). The test of whether 1 term is adequate is computed by

$$\begin{aligned}\chi^2 &= -2(\ln \mathcal{L}_1 - \ln \mathcal{L}_2) \\ &= -2(-157.5514 + 156.1836) \\ &= 2.7356.\end{aligned}$$

This value is not significant ($P = 0.10$), hence, one tentatively concludes that a 1-term FS model is adequate and that decision is confirmed by the chi-square goodness of fit test shown in Fig. 30a. That test statistic value is 11.95, which, with 8 degrees of freedom, is not significant ($P = 0.153$).

Having decided to use a 1-term FS model as the basis for the analysis of the data, one obtains results (from Fig. 30a) $\hat{a}_1 = 0.04537 \pm 0.00529$, and $f(0) = 0.09537 \pm 0.00529$. The estimated den-

SOME RESULTS DURING THE ITERATION FOR MLF VALUES ARE GIVEN BELOW
 ITER = THE ITERATION COUNT
 XLL = THE LOG-LIKELIHOOD VALUE AT THE PARAMETER VALUE(S) FOR THAT ITERATION
 ERROR = ERROR CONDITIONS ENCOUNTERED DURING THE ITERATION. THE NUMBERS GIVEN
 CORRESPOND TO CONDITIONS EXPLAINED BELOW.
 PARAMETER VALUES ARE PRINTED IN ORDER, I.E., A(1),...,A(NPAR)

ITER	XLL	ERROR	PARAMETER VALUES
0	-156.2430		.4994219E-01 .1302839E-01
1	-156.1937		.5196249E-01 .1315535E-01
2	-156.1936		.5196409E-01 .1318638E-01

CONVERGENCE ACHIEVED, ITERATIONS = 2

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.5197E-01	.4551E-02	12.6	.3911E-01 .6483E-01
A(2)	.1319E-01	.7512E-02	47.0	-.1536E-02 .2791E-01
F(0)	.1152	.1200E-01	10.5	.9145E-01 .1389
D	39.15	6.281	14.0	26.84 51.46

DENSITY(D) UNITS ARE NUMBER/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

1	2
1 1.000	.474
2 .474	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = SR. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1471E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1103E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 57.15

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
 THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1 0.00	1.00	8.	7.40	.506E-02
2 1.00	2.00	8.	7.53	.348
3 2.00	3.00	4.	7.29	1.49
4 3.00	4.00	13.	6.92	5.61
5 4.00	5.00	7.	6.22	.987E-01
6 5.00	7.00	8.	10.30	.551
7 7.00	9.00	7.	7.55	.398E-01
8 9.00	11.0	5.	5.04	.185
9 11.0	15.0	5.	5.37	.187E-01
10 15.0	20.0	4.	3.95	.722E-03

TOTAL CHI-SQUARE VALUE = 8.343 DEGREES OF FREEDOM = 7

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .90333469

FIG. 30b. Analysis of stake data (see Table 6) grouped into 10 intervals using a 2-term Fourier series model. See text for discussion.

sity of stakes is 32.43 (stakes/ha). To compute a standard error of that value, it is assumed that n is a Poisson random variable (that is reasonable because the stakes were randomly placed). Thus,

$$\text{var}(\hat{D}) = (32.43)^2 \left[\frac{1}{68} + (0.5543)^2 \right] \\ = 18.69,$$

or $\hat{\text{se}}(\hat{D}) = 4.32$. Because of the way $\hat{\text{var}}(\hat{D})$ is estimated (i.e., it is not based

on any replication), the 95 percent confidence interval on D is computed as $\hat{D} \pm 1.96 \hat{\text{se}}(\hat{D})$, which gives 23.95 to 40.90. That interval covers the true density. Finally, it is noted that the coefficient of variation of \hat{D} is 13.3 percent.

Those results can be compared to the analysis of the same data using the ungrouped FS method. However, the 2 analyses are not perfectly comparable because the ungrouped analysis used $w^* = 18.6$ m (the maximum observed perpen-

dicular distance), whereas $w^* = w = 20$ m was used for the grouped data. Also, in the analysis of the ungrouped data, $m = 2$ was chosen and we obtained $\hat{D} = 39.0 \pm 6.72$. The differences between the 2 analyses (grouped versus ungrouped) are mostly due to using $m = 2$ for ungrouped data versus $m = 1$ for grouped data, but the choice of w^* has some effect on the precision of the estimated density. Note that if $m = 2$ had been used for the grouped data (see Fig. 30b) the results would be $\hat{D} = 39.2 \pm 6.28$.

Duck Nests

Anderson and Pospahala (1970) presented line transect data on duck nests from a waterfowl production study. In 2 summers of data collection, a sample of 534 perpendicular distances was obtained. With a sample that large, it is reasonable to group the data for presentation and, in fact, it is convenient to analyze the data grouped rather than ungrouped. (Sample sizes that large are rare in line transect studies, but represent a case where grouping may be justified for ease of analysis.) Those data were grouped into 8 intervals all 1 ft (0.3048 m) long. (The data were taken in English units, hence we report our reanalysis of them in English units; however, program TRANSECT can print results in metric units if requested.) The data analysis for a 1-term FS model is presented in Fig. 31.

The log-likelihood values for 1-term and 2-term FS models are -1106.221 and -1105.539 , respectively. The 1 degree of freedom chi-square value testing whether to use $m = 1$ or 2 is 1.364 ($P = 0.2428$). Also the chi-square goodness of fit test for the 1-term model is not significant ($P = 0.628$) indicating a good fit. Thus we conclude that a 1-term FS model is suitable for analysis of the data.

From Fig. 31, $\hat{\alpha}_1$ is 0.02269 ± 0.0076 and $\hat{f}(0) = 1/8 + 0.02269 = 0.1477 \pm 0.0076$. The estimated nest density is 130.1 nests per square mile. This compares well to Anderson and Pospahala's

estimate of 127.1. To get some idea about the sampling variance of that estimate, it is assumed n is a Poisson variable, hence

$$\begin{aligned}\hat{\text{var}}(\hat{D}) &= (130.1)^2 \left[\frac{1}{543} + \left(\frac{0.0076}{0.1477} \right)^2 \right] \\ &= 76.0.\end{aligned}$$

Therefore, $\hat{s}e(\hat{D}) = 8.7$ and $\hat{cv}(\hat{D}) = 6.7$ percent. The Poisson assumption may not be true, but without a detailed analysis of the original data, one cannot derive an empirical estimator of $\text{var}(n)$. A more conservative approach would be to assume that nests are somewhat aggregated leading to $\hat{\text{var}}(n) > n$. For example, using $\hat{\text{var}}(n) = 2n$ leads to $\hat{s}e(\hat{D}) = 10.3$; this is perhaps a better indication of precision. For some parametric reanalyses of the Anderson and Pospahala (1970) duck nest data see Pollock (1978) and Gates (1979a).

Porpoise Study

In 1974, 1977, and 1979 the Southwest Fisheries Center of the National Marine Fisheries Service conducted aerial line transect studies of the porpoise stocks in the Eastern Tropical Pacific Ocean (Smith 1975, Holt and Powers In press). In particular, the 1977 effort was a large-scale study that involved 2 planes (a P2V and a PBY) and about 6 observers over a 3-month period. That study was motivated by an important problem (i.e., the incidental kill of porpoise during purse seine fishing for tuna) and is a good example of a very complex application of the line transect method. The reader interested in studying a comprehensive example of line transects should examine the report of that study (Holt and Powers In press).

The perpendicular distance data were taken grouped, without a finite w . The cut points (distance intervals), were defined before data collected, and with reference to both sides of the center line, using an interval width of 0.1 nautical miles (nm). The first interval was centered on the transect center line. Subse-

DATA TYPE	GROUPED DATA			
SAMPLE SIZE	534	LINE LENGTH	1600.	MILES
WIDTH	TRUNCATED DATA, WIDTH =		4.000	FEET
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.2269E-01	.7A05E-02	73.5	.7781E-02 .3759E-01
F(0)	.1477	.7A05E-02	9.1	.1324 .1626
D	130.1	9.757	6.7	113.0 147.3

DENSITY(D) UNITS ARE NUMBERS/SQUARE MILES

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 534. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1873E-02

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .2551E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 41.39

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	1.00	74.	.264
2	1.00	2.00	73.	.184
3	2.00	3.00	70.	.421
4	3.00	4.00	64.	.139
5	4.00	5.00	73.	.440
6	5.00	6.00	64.	.287
7	6.00	7.00	57.	.709E-01
8	7.00	8.00	56.	.396
9	8.00	9.00	54.	.162E-01

TOTAL CHI-SQUARE VALUE = 4.363 DEGREES OF FREEDOM = 6

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .62767275

FIG. 31. Analysis of the duck nest line transect of Anderson and Pospahala (1970) using a 1-term Fourier series model.

quent intervals on both sides of the line “build” on the first interval. Upon “folding over” the data (i.e., ignoring the side of the line on which a detection was made), the intervals for data analysis were defined by the cut points 0, 0.05, 0.15, 0.25 nm, etc. The first interval is only half the width of the subsequent ones. Correspondingly, in terms of the folded data, the first interval covers only half the ocean surface, as do the subsequent intervals. There is no problem in analyzing data from that scheme, which is seen to be one natural way to lay out intervals in the field where both sides of the line are considered.

Another complication of that survey is that porpoise are clustered in schools. This causes no problems in terms of estimation of the number of schools, but when an estimate of total porpoise numbers is desired, problems arise because

detection probability was related to school size. Therefore, the sample of school sizes obtained by line transect sampling is biased toward large schools. Some discussion of this general problem is given in APPENDIX D.

An example of the data collected is shown in Fig. 32. For that example we used only data on large schools (more than 15 animals in the cluster) detected from the P2V plane. With a truncation point of $w^* = 1.55$ nm, the 3-term model was chosen as appropriate for the data. The 3-term FS model fits the data very well ($\chi^2 = 2.06$, 5 df). The estimated $f(0)$ is 2.79 ± 0.24 with a coefficient of variation of 8.6 percent. Based on a sample of size 66 this is a good (reasonably precise) estimator of $f(0)$. The question of density estimation will not be pursued here, rather we refer the interested reader to Holt and Powers (1980).

DATA TYPE GROUPED DATA
 SAMPLE SIZE 66
 LINE LENGTH .1000E+04 NAUTICAL MILES
 WIDTH UNTRUNCATED DATA, UNBOUNDED WIDTH
 **NOTE - THE DATA ARE UNTRUNCATED BUT A WIDTH MUST BE CHOSEN FOR THE ESTIMATOR.
 THE LAST ORDER STATISTIC(LARGEST MEASUREMENT) WAS USED FOR THE WIDTH = 1.950

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	1.100	.1315	12.0	.8425 1.358
A(2)	.3668	.2434	66.5	-.1110 .8447
A(3)	.6921	.2924	42.9	.1063 1.256
F(0)	2.794	.2406	9.6	2.323 3.266
D	.9222E-02	.1345E-02	14.0	.6507E-02 .1194E-01

DENSITY(D) UNITS ARE NUMBERS/SQUARE NAUTICAL MILES

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2	3
1	1.000	-.745	.454
2	-.746	1.000	-.456
3	.456	-.456	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 66. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1515E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .7410E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 67.16

CHI-SQUARE GOODNESS OF FIT TEST OF THE NULL HYPOTHESIS
 THAT THE MODEL PROVIDES AN ADEQUATE FIT TO THE PERPENDICULAR DISTANCE DATA.

CUT POINT SET 1

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	.500E-01	11.	9.17
2	.500E-01	.150	14.	17.21
3	.150	.250	14.	14.12
4	.250	.350	9.	9.93
5	.350	.450	4.	9.77
6	.450	.550	4.	7.62
7	.550	.650	1.	1.04
8	.650	.750	1.	1.11
9	.750	1.55	9.	5.00

TOTAL CHI-SQUARE VALUE = 2.057 DEGREES OF FREEDOM = 5

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .84117554

FIG. 32. Example of the data collected on the 1977 aerial line transect study of porpoise. This subset of the total data comes from the P2V plane sightings of schools of 15 or more porpoise. A 3-term Fourier series was indicated as the appropriate model.

Discussion

It should be evident that rigorous analysis of grouped data is possible; however, it is not as easy as the FS analysis of ungrouped data. In particular, the lack of simple equations precludes analysis by hand calculation, and makes the use of computer programs, such as TRANSECT, mandatory.

The FS method applied to grouped data is *model robust*. Although technically the estimation method does not exactly satisfy our *pooling robust* criterion (because the estimator of $\hat{f}(0)$ is nonlinear), for practical purposes the method may also be considered *pooling robust*.

It also satisfies the *shape criterion* and is a statistically efficient method.

A natural question to raise is that of the efficiency (precision) of the FS method for grouped versus ungrouped data. To address that question would require more material on mathematical statistics than will be given here. Moreover, a problem arises because although the same model is used in both cases, the statistical estimation method differs: ML estimation for grouped data and a method of expectation (*not* ML) for ungrouped data (ML cannot be used for ungrouped data, see PART 3). That difference confounds the comparison of efficiency and precludes a simple theoretical analysis.

Even if the estimation methods were the same, some loss of efficiency would result from grouping the data. However, because the ML method is more efficient (for large samples) than the method of expectation, there may be little, if any, loss in efficiency when analyzing grouped data if the number of groups is at least 6 and the sample size is large (say, greater than 60).

ANALYSIS AND INFERENCE PROCEDURES FOR SIGHTING ANGLES AND SIGHTING DISTANCES

Much of the early line transect work concentrated on sighting distances and, rarely, sighting angles. This was partly because the importance of having a well-defined center line had not been recognized. To some extent, investigators would meander rather than follow a well-defined line; thus, the only definable measurement was the sighting distance. It is also significant that the first justifiable estimator, and one that does have a claim to robustness, was based on sighting distances (Hayne 1949). Under the Hayne model, the expected sighting angle is 32.7° . However, when angles were measured in a variety of line transect sampling, it was found (usually by a t -test of θ vs. 32.7°) that that fundamental assumption of Hayne's method was often false.

Until recently, if the average sighting angle deviated significantly from 32.7° , no robust estimator was available for use with sighting angles and distances. Gates (1969) presented a nonrobust parametric estimator for such a case. As a general approach, one can always compute the perpendicular distances from the sighting angles and sighting distances and base an estimate on those perpendicular distances. In fact, it is recommended that estimation be based on perpendicular distances. To present a comprehensive monograph on line transects, however, 2 alternatives to the Hayne estimator (i.e., estimators using sighting distances and angles) are included. It is emphasized

that those alternatives require sighting angles to be recorded in addition to sighting distances. Although Hayne's estimator can be computed from just sighting distances, the critical assumption that underlies the Hayne estimator cannot be tested without angles, and assumptions should be tested.

Estimators based on sighting distances and angles require more assumptions for their development than do methods based on perpendicular distances. The latter methods (such as the Fourier series) do not require any assumptions about the nature of the detection process. Under the general assumptions of line transect sampling, the perpendicular distance does not depend on the specific nature of the detection process. However, to model the sighting distance and angle data, understanding of, or assumptions about, the detection process are necessary. Unfortunately, we do not know much about that process; it probably is complex and involves characteristics of both the observer and the objects sought. When living animals are sampled, the behavior of the target species may also influence the detection process. Clearly, the detection process can vary drastically for different applications of line transects. The 3 methods presented below represent some approaches to estimation based on the sighting distances and angles. However, the underlying models are based neither on substantive information nor on an understanding of the detection process.

Modeling Sighting Angle and Distance Data

Two different conceptual approaches to modeling sighting angle and distance data are possible: the "fixed flushing radius" approach, and the "variable sighting distance" approach (Eberhardt 1978a). The Hayne estimator usually is selected when the fixed flushing radius model is used. (See Overton and Davis 1969 for an explanation of the derivation of the Hayne estimator by that approach.) Such

a model is not logical unless the objects being sought are animals that respond dramatically to the searching observer (e.g., grouse that flush) and their detection is a result of that response. The basic assumption of the fixed flushing radius approach is that for each animal (or group of animals) there is a fixed distance, r , such that if the observer comes closer than r units to the animal it flushes and is then observed. Thus, the sighting distance for the i^{th} animal observed will be r_i , regardless of the orientation of the animal to the line. That flushing radius model can be extended to allow a non-circular (e.g., elliptical) flushing "envelope" to exist and to determine whether or not the animal is detected (cf. Overton and Davis 1969, Burnham 1979).

Our approach to modeling sighting angle and distance data is based on the concept that sighting distance is a random variable with some unspecified probability distribution. As discussed in PART 1 (MATHEMATICAL BACKGROUND), a general model then depends only on the conditional probability distribution of perpendicular distance, x , given sighting distance, r . Let the corresponding probability density function be $f(x|r)$. The 3 methods all require the assumption that in the conditional pdf $f(x|r)$, r is a scale parameter. (This is a technical matter in mathematical statistics; no attempt is made to explain it here.) Although it is not a testable assumption, it implies that sighting angles and distances are independent random variables, which is equivalent to assuming that the $y \equiv x/r = \sin(\theta)$ and r are independent. That "scale parameter assumption" allows one to write $f(x|r)$ as $f(x|r) = (1/r)c(y)$. Here, $c(y)$ is the unknown probability density function of y . To get a computable estimator of density D , the form of $c(y)$ must be modeled; equivalently, one could model the distribution of sighting angles (but that is no easier).

For practical purposes, the assumption that sighting angles and distances are independent may be tested by examining the correlation between θ_i and r_i , $i = 1$,

..., n . That correlation (and its significance level) is computed by program TRANSECT. That assumption has been tested for all real (including field simulated) data we have obtained and have found it generally acceptable.

The real problem is, therefore, to model the distribution of $y = \sin(\theta)$. Three such models and the associated estimators are discussed below. For each model the general form of the resultant estimator of D is

$$\hat{D} = \frac{n}{2L} \hat{c}(0) \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right], \quad (2.10)$$

where $\hat{c}(0)$ is an estimator of $c(y)$ at $y = 0$. The estimate $\hat{c}(0)$ is based on the sighting angles. Note that for the estimator in Eq. (2.10),

$$\hat{f}(0) = \hat{c}(0) \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right], \quad (2.11)$$

and that a general equation for $\hat{\text{var}}(\hat{f}(0))$ is

$$\hat{\text{var}}(\hat{f}(0)) = (\hat{f}(0))^2 \left[\frac{\hat{\text{var}}(\hat{c}(0))}{(\hat{c}(0))^2} + \frac{\hat{\text{var}}(\bar{r}_h)}{(\bar{r}_h)^2} \right] \quad (2.12)$$

where,

$$\bar{r}_h = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \quad (2.13)$$

$$\hat{\text{var}}(\bar{r}_h) = \frac{\sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2}{n(n-1)}.$$

(Note that the subscript h in \bar{r}_h distinguishes it as the mean of the reciprocals of the r_i ; \bar{r}_h is related to the harmonic mean). It is emphasized that the only differences between the following 3 models are the assumptions about $c(y)$, hence only $\hat{c}(0)$ and $\hat{\text{var}}(\hat{c}(0))$ differ among the models. Finally, specific theoretical background for the results presented below is given in PART 3.

The Hayne Estimator

Assumptions

Only 1 critical assumption underlies the Hayne (1949) estimator; specifically,

it is assumed that $\sin(\theta)$ is a uniform random variable on the interval [0,1]. That assumption implies that the expected average sighting angle is 32.7° . The test statistic

$$z = \frac{\sqrt{n}(\bar{\theta} - 32.7)}{21.56}$$

is a standard normal variable under the null hypothesis that $E(\bar{\theta}) = 32.7^\circ$. The null hypothesis is rejected for large absolute values of z . For example, reject $E(\bar{\theta}) = 32.7^\circ$ at the 5 percent level if $|z| > 1.96$. If that hypothesis is rejected, the Hayne estimator should not be used, because it is not robust to failure of the fundamental assumption that $\sin(\theta)$ is a uniform random variable on [0,1].

Other, more direct tests should be based on the actual values of y_i . Specifically, a competitor to the above test is simply to test that $E(\bar{y}) = 0.5$ by using the test statistic,

$$z = \sqrt{12n}(\bar{y} - 0.5).$$

Again, reject at the 5 percent level if $|z| > 1.96$. It is necessary to perform only 1 of the 2 tests; both are presented because some people may work with only 1 of the 2 statistics, θ or $y = \sin(\theta)$.

Another test is a chi-square goodness of fit test to the uniform distribution of the y_i . Given k equal intervals on [0,1] (each of length $1/k$), let n_j be the number of y_i values falling in interval j . Note that $n = n_1 + \dots + n_k$. The expected number of observations in each interval is n/k . The chi-square test statistic is

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - (n/k))^2}{(n/k)},$$

with $k - 1$ degrees of freedom. Large values of χ^2 imply rejection of the null hypothesis that y is a uniform variable on [0,1].

More sophisticated tests can be done to compare the distribution of the y_i to the theoretical uniform distribution. Only 1 such test is computed by program TRANSECT; specifically, a special type

of data plot of the y_i is made (the sample cumulative distribution function). Under the assumption that $\sin(\theta)$ is uniform, the expected behavior of the plot is a straight line between the points (0,0) and (1,1). Consequently, such a theoretical "line" is plotted along with the plot of the sample cumulative distribution function of the y_i . A visual inspection of the actual and theoretical plots is recommended. Large deviations from the theoretical straight line, especially near values of $y = 0$ or $y = 1$, suggest that the null hypothesis of y being a uniform variable on [0,1] may be false. Also, that null hypothesis is suspect if the actual data plot lies consistently well above or below the theoretical line.

Estimation

Under the assumption of the Hayne estimator, $c(y) = 1$, $0 < y < 1$; hence $\hat{c}(0) = 1$ with $\hat{\text{var}}(\hat{c}(0)) = 0.0$. The Hayne estimator of density is

$$\hat{D}_H = \frac{n}{2L} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right]. \quad (2.15)$$

From Eqs. (2.12), (2.13), (2.14) and the general equation for $\hat{\text{var}}(\hat{D})$, Eq. (1.17),

$$\begin{aligned} \hat{\text{var}}(\hat{D}_H) &= (\hat{D}_H)^2 \\ &\cdot \left[\frac{\hat{\text{var}}(n)}{n^2} + \frac{\sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2}{(\bar{r}_h)^2 n(n-1)} \right]. \end{aligned} \quad (2.16)$$

The Generalized Hayne Estimator

Assumptions

A logical way to generalize the Hayne approach is to assume animals are sighted whenever the observer crosses an imaginary flushing curve (but not necessarily a circle) about the animal. Such a curve is meant to be representative of the animal's perception of the approaching observer, and as such it may be reasonable to assume an elliptical flushing

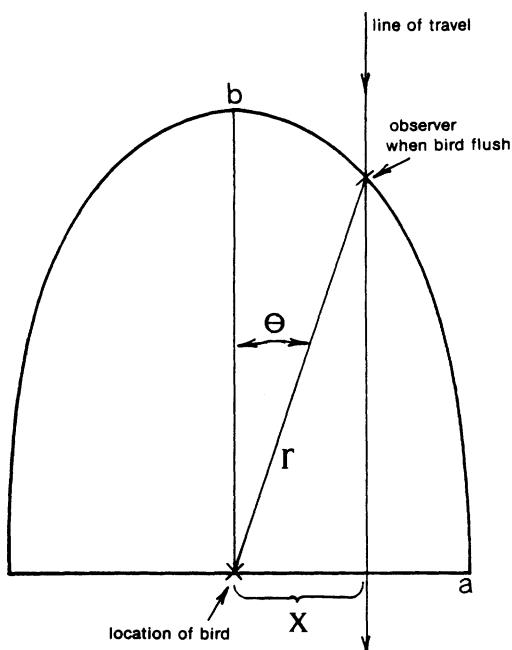


FIG. 33. Diagram of an elliptical flushing curve used in visualizing the basis for the generalized Hayne estimator.

curve with one axis parallel to the line of the observer's travel (see Fig. 33).

From Fig. 33, let a and b be the half length of the axis of the flushing ellipse. For $b > a$, the curve is extended toward the approaching observer, as if the bird would flush at greater distances if it were approached directly, rather than if the observer passed to one side or the other. Such a model has been conceptualized before in the literature (Overton and Davis 1969:422), but not developed. By assuming that the ratio b/a is fixed and equal to the unknown parameter c , but the actual values of a and b may vary among individual animals, the probability density function $c(y)$ of y can be derived:

$$c(y) = \frac{c}{(1 + (c^2 - 1)y^2)^{3/2}}, \quad 0 < y < 1. \quad (2.17)$$

Note that $c(0) = c$. For $a = b$, $c = 1$, which gives the Hayne circular model.

There is no simple test of the assumptions of the model. Moreover, before any test can be made, c must be estimated. A chi-square goodness of fit test is possible (discussed in PART 3). Program TRANSECT computes the ML estimator \hat{c} and then gives a plot of the data that may be used visually to judge the goodness of fit of this elliptic model to the observed y_i . In spirit, the plot is exactly like the one based on the sample cumulative distribution function used for testing the Hayne model, but the values on which the plot is based derive from a complex mathematical transformation of the y_i .

The generalization from Hayne's (1949) circular flushing area model to Burnham's (1979) elliptical flushing area model seems conceptually straightforward. Alternatively, the elliptical model may be viewed as a generalization of the distribution of $\sin(\theta)$. Under such a viewpoint, it can be seen that the generalized Hayne model is applicable in principle, whether or not the objects sought actually flush in response to the observer.

Estimation

Burnham (1979) has investigated the model in Eq. (2.17), including the estimation of c from the sample of values, y_1, \dots, y_n . The ML estimator, \hat{c} , does not exist in closed form. Program TRANSECT computes \hat{c} , which is equivalent to $\hat{c}(0)$ in Eq. (2.10). Let \hat{D}_{GH} represent the generalized Hayne estimator of density.

The sampling variance of the ML estimator \hat{c} does exist in simple closed form,

$$\hat{\text{var}}(\hat{c}) = \frac{5(\hat{c})^2}{4n}.$$

Thus, an estimator for $\text{var}(\hat{D}_{GH})$ can readily be obtained. Equation (1.17) should be used with $\hat{\text{var}}(\hat{f}(0))$ as given by Eq. (2.12). For the generalized Hayne estimator,

$$\hat{\text{var}}(\hat{f}(0)) = (\hat{f}(0))^2$$

$$\cdot \left[\frac{1.25}{n} + \frac{\sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2}{(\bar{r}_h)^2 n(n-1)} \right], \quad (2.18)$$

where $\hat{f}(0) = \hat{c}\bar{r}_h$. Substituting Eq. (2.18) into Eq. (1.17),

$$\hat{\text{var}}(\hat{D}_{GH}) = (\hat{D}_{GH})^2 \left[\frac{\hat{\text{var}}(n)}{n^2} + \frac{1.25}{n} + \frac{\sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2}{(\bar{r}_h)^2 n(n-1)} \right]. \quad (2.19)$$

Comparison of Eq. (2.19) with the variance of the Hayne estimator in Eq. (2.16) shows that the only difference is the added term $1.25/n$, which arises because $\hat{c}(0)$ is estimated rather than known. Thus, $\hat{\text{var}}(\hat{D}_{GH})$ is always larger than $\hat{\text{var}}(\hat{D}_H)$. Our experience with real and simulated data indicates that the increase in variance will not be large; moreover, the moderate loss of precision is accompanied by a considerable increase in robustness against the unknown distribution of sighting angles. (Note that *Example 8* under ILLUSTRATIVE EXAMPLES is based on the elliptic flushing model).

The Modified Hayne Estimator

Assumptions

An ad hoc estimator based on sighting angles and distances was given by Burnham and Anderson (1976). Although it does not have as firm a theoretical foundation as the generalized Hayne estimator discussed above, it is not based on a restricted flushing model (such as the elliptic model).

We have seen sighting angles in numerous data sets, and usually the average angle, $\bar{\theta}$, either falls within the limits 32.7° to 45° or is not significantly outside that interval. Hayne's model is based on the idea that animals flush in response to the observer and that the flushing curve

is circular. A quite different model may be applicable if detection depends on active searching, which may be the case if objects are inanimate (nonresponsive). Then the sighting angle may be a uniform random variable on the interval $[0, 90^\circ]$ rather than $\sin(\theta)$ being uniform $[0, 1]$.

If such a searching model were true, the appropriate estimator of density would be

$$\hat{D} = \frac{n}{2L} \left[\left(\frac{2}{\pi} \right) \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right],$$

which is just Hayne's estimator multiplied by $(2/\pi) = 0.63661$. If the data represent a mixture of those 2 responses, the "true" $c(0)$ may lie between 1 (Hayne) and $2/\pi$. There is an explicit probability function that represents the assumption that the sighting process is somehow a mixture of the Hayne circular flushing model (passive detection) and the searching model (active detection); it is not necessary to formalize that model, nor to make any use of it to derive the estimator given below. To derive any test of the assumption behind this model would require more rigor than seems warranted here (and it would not be a simple test). Consequently, we are inclined to consider this as a potentially very useful model; however, the estimator of c could be improved upon. The modified Hayne estimator is more robust than the Hayne estimator.

The generalized Hayne and modified Hayne estimators are not based on the same assumptions. Consequently, these two estimators do not give identical results for \hat{D} when the average sighting angle $\bar{\theta}$ differs from 32.7° .

Estimation

The modified Hayne estimator was conceived originally by Burnham and Anderson (1976) as

$$\hat{D}_{MH} = \hat{c} \frac{n}{2L} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right] = \hat{c} \hat{D}_H,$$

PD	SD	MESANG	COMANG	SINE	
5.644	9.206	40.34	41.93	.003 MEANS	
6.613	6.025	21.21	21.93	.26 STANDARD DEVIATIONS	
.5596	.7307	2.57	2.66	.032 STANDARD ERRORS OF THE MEANS	
CORRELATIONS (LOWER OFF DIAGONAL), AND THEIR TWO TAILED SIGNIFICANCE LEVELS (UPPER OFF DIAGONAL)					
PD	SD	MESANG	COMANG	SINE	
*****	0.0000	.1762	.2190	.1751	PD
.7762	*****	.9882	.7497	.9521	SD
.5131	.0018	*****	0.0000	0.0000	MESANG
.5613	.0408	.7365	*****	0.0000	COMANG
.5118	-.0074	.9852	.9328	*****	SINE

FIG. 34a. Summary statistics for the basic stake data example (Laake pers. comm.). Sample size is 68. Notation is PD=perpendicular distances, SD=sighting distances, MESANG=measured angles, COMANG=computed angles, SINE=sine of measured angles.

where \hat{c} is an adjustment factor to account for an average sighting angle greater than 32.7° (but not significantly greater than 45°). After some simplification, \hat{c} can be expressed as

$$\hat{c} = 1.9661 - 0.02954(\bar{\theta}).$$

The sampling variance of \hat{c} is

$$\text{var}(\hat{c}) = \frac{0.0008726}{n(n-1)} \sum_{i=1}^n (\theta_i - \bar{\theta})^2.$$

From the above equation and using Eqs. (1.17), and (2.12),

$$\begin{aligned} \text{var}(\hat{D}_{\text{MH}}) &= (\hat{D}_{\text{MH}})^2 \\ &\cdot \left\{ \frac{\text{var}(n)}{n^2} \right. \\ &+ \left[\frac{0.000876}{(\hat{c})^2} \frac{\sum_{i=1}^n (\theta_i - \bar{\theta})}{n(n-1)} \right] \\ &+ \left. \frac{\sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2}{(\bar{r}_h)^2 n(n-1)} \right\} \quad (2.20) \end{aligned}$$

is derived.

Examples

Stake Data

The stake example (Laake unpublished thesis) used previously to illustrate methods based on perpendicular distances is used again here. In that simulated field sampling, all 3 measure-

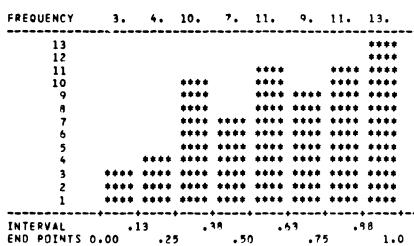


FIG. 34b. Histogram of the sighting angle sines for the stake data example.

ments were taken. Note that the true density was 37.5 stakes per ha. The FS method applied to the ungrouped perpendicular distances gave $\hat{D} = 39.0 \pm 6.7$. For grouped data the FS method produced $\hat{D} = 32.4 \pm 5.5$.

Simple summary statistics for the sample of 68 measurements are given in Fig. 34a. In particular, note that the measured angles have a mean of 40.34° . The z test of $E(\bar{\theta}) = 32.7^\circ$ is thus

$$z = \frac{\sqrt{68}(40.34 - 32.7)}{21.56} = 2.92.$$

The (2-tailed) significance level of that z value is 0.0034 ($P = 0.0034$). It is concluded that the Hayne model does not fit those data (the Hayne estimator for those data is $\hat{D} = 77.2 \pm 16.5$, which is quite poor; in fact, PRB = 106%).

The various correlations of the measured (and computed) data and the significance levels of those correlations are also shown in Fig. 34a. The only correlation of theoretical concern is that of sighting angles and distances; it is $r = 0.0018$ ($P = 0.998$). The assumption that sighting angle and distance are uncorrelated is not rejected, hence their independence is operationally acceptable.

A correlation of practical interest is that of measured and computed sighting angles. In that example, it is $r = 0.937$ leading one to believe the data were taken with reasonable care, although typically the correlation of measured and computed sighting angle was greater than 0.94 in those stake surveys (see also that same example in the section on the FS for un-

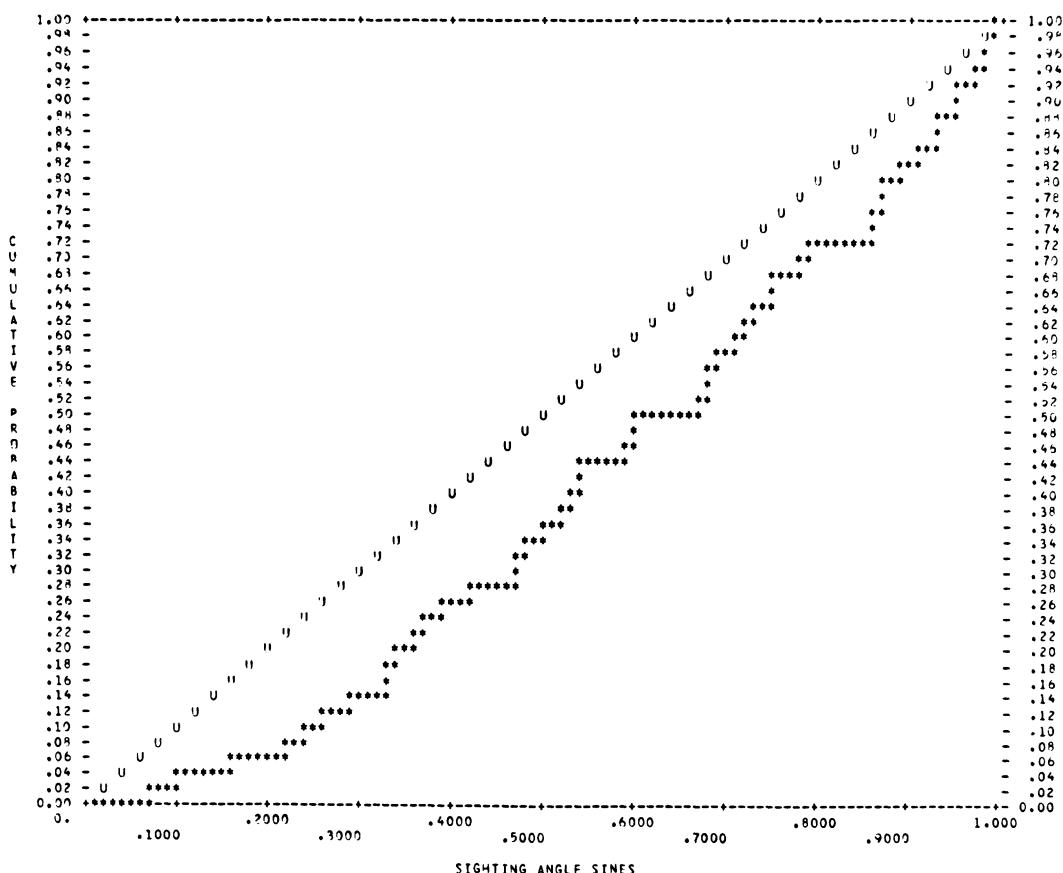


FIG. 34c. Sample cumulative distribution function of sighting angle sines for the stake data example. The straight line (U) is the theoretical plot if the Hayne model is true.

grouped data, ANALYSIS AND INFERENCE PROCEDURES FOR PERPENDICULAR DISTANCES, UNGROUPED DATA).

The other 2 tests of the Hayne model are illustrated in Figs. 34b and 34c. The frequency data and a corresponding histogram of the $\sin(\theta)$ data grouped into 8 equal intervals are shown in Fig. 34b. Under the assumption that $\sin(\theta)$ is uniformly distributed, the expected count in each interval is $(68/8) = 8.5$. The chi-square test statistic is 17.90; it has $k - 1 = 7$ degrees of freedom and is significant ($P < 0.025$). Again, the Hayne model is rejected, even though this chi-square test is not very powerful.

The sample cumulative distribution function of the $\sin(\theta)$ data and the theo-

retical plot expected if $\sin(\theta)$ is uniform are shown in Fig. 34c. The actual data lie entirely below the theoretical line (well below, compared to what would be expected if the assumption of the Hayne model were true). The Kolmogorov-Smirnov test (see Hollander and Wolfe (1973)) can be used validly to compare the observed and theoretical distributions of $\sin(\theta)$. The tests already computed in TRANSECT (e.g., test that $E(\theta) = 32.7$) are adequate for examining Hayne's model, therefore the Kolmogorov-Smirnov test in the program was not implemented.

The results of applying the generalized Hayne method to these stake data are given in Fig. 34d. The estimate of c is

DATA TYPE		UNGROUPED DATA		
SAMPLE SIZE	58	KILOMETERS		
LINE LENGTH	1.000	TRUNCATED DATA, WIDTH =	20.00	METERS
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.7123	.0966E-01	13.6	.5230 .9016
F(0)	.1617	.3504E-01	22.9	.8927E-01 .2341
D	54.98	14.23	25.9	27.10 82.86

DENSITY(D) UNITS ARE NUMBERS/HECTARE

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = λ^2 . THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1471E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .5224E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 21.97

WHEN THE PARAMETER A(1)(T,E,C) IS ESTIMATED IN A SIGHTING DISTANCE MODEL THERE IS A CORRESPONDING INCREASE IN THE VARIANCE. THERE IS A TRADE-OFF BETWEEN ROBUSTNESS AND VARIANCE. THE VARIANCE ATTRIBUTABLE TO THE ESTIMATION OF THE PARAMETER IS GIVEN BELOW.

SQUARED COEFFICIENT OF VARIATION OF A(1)(T,E,C) = .1838E-01

SQUARED COEFFICIENT OF VARIATION OF THE MEAN SD RECIPROCAL (1/S0) = .3386E-01

PERCENT OF SAMPLING VARIANCE OF F(0) ATTRIBUTABLE TO ESTIMATING A(1) 35.19

PERCENT OF SAMPLING VARIANCE OF D ATTRIBUTABLE TO ESTIMATING A(1) 27.46

FIG. 34d. The generalized Hayne estimate for the stake data example. In the output of TRANSECT, A(1) denotes the parameter c in the sighting angle, sighting distance models, while F(0) denotes f(0).

0.7123 ± 0.0966 , and of density (\hat{D}_{GH}) is 55.0 ± 14.2 . The 95 percent confidence interval on D is computed as $55.0 \pm (1.96)14.2$, or 27.1 to 82.9 (PRB = 46.7%). In that example, the sampling variance of n is based on the random distribution of stakes, so $\hat{\text{var}}(n) = 68$ is used. Fig. 34d includes information on the sensitivity of $\text{var}(\hat{D})$ to this assumption (TRANSECT routinely gives this information). Of the total sampling variance of \hat{D} in that example, only 22 percent is due to the variable nature of n, whereas 78 percent is attributed to the estimation of f(0); thus, the variance of \hat{D} is not critically dependent upon the assumed $\text{var}(n)$.

A transformation and subsequent plot of the $\sin(\theta)$ data can be made to examine the fit of the generalized Hayne model visually (see Fig. 34e). If the elliptic model for c(y) is true, the data plot will not be significantly different from a straight line (as also shown on Fig. 34e). In the example of Fig. 34e, it is judged that the elliptic model fits the data.

Figure 34f gives the results for the modified Hayne estimator applied to the same stake data. There, $\hat{c} = 0.7743 \pm 0.0760$ and $\hat{D}_{MH} = 59.8 \pm 14.4$. Thus, \hat{D}_{MH}

is not greatly different here from \hat{D}_{GH} in this example.

Hemingway's Data

Hemingway (1971) discussed the application of line transect sampling to African ungulates. Some of those data were obtained (P. Hemingway pers. comm.) and one set used as an example. The actual data ($n = 73$) are sighting angles and distances for the impala *Aepyceros melampus*, taken in October 1971. Given both measurements, one can compute perpendicular distances and apply the FS estimator, (which was done as an example with the same data in ANALYSIS AND INFERENCE PROCEDURES FOR PERPENDICULAR DISTANCES, UNGROUPED DATA). Because the impala were in groups (of size 1 to about 25), those methods produce estimates of the density of impala groups, not the density of individuals.

Simple summary statistics for those data are shown in Fig. 28a. From that information, the correlation of θ_i and r_i is 0.1522 ($P = 0.1988$); that correlation is

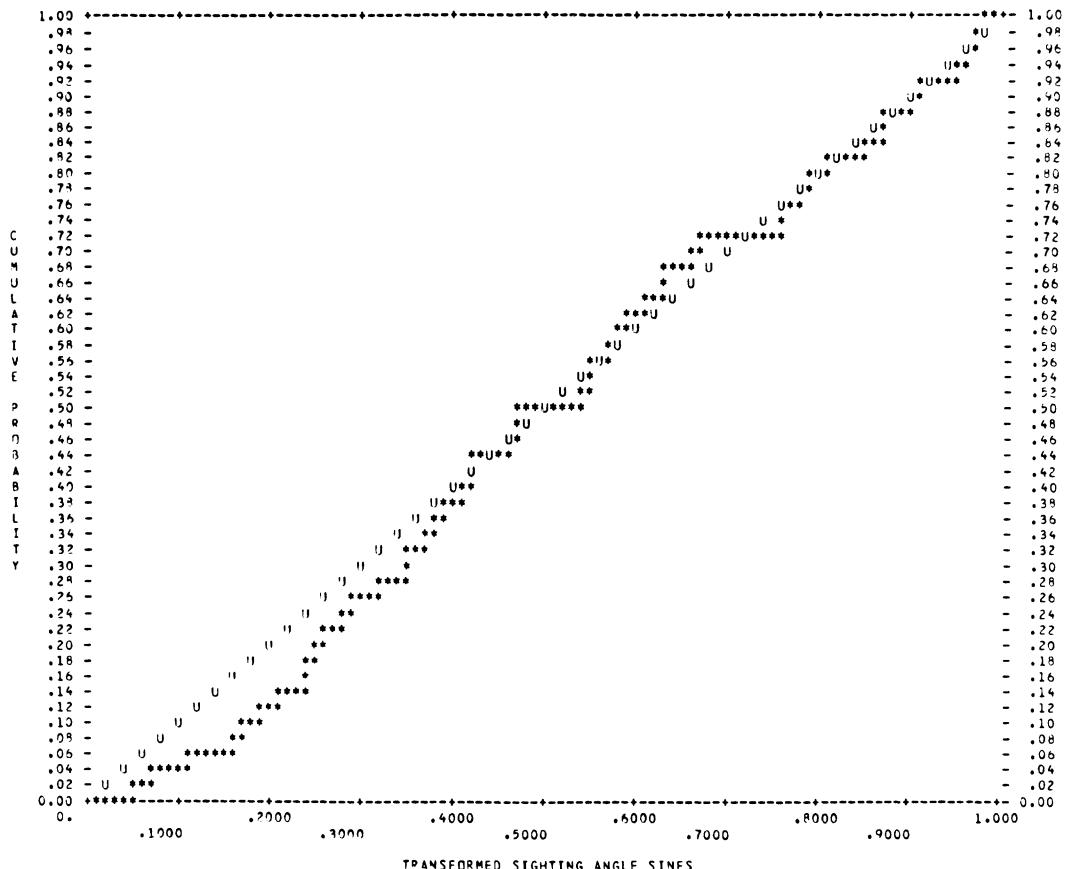


FIG. 34e. Plot of the transformed $\sin(\theta)$ data used to examine the goodness of fit of the generalized Hayne model for the stake data example.

not significant and one is therefore willing to treat θ and r as independent. Average sighting angle is 33.51; the z-test of that against 32.7 is 0.320 ($P = 0.75$). The sample cumulative distribution function of the $\sin(\theta)$ is shown in Fig. 35a. It is apparent that the fit to a uniform distribution is good. From those tests and plots it is concluded that the Hayne estimator may be used validly with the data.

The results of applying the Hayne, generalized Hayne, and modified Hayne estimators to these impala data are given in Fig. 35b. We have no empirical estimator of $\text{var}(n)$; program TRANSECT therefore uses the estimator $\text{var}(n) = n$, which is based on the assumption of a random spatial distribution of impala

groups. The sampling variances of \hat{D} in Fig. 35b are thus computed on the basis of that assumption (note that the complete output of TRANSECT warns that that has been done). From the basic output in Fig. 35b, one can recompute $\text{var}(\hat{D})$ for any estimator of $\text{var}(n)$ one wants to use.

For example, let it be assumed that $\text{var}(n) = 2n$ ($= 146$), which corresponds to some aggregation in the spatial distribution of impala groups. Table 8 shows the coefficients of variation of the density estimates for Hemingway's data under this assumption. For any estimator \hat{D} ,

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

DATA TYPE		UNGROUPED DATA			
SAMPLE SIZE	68	LINE LENGTH	1.000	KILOMETERS	
WIDTH		TRUNCATED DATA, WIDTH =	30.00	METERS	
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	.7743	.7594E-01	9.8	.6254	.9233
F(0)	.1759	.3444E-01	20.9	.1039	.2476
D	59.77	14.42	24.1	31.51	88.02

DENSITY(D) UNITS ARE NUMBERS/HECTARE

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 68. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1471E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .4348E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 25.27

WHEN THE PARAMETER A(1)(T,E+,C) IS ESTIMATED IN A SIGHTING DISTANCE MODEL THERE IS A CORRESPONDING INCREASE IN THE VARIANCE. THERE IS A TRADE-OFF BETWEEN ROBUSTNESS AND VARIANCE. THE VARIANCE ATTRIBUTABLE TO THE ESTIMATION OF THE PARAMETER IS GIVEN BELOW.

SQUARED COEFFICIENT OF VARIATION OF A(1)(T,E+,C) = .9629E-02

SQUARED COEFFICIENT OF VARIATION OF THE MEAN SD RECIPROCAL (1/S0) = .3386E-01

PERCENT OF SAMPLING VARIANCE OF F(0) ATTRIBUTABLE TO ESTIMATING A(1) 22.14

PERCENT OF SAMPLING VARIANCE OF D ATTRIBUTABLE TO ESTIMATING A(1) 16.55

FIG. 34f. The modified Hayne estimate for the stake data example. In the output of TRANSECT, A(1) denotes the parameter c in the sighting angle-sighting distance models.

For example, to compute $\hat{cv}(\hat{D}_{MH})$ in Table 8,

$$\begin{aligned} (\hat{cv}(\hat{D}_{MH}))^2 &= \frac{(2)(73)}{(73)^2} + \left[\frac{0.000581}{0.00606} \right]^2 \\ &= 0.0274 + 0.0092 = 0.0366, \end{aligned}$$

hence, $\hat{cv}(\hat{D}_{MH}) = (0.0366)^{1/2} = 0.191$.

From the information in Table 8, it can be seen that all 4 methods produce very similar results, and is to be expected when the assumptions required for the Hayne model are met. The quantities of ultimate interest are \hat{D} and its precision, summarized here through the coefficient of variation of \hat{D} . Note that $\hat{se}(\hat{D}) = \hat{D}\hat{cv}(\hat{D})$. There is remarkably little difference in the 4 coefficients of variation, but no meaningful conclusion can be drawn about the matter on the basis of 1 (or even a few) such examples. In our experience, the FS estimator and \hat{D}_H (when it is applicable) have similar standard errors, though usually that of the FS method is slightly higher (unless 1 of the sighting distances is very small, then, the Hayne estimator variance becomes very large); this is not surprising because the FS method is more robust than \hat{D}_H and the

“price” of robustness is typically some loss in precision. Similarly, on theoretical grounds, the standard error of \hat{D}_{GH} and \hat{D}_{MH} are higher than that of \hat{D}_H . It has been found, however, that \hat{D}_{MH} and \hat{D}_{GH} estimators are only slightly more variable than \hat{D}_H , but the loss in precision is again accompanied by a considerable gain in robustness.

Discussion

For a method based on sighting angles and distances, the following guidelines are provided. First, test the correlation of θ_i and r_i ; if it is significant, none of the 3 estimators discussed above should be used (convert to perpendicular distances). If that correlation is not significant, next test the assumptions of the Hayne model. If those assumptions are clearly met, use \hat{D}_H rather than \hat{D}_{GH} or \hat{D}_{MH} . If the assumptions of the Hayne model are not met, examine $\bar{\theta}$ further. If that mean sighting angle is in the interval [32.7° to 45°] (or not significantly greater than 45°), use the modified Hayne estimator. If $\bar{\theta}$ is significantly less than 32.7°

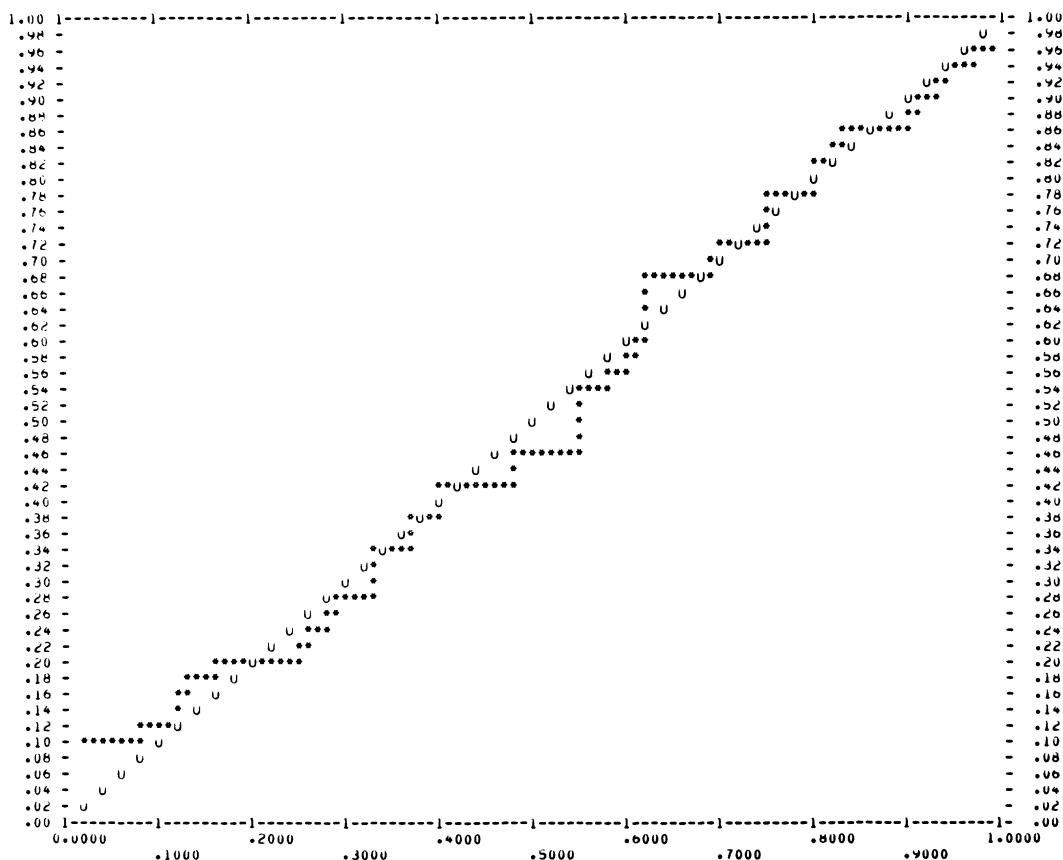


FIG. 35a. The sample cumulative distribution of sighting angle sines for Hemingway's (1971) data on impala.

or significantly greater than 45° , use the generalized Hayne method.

In PART 1, some criteria were specified that line transect estimators should satisfy: *model robustness* is foremost of those. None of those 3 estimators is truly *model robust*; all require an assumption about r being a scale parameter in $f(x|r)$. However, that assumption does not appear critical if θ and r are uncorrelated. Even then \hat{D}_H is not *model robust* because it requires $E(\bar{\theta}) = 32.7^\circ$. However, given those 2 conditions, \hat{D}_H is *pooling robust* and reasonably efficient. The other 2 estimators, \hat{D}_{GH} and \hat{D}_{MH} , are moderately *model robust* if θ and r are uncorrelated. They are also reasonably efficient, given their greater robustness. The *shape*

criterion makes a requirement of $f(x)$ but those 3 estimators assume nothing about $f(x)$; in that sense, they can therefore be said to meet the *shape criterion*. All 3 of the estimators fail to be robust to small sighting distances. One value of r_i very near to zero relative to the rest of the data will cause an enormous increase in estimated density and its variance as compared to the result without that datum.

Finally, it is emphasized that using the FS method based on the perpendicular distance data is recommended. The FS method makes no assumptions about the detection process and is applicable in cases when the 3 methods discussed above are not usable; it is therefore more robust and its efficiency is good.

HAYNE ESTIMATOR					
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
F(0)	.6204E-02	.3055E-03	4.9	.5608E-02	.6803E-02
D	.3774E-01	.4798E-02	12.7	.2835E-01	.4715E-01

GENERALIZED HAYNE ESTIMATOR					
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	.9476E-02	.1240	13.1	.7046	1.191
F(0)	.5879E-02	.8228E-03	14.0	.4268E-02	.7491E-02
D	.3577E-01	.6522E-02	18.2	.2298E-01	.4855E-01

MODIFIED HAYNE ESTIMATOR					
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	.9762	.8030E-01	8.2	.8186	1.134
F(0)	.6057E-02	.5811E-03	9.6	.4918E-02	.7196E-02
D	.3684E-01	.5576E-02	15.1	.2591E-01	.4771E-01

FIG. 35b. The results of applying the Hayne, generalized Hayne, and modified Hayne estimators to Hemingway's (1971) data on impala. Variance estimators for \hat{D} based on assuming $\text{var}(n)=n$.

ILLUSTRATIVE EXAMPLES

In this section, 9 examples to illustrate various points in the analysis of line transect data are presented. Most of those examples are from samples of artificial populations. In those examples, N (population size) and D (density) are known and the

estimated PRB can be computed. All analyses were performed by program TRANSECT and the figures are output from that program.

The examples were chosen to illustrate a number of important points. However, in an actual field study, the analysis usually will not be as simple as any one of

TABLE 8.—SUMMARY OF ESTIMATORS OF $f(0)$ AND D APPLIED TO HEMINGWAY'S DATA; UNITS OF \hat{D} ARE GROUPS OF IMPALA PER SQUARE KILOMETER. $\hat{\text{var}}(n) = 2n$ TO COMPUTE $\hat{\text{cv}}(\hat{D})$ IS ASSUMED

Estimator	$\hat{f}(0)$	$\hat{\text{cv}}(\hat{f}(0))$	\hat{D}	$\hat{\text{cv}}(\hat{D})$
FS, ungrouped data	0.00652	0.0920	3.96	0.189
Hayne	0.00620	0.0498	3.77	0.173
Generalized Hayne	0.00588	0.1400	3.58	0.217
Modified Hayne	0.00606	0.0959	3.68	0.191

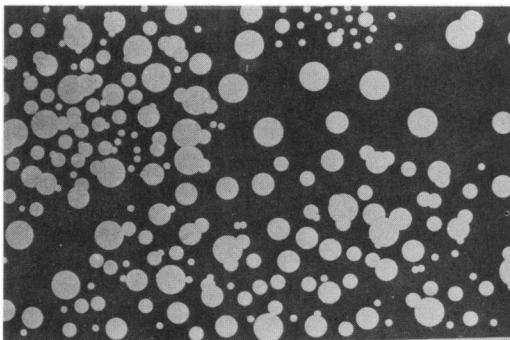


FIG. 36. An artificial population of nonrandom points and their "circular flushing radii." $N=319$ and $D=107.3 \text{ points/m}^2$.

the following 9 examples. Instead, the study design and sampling procedures should be done carefully and lead to an integrated analysis. The analysis would typically include data plotting, examination of various summary statistics, tests of assumptions, consideration of alternatives in estimating the sampling variances of $f(0)$ and D (i.e., replicate lines, jackknifing, etc.), and (possibly) computing \hat{D} for more than 1 estimation method. Often, time effects, spatial effects, ancillary data, etc., will need consideration. Examples of only the various "tools" in the material that follows are given. A comprehensive example is not attempted due to space limitations, however, such an analysis would build on the principles illustrated here.

It is hoped that biologists will study this section and follow the interpretation of the computer analysis. Many important points are illustrated (e.g., computing $\hat{\text{var}}(\hat{D})$ by using several approaches). Program TRANSECT makes many options available (particularly for the more sophisticated user) and they become more accessible as the user becomes more familiar with the technique.

Example 1, Ungrouped Perpendicular Distance Data

An artificial population of known size N and density D was established by plac-

PD	SD	MESANG	COMANG	SINE	MEANS
.1698E-01	.3223E-01	35.41	35.31	.931	MEANS
.1188E-01	.1261E-01	29.98	29.31	.297	STANDARD DEVIATIONS OF THE MEANS
.1328E-01	.1410E-02	2.68	2.61	.033	STANDARD ERRORS OF THE MEANS
CORRELATIONS (LOWER OFF DIAGONAL) AND THEIR TWO TAILED SIGNIFICANCE LEVELS (UPPER OFF DIAGONAL)					
PD	SD	MESANG	COMANG	SINE	PD
*****	.2283	0.0000	0.0000	0.0000	PD
.9229	*****	.7004	.8076	.8727	SD
.7420	-.0436	*****	0.0000	0.0000	MESANG
.7691	-.0276	.9945	*****	0.0000	COMANG
.7872	-.0182	.9816	.9853	*****	SINE

FIG. 37a. Summary statistics of data from Survey D. The mnemonic abbreviations are Perpendicular Distance (PD), Sighting Distance (SD), MEasured ANGLE (MESANG), COMputed ANGLE (COM-ANG), and COMputed SINE (COMSINE) of the angle.

ing 319 circles cut from white paper on a $1.219 \times 2.438\text{-m}$ plywood board (Fig. 36). The area was 2.973 m^2 and the density was $107.3 \text{ objects/m}^2$. The population consisted of points surrounded by circles with diameters of 2.54, 5.08, 7.62, and 10.16 cm glued to the surface of the board in a nonrandom spatial distribution. The basis for line transect sampling of the population was the "circular flushing radius" concept, and the objects of interest were assumed to be at the center of each circle.

For this table top survey, small nails were driven into the edge of the board so that a thin nylon cord could be stretched across the board in various positions to represent either random or systematic line transects. When the line transect (nylon cord) touched the perimeter of the circle, the object was detected ("flushed"); careful measurements of the perpendicular and sighting distances were made with a caliper and of sighting angles with a protractor. The sighting distances always fell into 1 of 4 classes because the circles (flushing radii) were of 4 sizes. Ten surveys were conducted on the artificial population and labeled Survey A, B, . . . , J. Sampling varied only in the placement of the lines for each survey, with other factors unchanged. Data from line transect sampling of this population are used in several of the following examples.

The data from Survey D is used to il-

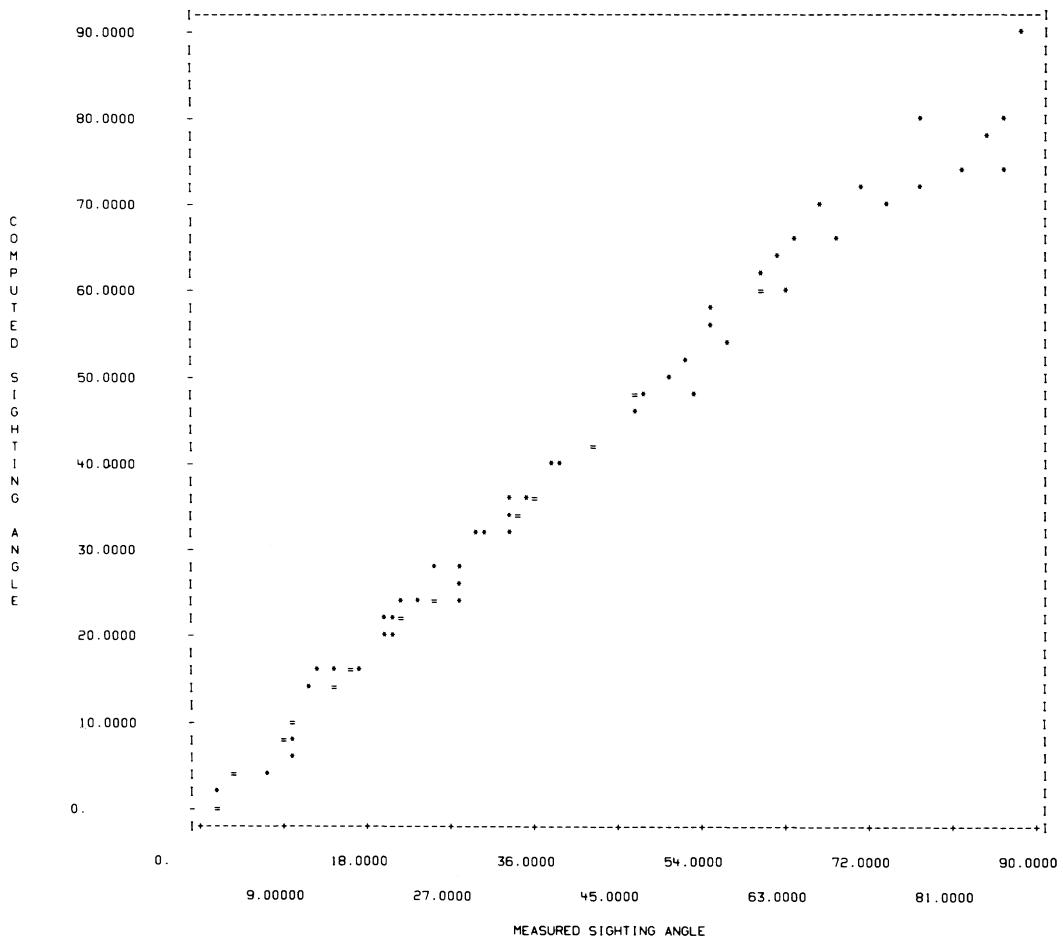


FIG. 37b. Plot of measured and computed sighting angles for Survey D. The correlation coefficient is high ($r=0.99$) and indicates accurate measurement and record keeping of the distance and angle data.

lustrate the estimation of density by the FS estimator for ungrouped perpendicular distance data. Survey D consisted of 10 lines across the width of the board in a systematic design with a random first start. Each line was 1.2192 m long for a total line length of $L = 12.192$ m. The lines were parallel and about 20.3 cm apart. On that survey, 80 objects were detected and their perpendicular distances were recorded separately for each separate line (i.e., $\ell_1, \ell_2, \dots, \ell_{10}$). The results of the analysis of those data (shown in Figs. 37a–37f) are the output from program TRANSECT. The mean, standard deviation, standard error, and

correlations for each variable are given in Fig. 37a. In that particular example, one focuses primarily on the ungrouped perpendicular distance data, although measurements were also taken of sighting distances and angles. A convenient error check can be made by plotting the measured sighting angle versus the computed sighting angle. Fig. 37b gives no indication that serious measurement or recording errors occurred; however, there were slight difficulties in measuring angles between 80 and 90°. The correlation between the computed and observed sighting angles was 0.99 (Fig. 37a). A sample histogram of the perpen-

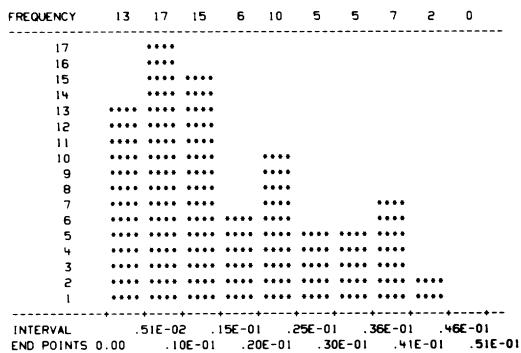


FIG. 37c. Sample histogram of the perpendicular distance data for Survey D. The 80 observations were grouped into 10 equal intervals of width 0.00508 m.

dicular distances is shown in Fig. 37c for intervals of 0.00508 m (0.508 cm). Of course, other interval sizes should be examined because the sample histograms for those small samples can change somewhat depending on the size and number of intervals chosen. Program TRANSECT allows several histograms to be prepared as an option. Ideally, parameter estimation should be done with the ungrouped data, and grouping the perpendicular distance data in the field is not

recommended; instead, it is urged that accurate data be taken in the field and investigated by means of sample histograms.

It is necessary to specify the transect width w before using the FS estimation procedure. We used 5.08 cm (0.0508 m) as it was the radius of the largest circle. In general, the selection of w is fairly easy and the estimation of density does not critically depend on this choice.

The results of program TRANSECT for the FS estimator appear in Fig. 37d. The stopping rule indicates that only 1 term of the Fourier series is required; i.e., from Eq. (2.7), $(1/0.0508)\sqrt{2/81} = 3.09 > \hat{a}_2$. The estimate of the Fourier coefficient a_1 is 15.873 ± 2.587 and $1/w = 1/0.0508 = 19.685$. Using Eq. (2.2), $\hat{f}(0) = \frac{1}{w} + \sum_{k=1}^m \hat{a}_k$, it is seen that $\hat{f}(0) = 35.558 \pm 2.587$ (i.e., $\hat{f}(0) = 19.685 + 15.873$) and

$$\begin{aligned}\hat{D} &= \frac{n\hat{f}(0)}{2L} \\ &= \frac{80 \times 35.558}{2 \times 12.192} \\ &= 116.667/m^2.\end{aligned}$$

DATA TYPE		UNGROUPED DATA		
SAMPLE SIZE	80	LINE LENGTH	12.19	METERS
WIDTH	TRUNCATED DATA, WIDTH = .5080E-01 METERS			
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	15.87	2.587	16.3	10.80 20.94
F(0)	35.56	2.587	7.3	30.49 40.63
D	116.7	15.41	13.2	81.80 151.5

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 77.8). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .1215E-01
SQUARED COEFFICIENT OF VARIATION FOR F(0) = .5295E-02
PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 69.65

FIG. 37d. Summary of the Fourier series estimation procedure for Survey D.

This estimate compares well with the actual density of 107.3 points/m², with an estimated PRB of only 8.7. The estimated size of the population is $\hat{D} \times A = 347$ (true N = 319).

An estimate of the sampling variance of \hat{D} is more difficult because the objects were deliberately placed in a nonrandom spatial distribution (Fig. 36). From Eq. (1.17), the estimator of the sampling variance of the estimated density is

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

where cv is the coefficient of variation. Equivalently, this can be written as

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

In the example, \hat{D} , n, $\hat{f}(0)$, and $\text{var}(\hat{f}(0))$ are known or have been computed; only $\text{var}(n)$ is lacking and it can be computed from the number of objects seen on each of the R individual lines ℓ_i , $i = 1, \dots, 10$. In Survey D, the numbers of objects observed on the 10 replicate lines were 6, 9, 11, 8, 13, 7, 8, 9, 6, and 3, respectively, with a mean of $\bar{n} = 8$ objects per line. The variance of n was estimated by using Eq. (1.24)

$$\begin{aligned} \text{var}(n) &= \frac{R}{R-1} \sum_{i=1}^R (n_i - \bar{n})^2 \\ &= \frac{10}{9} \times 70 \\ &= 77.778. \end{aligned}$$

That equation can also be used to estimate $\text{var}(\hat{D})$. In the example, $\hat{se}(\hat{D}) = 15.41$ and an approximate 95 percent confidence interval is 81.8–151.5. Because $\text{var}(n)$ is based on 10 replicate lines, the confidence interval is computed from the t-distribution with 9 degrees of freedom.

Program TRANSECT can compute $\text{var}(\hat{D})$ from the number of objects seen on individual replicate lines to obtain $\hat{var}(n)$. Final results appear in Fig. 37d. The single cosine term was selected by the stopping rule, and Fig. 37e indicates that a satisfactory fit has been achieved. Goodness of fit tests are possible and are

computed by program TRANSECT (see e.g., Fig. 37f).

Eq. (1.17) can be used to examine the relative contribution of the variation of n versus $\hat{f}(0)$ to the sampling variance of \hat{D} . The relevant computations are

$$\begin{array}{ll} \hat{cv}(n) = \frac{\hat{se}(n)}{n} & \hat{cv}(\hat{f}(0)) = \frac{\hat{se}(\hat{f}(0))}{\hat{f}(0)} \\ = \frac{8.82}{80} & = \frac{2.587}{35.558} \\ = 0.110 \text{ or} & = 0.073 \text{ or} \\ 11.0\% & 7.3\%. \end{array}$$

In that example, the 2 coefficients of variation are similar in magnitude, although the variation in n contributes more to the sampling variance of \hat{D} . Also of interest is the fact that the estimate of the variance of n (77.8) is similar to what it would have been if the population had been distributed randomly and the Poisson assumption had been employed ($\text{var}(n) = 80$ under the Poisson assumption).

Example 1 represents the ideal situation in many respects. Much can be done if accurate measurements of the perpendicular distance and sighting distance and angle are made for each observation. The sample size (n = 80) was quite adequate. It is realized that many factors will prohibit such data from surveys in the field. However, the field procedures and survey design guidelines presented in PART 1 should be followed as closely as possible.

If the design of the survey and the collection of data have been done properly, the analysis using program TRANSECT is relatively straightforward. Although caution must be exercised in selecting the scale of the various measurements (i.e., L, x_i , and r_i), use of program TRANSECT simplifies the task. The remaining examples explore a variety of other situations that commonly arise.

Example 2, Analysis of Sighting Angles and Distances

This example also uses the data collected from Survey D, and illustrates the

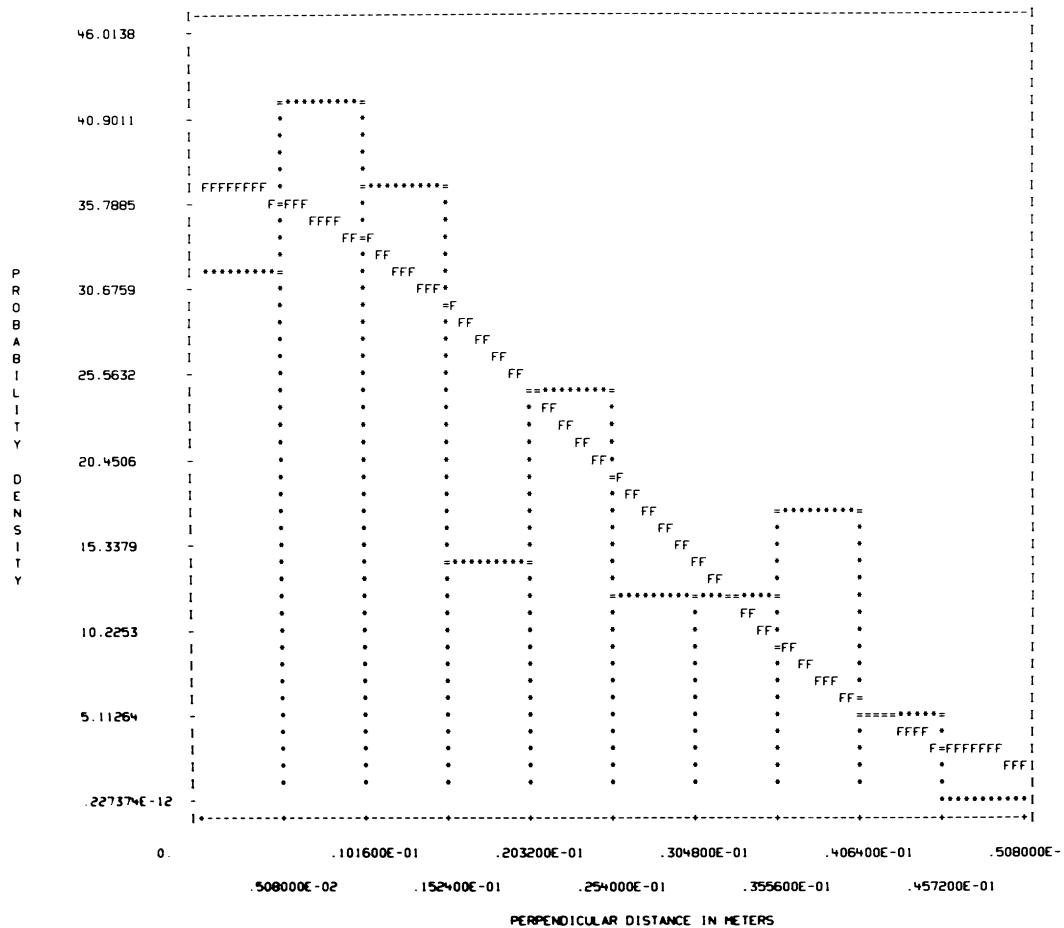


FIG. 37e. The 1-term Fourier series pdf plotted against a sample histogram of the data from Survey D. Note, it is the function at 0 distance (the intercept) that is crucial in estimating density.

use of the sighting distance and angle data. Because Hayne's (1949) method is based on the concept of a circular flushing radius, it is not surprising that Hayne's estimator performs well on this example. First, 2 assumptions that are absolutely crucial to Hayne's method are examined: the expected (average) angle is $= 32.7^\circ$ and $\sin(\theta)$ is uniformly distributed in the interval $[0,1]$. If $\sin(\theta)$ is uniform $[0,1]$, then $E(\theta) = 32.7^\circ$ is assured, but the converse is not true. Hayne's estimator is very sensitive to departures from those assumptions. One commonly finds that neither assumption is met with real data. The results of the analysis appear in Figs. 38a–38d.

The average measured sighting angle in this example is $35.81^\circ \pm 2.68$ (Fig. 37a) which does not differ significantly from

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	.508E-02	13	14.86
2	.508E-02	.102E-01	17	14.19
3	.102E-01	.152E-01	15	.337
4	.152E-01	.203E-01	6	11.15
5	.203E-01	.254E-01	10	9.09
6	.254E-01	.305E-01	5	6.91
7	.305E-01	.356E-01	5	4.85
8	.356E-01	.406E-01	7	3.09
9	.406E-01	.457E-01	2	1.81
10	.457E-01	.508E-01	0	1.14

TOTAL CHI-SQUARE VALUE = 10.25 DEGREES OF FREEDOM = 8

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .2478617

FIG. 37f. Chi-square goodness of fit test of the Fourier series based on 10 equal intervals. The test indicates a good fit ($P=0.25$).

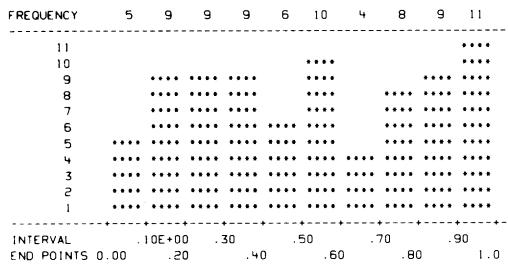


FIG. 38a. Sample histogram of the sine of the sighting angles for Survey D. $\sin(\theta)$ must be uniform $[0,1]$ for Hayne's method to be useful.

32.7° ($z = 1.29$, $P = 0.20$). Generally, and in Fig. 38a, which is a sample histogram of $\sin(\theta)$, no marked departure from the uniform distribution is apparent. This important assumption is tested further and the results from program TRANSECT are shown in Fig. 38b. From those tests and a plot of the sample cumulative distribution function (cdf) versus the uniform cdf (Fig. 38c), one sees no reason to reject the assumption that the sine of the sighting angle is uniform.

Having thoroughly investigated the underlying assumptions, one can proceed to the results, using Hayne's estimator (Fig. 38d). The parameter $f(0)$ is estimated to be 37.40. The estimator of the sampling variance of $f(0)$ is (Eq. 1.12)

$$\hat{v}ar(\hat{f}(0)) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2$$

$$= 4.24$$

and $\hat{s}e(\hat{f}(0)) = \sqrt{4.24} = 2.06$ (Fig. 38d). The estimate of D (from Eq. 2.1) is

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

$$= \frac{80 \times 37.40}{2 \times 12.192}$$

$$= 122.7 \text{ objects/m}^2.$$

Because one now has \hat{D} , $\hat{cv}(n)$ (from Example 1), $\hat{f}(0)$ and $\hat{v}ar(\hat{f}(0))$, $\hat{v}ar(\hat{D})$ by using Eq. (1.17) can be computed. First one computes

$$\hat{cv}(\hat{f}(0)) = \frac{\hat{s}e(\hat{f}(0))}{\hat{f}(0)}$$

$$= \frac{2.06}{37.40}$$

$$= 0.05508.$$

Then,

$$\hat{v}ar(\hat{D}) = (\hat{D})^2 \{ (\hat{cv}(n))^2 + [\hat{cv}(\hat{f}(0))]^2 \}$$

$$= 122.7^2 [0.1102^2 + 0.05508^2]$$

$$= 228.6$$

$$\hat{s}e(\hat{D}) = \sqrt{228.6} = 15.1.$$

Again, the variation associated with n

DATA TYPE UNGROUPED DATA
SAMPLE SIZE 80
LINE LENGTH 12.19 METERS
WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

ORDERED SIGHTING ANGLE STYLES

.03	.03	.03	.07	.07	.12	.15	.15	.16	.16	.17	.17	.17	.17	.17	.21	.22	.24	.24	.24	.26
.28	.29	.29	.34	.34	.35	.35	.37	.37	.37	.37	.39	.42	.42	.42	.47	.47	.47	.50	.52	
.54	.54	.54	.56	.56	.57	.59	.59	.62	.63	.67	.67	.73	.73	.73	.74	.77	.79	.80		
.92	.82	.84	.87	.87	.87	.89	.89	.90	.92	.93	.95	.96	.97	.97	.99	1.00	1.00	1.00	1.00	

SAMPLE MEAN = 0.5310 SAMPLE VARIANCE = 0.088505
Z-TEST VALUE THAT THIS AVERAGE = 0.5 IS 0.9619 SIGNIFICANCE LEVEL = 0.68

FIG. 38b. Test that $\sin(\theta)$ is uniform for the data collected on Survey D.

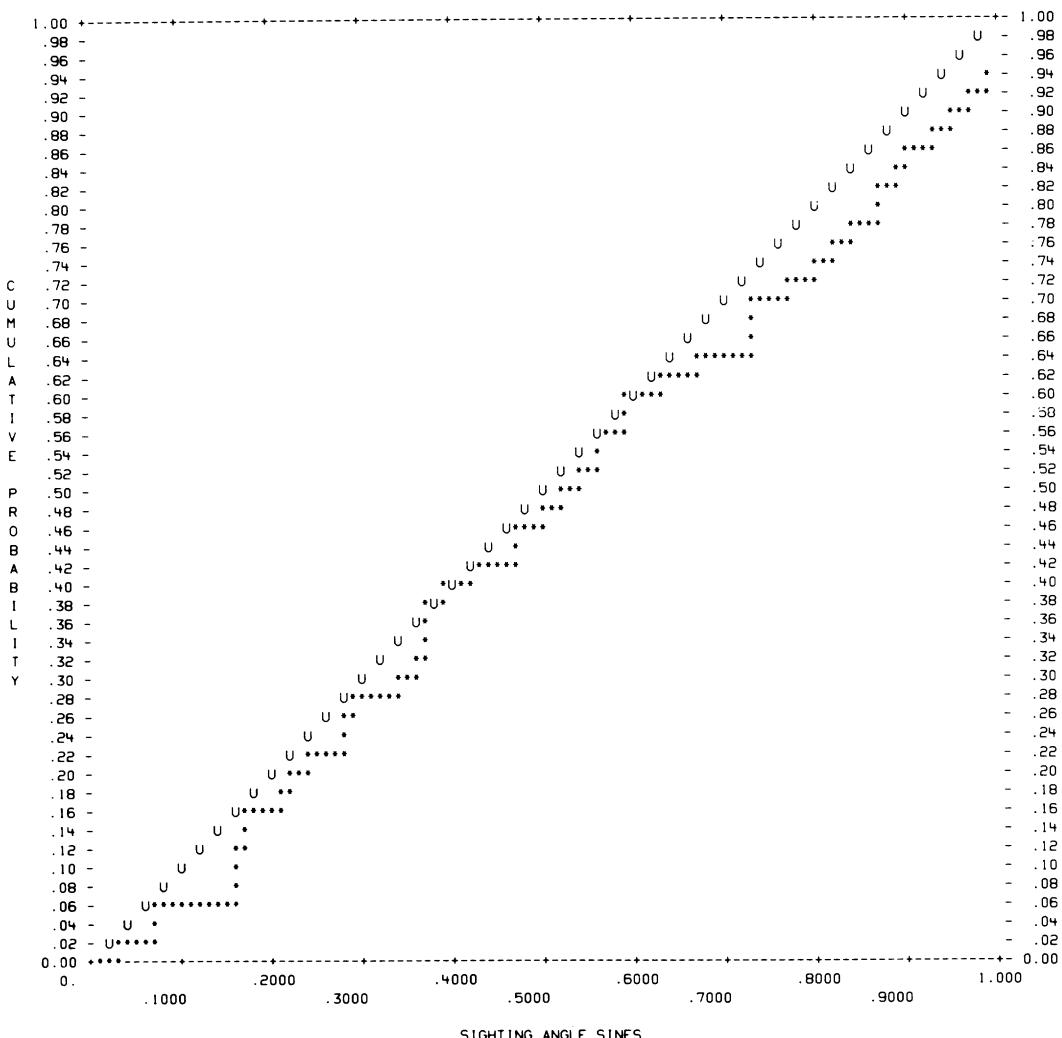


FIG. 38c. A plot of the sample cumulative distribution function (cdf) of the sine of the sighting angle versus the cdf for the uniform distribution.

dominates the expression for $\hat{\text{var}}(\hat{D})$. In summary, using Hayne's estimation method, density is estimated at 122.7 ± 15.1 objects/m² and $\hat{N} = 364.8$ objects in the population. It is interesting to note that $\hat{\text{se}}(\hat{D})$ for the Fourier series (Example 1) is only 15.4 and the estimated PRB is 14.4. The results obtained by using the estimators compare fairly well with the true population size and density. The 95 percent confidence interval for \hat{D} is 88.5–156.9 and the true density lies well within

in that interval. Furthermore, if $\text{var}(\hat{D})$ is estimated from 10 replicate lines by using Hayne's method (Eq. 1.22), $\hat{\text{se}}(\hat{D}) = 15.1$. The approaches to estimating $\text{var}(\hat{D})$ compare well in this example.

Finally, it is noted that the precision for the Hayne estimator, which makes fairly restrictive assumptions and is non-robust to failure of the assumptions, is very similar to the general FS procedure, which makes no such assumptions and is a general, robust procedure. The various

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 80
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

PARAMETER	POINT	STANDARD	PERCENT COEFF.	95 PERCENT
	ESTIMATE	ERROR	OF VARIATION	CONFIDENCE INTERVAL
F(0)	37.40	2.060	5.5	33.36 41.44
D	122.7	15.12	12.3	88.50 156.9

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 77.8). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .1215E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .3033E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 80.03

FIG. 38d. Summary of the Hayne (1949) estimation procedure for Survey D.

Hayne type estimators are quite sensitive to very small sighting distances. For those reasons and others, estimation based on sighting distances and angles is not recommended. If only r_i and θ_i are taken in the field, converting to the perpendicular distances x_i and basing inference on those is recommended.

The Hayne estimator as modified by Burnham and Anderson (1976:334) represents 1 of 2 approaches to deal with cases for which the average sighting angle differs from 32.7° . The modified Hayne estimator is useful when the average sighting angle lies between 32.7 and 45° . The modification consists of estimating a constant c where a simple linear interpolation is used,

$$\delta = (\bar{\theta} - 32.7)/(45 - 32.7)$$

and

$$\hat{c} = (1 - \delta) + \delta(2/\pi).$$

A simpler computational formula is given in PART 2 as $\hat{c} = 1.966 - 0.02954\bar{\theta}$. The estimate of the constant c is multiplied by the usual Hayne estimator. For example,

$$\hat{f}(0) = \hat{c} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \right)$$

or

$$= \hat{c}\bar{r}_h.$$

Then,

$$\begin{aligned} \hat{v}\text{ar}(\hat{f}(0)) &= (\hat{f}(0))^2 [(\hat{c}v(\hat{c}))^2 + (\hat{c}v(\bar{r}_h))^2] \\ &= (3.5023)^2 \end{aligned}$$

where

$$\begin{aligned} (\hat{c}v(\hat{c}))^2 &= \frac{\hat{v}\text{ar}(\hat{c})}{(\hat{c})^2} \\ &= 0.007602 \end{aligned}$$

$$\begin{aligned} \hat{v}\text{ar}(\hat{c}) &= (0.02954)^2 \hat{v}\text{ar}(\bar{\theta}) \\ &= (0.07917)^2 \end{aligned}$$

$$\begin{aligned} \hat{v}\text{ar}(\bar{\theta}) &= \frac{1}{n(n-1)} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 \\ &= (2.68)^2 \end{aligned}$$

and

$$\begin{aligned} (\hat{c}v(\bar{r}_h))^2 &= \frac{\hat{v}\text{ar}(\bar{r}_h)}{(\bar{r}_h)^2} \\ &= 0.003038 \end{aligned}$$

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 80
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

PARAMETER	POINT	STANDARD	PERCENT COEFF.	95 PERCENT
	ESTIMATE	ERROR	OF VARIATION	CONFIDENCE INTERVAL
A(1)	.9080	.7920E-01	8.7	.7528 1.063
F(0)	33.96	3.503	10.3	27.10 40.83
D	111.4	16.82	15.1	73.37 149.5

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 77.8). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .1215E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1064E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 53.32

WHEN THE PARAMETER A(1)(I.E.,C) IS ESTIMATED IN A SIGHTING DISTANCE MODEL THERE IS A CORRESPONDING INCREASE IN THE VARIANCE. THERE IS A TRADE-OFF BETWEEN ROBUSTNESS AND VARIANCE. THE VARIANCE ATTRIBUTABLE TO THE ESTIMATION OF THE PARAMETER IS GIVEN BELOW.

SQUARED COEFFICIENT OF VARIATION OF A(1)(I.E.,C) = .7607E-02
 SQUARED COEFFICIENT OF VARIATION OF THE MEAN SD RECIPROCAL (1/SD) = .3033E-02

PERCENT OF SAMPLING VARIANCE OF F(0) ATTRIBUTABLE TO ESTIMATING A(1) 71.49

PERCENT OF SAMPLING VARIANCE OF D ATTRIBUTABLE TO ESTIMATING A(1) 33.38

FIG. 39. Summary of the modified Hayne estimator for the data from Survey D.

$$\hat{\text{var}}(\bar{r}_h) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{1}{r_i} - \bar{r}_h \right)^2 \\ = (2.06)^2.$$

The results of such an approach are shown in Fig. 39 for Survey D. The value of c is estimated to be 0.908 ± 0.0792 . Density is estimated to be 111.4 objects/ m^2 and the estimated population size ($\hat{N} = \hat{D} \times A$) is 331.3, with an estimated PRB of 3.8.

If $\bar{\theta} = 32.7^\circ$, the value of c = 1 and the usual Hayne estimator results. In that example, the modified Hayne estimator was less biased and the result can be expected when $\bar{\theta} > 32.7^\circ$. The modified Hayne estimator of density has a larger standard error (16.82) than the usual Hayne estimator (15.12). Because the modified Hayne estimator is a generalization of Hayne's estimator and includes the special case $\bar{\theta} = 32.7^\circ$, its use is suggested as a general procedure for the analysis of

data on sighting angles and distances. Note, however, that the FS estimator based on the perpendicular distance data also had a small PRB, a standard error of \hat{D} of only 15.41, and makes no restrictive assumptions.

The generalized Hayne estimator (Burnham 1979) is a second approach for the analysis of data on sighting angles and distances. It is based on the concept of elliptical flushing envelopes and also includes the usual Hayne estimator as a special case. Information shown in Fig. 40a provides a visual test to judge the appropriateness of this procedure. As has already been seen, the average sighting angle is not different from 32.7° in this example. The other tests are also not significant.

The results of this analysis scheme for Survey D are shown in Fig. 40b and appear quite good. Density is estimated to be $106.0/\text{m}^2$, giving $\hat{N} = 315$. Estimators

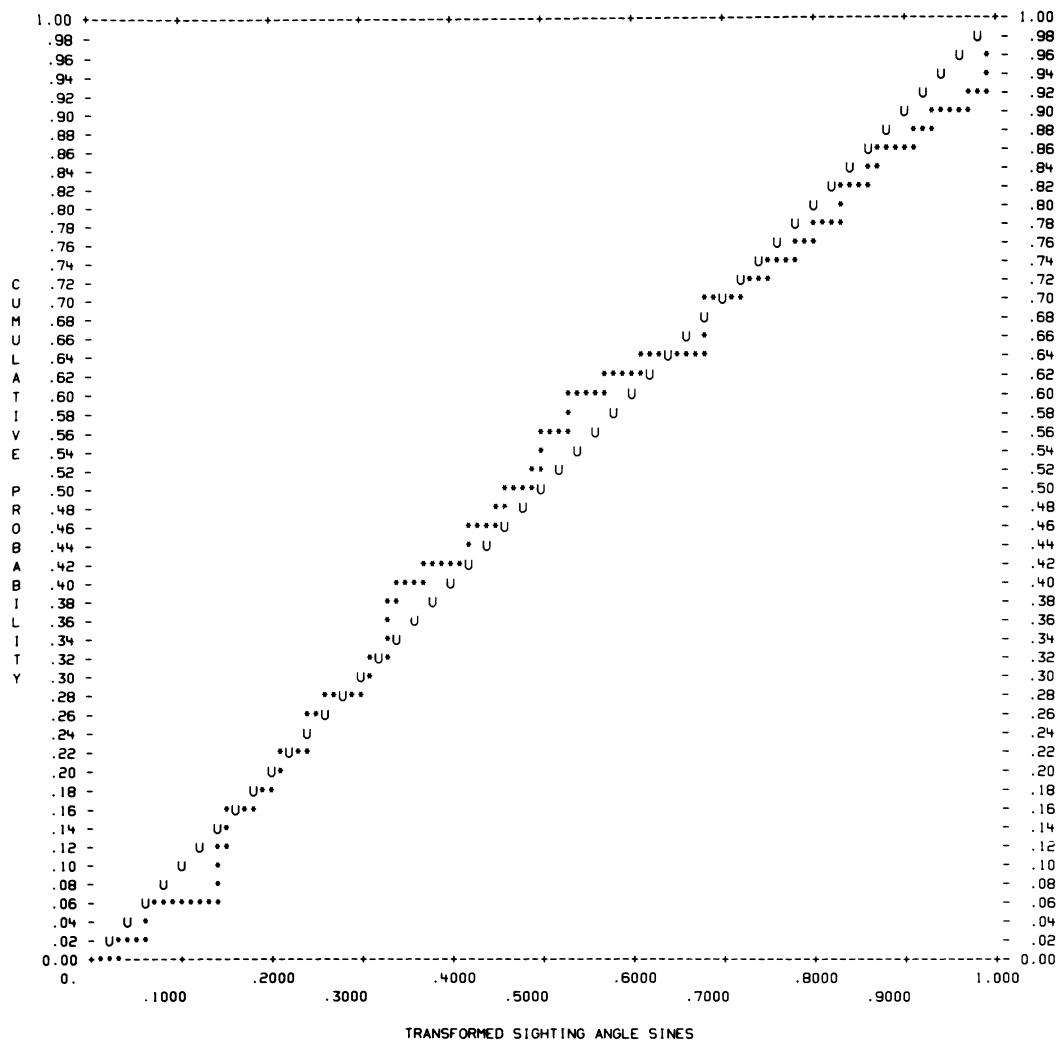


FIG. 40a. A plot of the sample cumulative distribution function (cdf) for probability integral transforms of the sine of the sighting angles versus the cdf for the uniform distribution. Data were taken as part of Survey D.

for $\hat{\text{var}}(\hat{f}(0))$ and $\hat{\text{var}}(\hat{D})$ are the same as for the modified Hayne estimator except that

$$\hat{\text{var}}(\hat{c}) = \frac{5}{4} \frac{(\hat{c})^2}{n}.$$

The estimated relative bias with the generalized Hayne method is only -1 percent, however, the $\hat{s}e(\hat{D})$ is larger at 18.6, giving a wider 95 percent of confidence interval, 63.9–148.0. In general, $se(\hat{D}_H) < se(\hat{D}_{MH}) < se(\hat{D}_{GH})$. In that example, the generalized Hayne method has a relative

efficiency of 82 percent relative to the modified Hayne method. Furthermore, the generalized Hayne method has an efficiency of only 64 percent relative to the FS or Hayne methods.

Several summary comments concerning these 3 estimators are offered for the analysis of sighting angles and distances. Taking ungrouped perpendicular distance data and performing the analysis on such data is recommended. If only sighting angles and distances are taken, it is

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 80
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.8635	.1079	12.5	.6519 1.075
F(0)	32.30	4.411	13.7	23.65 40.94
D	106.0	18.60	17.6	63.89 148.0

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 77.8). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .1215E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1866E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 39.44

WHEN THE PARAMETER A(1)(I.E.,C) IS ESTIMATED IN A SIGHTING DISTANCE MODEL THERE IS A CORRESPONDING INCREASE IN THE VARIANCE. THERE IS A TRADE-OFF BETWEEN ROBUSTNESS AND VARIANCE. THE VARIANCE ATTRIBUTABLE TO THE ESTIMATION OF THE PARAMETER IS GIVEN BELOW.

SQUARED COEFFICIENT OF VARIATION OF A(1)(I.E.,C) = .1563E-01

SQUARED COEFFICIENT OF VARIATION OF THE MEAN SD RECIPROCAL (1/SD) = .3033E-02

PERCENT OF SAMPLING VARIANCE OF F(0) ATTRIBUTABLE TO ESTIMATING A(1) 83.74

PERCENT OF SAMPLING VARIANCE OF D ATTRIBUTABLE TO ESTIMATING A(1) 50.71

FIG. 40b. Summary of the generalized Hayne estimator for the data from Survey D.

recommended that they be converted to perpendicular distances (i.e., $x_i = r_i \cdot \sin(\theta_i)$) and the analysis be based on those perpendicular distances. If the sighting angle and distance data must be analyzed, either the modified or the generalized Hayne estimation methods are recommended. All 3 estimators and the various test statistics are available in program TRANSECT as options.

Example 3, Grouped Perpendicular Distance Data

This example is also based on data from Survey D, but the perpendicular distance data are grouped arbitrarily into 5 intervals, each of width 1.02 cm. For the purpose of illustration, assume that the data were necessarily taken in the field only by intervals or that, for some reason, the data must be analyzed as grouped data (e.g., see *Example 5* below). The FS estimator for grouped data was the basis for estimation; a sample

histogram is shown in Fig. 41a and the results are summarized in Fig. 41b. The simple chi-square goodness of fit test is not significant, indicating that the model

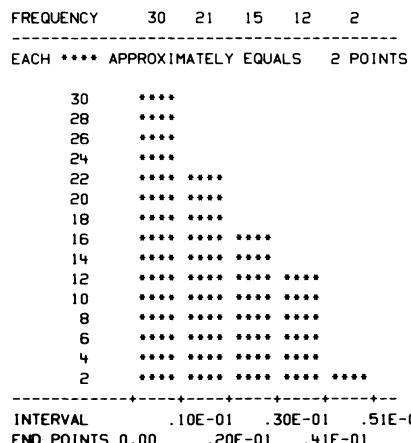


FIG. 41a. Sample histogram of the perpendicular distance data for Survey D. The 80 observations were grouped into 5 equal intervals of width 0.0102 m.

DATA TYPE GROUPED DATA
 SAMPLE SIZE 80
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

PARAMETER	POINT	STANDARD	PERCENT COEFF.	95 PERCENT
	ESTIMATE	ERROR	OF VARIATION	CONFIDENCE INTERVAL
A(1)	17.01	2.019	11.9	13.05 20.97
F(0)	36.69	2.019	5.5	32.73 40.65
D	120.4	14.83	12.3	86.83 153.9

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 77.8). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .1215E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .3029E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 80.05

FIG. 41b. Summary of the Fourier series estimation procedure for Survey D. The analysis is based on data grouped into 5 equal intervals.

fits the data ($\chi^2 = 2.84$, $P = 0.42$). Examination of the log-likelihood values indicates that a 1-term cosine series is adequate for the data. The values of the log-likelihood function at its maximum are -114.09 and -113.86 for 1- and 2-term FS models, respectively. The quantity

$-2(\ln \mathcal{L}_1 - \ln \mathcal{L}_2)$ used as a test of the adequacy of the 1-term model is asymptotically distributed as a chi-square variable. The value $-2[-114.09 - (-113.86)] = 0.46$ is not significant; it is concluded that a model with 1 term is sufficient. From the information in Fig.

DATA TYPE GROUPED DATA
 SAMPLE SIZE 80
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

PARAMETER	POINT	STANDARD	PERCENT COEFF.	95 PERCENT
	ESTIMATE	ERROR	OF VARIATION	CONFIDENCE INTERVAL
A(1)	17.17	1.903	11.1	13.43 20.90
F(0)	36.85	1.903	5.2	33.12 40.58
D	120.9	14.72	12.2	87.61 154.2

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 77.8). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .1215E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .2668E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 82.00

FIG. 41c. Summary of the Fourier series estimation procedure for Survey D. The analysis is based on data grouped into 10 equal intervals.

TABLE 9.—SUMMARY OF ESTIMATES FROM LINE TRANSECT DATA FOR SURVEY D USING THE FOURIER SERIES ESTIMATOR FOR GROUPED DATA

Number of groups	$\hat{f}(0)$	$\hat{s}e(\hat{f}(0))$	$\hat{cv}(\hat{f}(0)) \times 100$	\hat{D}	$\hat{s}e(\hat{D})$	$\hat{cv}(\hat{D}) \times 100$	Percentage loss of efficiency ¹
10	36.85	1.903	5.2	120.9	14.72	12.2	—
5	36.69	2.020	5.5	120.4	14.83	12.3	98.5
3	34.85	2.507	7.2	114.3	15.05	13.2	95.7
2	35.92	2.942	8.2	117.8	16.18	13.7	82.8

¹ Compared to the analysis based on 10 groups, i.e., the ratio of the sampling variances.

41b, it is apparent that $\hat{f}(0) = 36.69 \pm 2.02$; this gives an estimate of density of $120.4/m^2$ or $\hat{N} = 358$, similar to the FS estimator for ungrouped perpendicular distance data (see *Example 1*).

The analysis of grouped data poses no real problems other than that a complicated numerical procedure is required to obtain the exact ML estimates, making hand calculation virtually impossible. In practice, 2 disadvantages are encountered. First, grouping of the data is arbitrary. To some extent $\hat{f}(0)$ and, therefore, \hat{D} depend on the group size and the number of groups. That is especially true of small samples, say $n = 25-40$ or when only 2–3 groups are used. One is then faced with decisions, for example, as to whether the groups should be of equal or unequal size. Second, some loss of estimation efficiency (precision) occurs when the data are grouped into distance intervals; the larger the groups, the larger will be the $\text{var}(\hat{f}(0))$, for example.

As a second illustration, the same data set was considered but was grouped into 10 equal intervals of 0.508 cm; (Fig. 37c). The results of the estimation procedure appear in Fig. 41c: $\hat{f}(0) = 36.85 \pm 1.90$ for a 1-term cosine series. Those data were also analyzed with 2 and 3 groups and the results are summarized in Table 9. In all 4 cases, perpendicular distance data from Survey D were analyzed by using a 1-term cosine series. The loss of precision due to grouping can be substantial and is as much as 17 percent for this example.

Although the analysis of grouped data is not recommended (if one has a choice),

certainly instances occur for which such analyses are appropriate. Further examples will bear this out.

Example 4, Transects of Unequal Length

The design of line transect surveys often involves lines of unequal length. This may arise because the line lengths themselves are randomly chosen or because the transects span the width of the area but the area to be surveyed is irregularly shaped. Use of perpendicular distance data from Survey O illustrates that unequal line lengths pose no obstacle in the analysis.

Survey O is identical to Survey D except that 14 lines of unequal length were used and run diagonally across the board.

TABLE 10.—SUMMARY OF LINE LENGTHS AND NUMBER OF OBJECTS DETECTED ON EACH LINE FOR SURVEY O

Line	$\ell_i(\text{cm})$	n_i
1	149.86	14
2	149.86	9
3	149.86	10
4	149.86	9
5	149.86	7
6	149.86	10
7	149.86	9
8	143.51	6
9	71.12	6
10	86.36	4
11	21.59	1
12	116.84	8
13	76.20	2
14	64.77	1
Total	$L = 1629.4$	$n = 96$

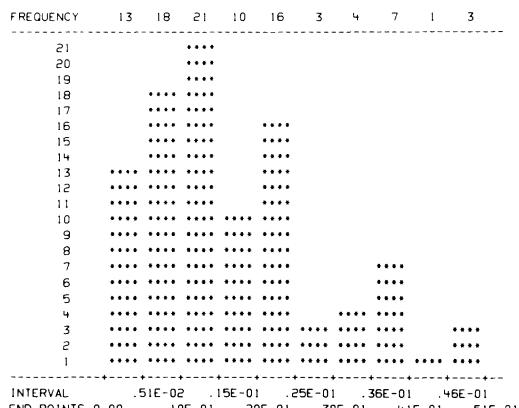


FIG. 42a. Histogram of perpendicular distance data for Survey O.

The lines ranged in length from 21.6 to 149.9 cm and the number of objects detected on these lines ranged from 1 to 14 (Table 10). Program TRANSECT provided the analysis (Figs. 42a-42d). A histogram of the distance data is shown in Fig. 42a and the FS analysis based on the ungrouped data appears in Fig. 42b. In this example, $\hat{v}ar(n)$ is needed to compute $\hat{v}ar(\hat{D})$. The variance of n was computed from the number of objects observed on the R replicate lines of unequal length. From Eq. (1.23),

$$\hat{v}ar(n) = \left(\frac{L}{R-1} \right) \sum_{i=1}^R \ell_i \left(\frac{n_i}{\ell_i} - \frac{n}{L} \right)^2 \\ = 67.58$$

where $R = 14$ in this example.

Thus, $\hat{D} = 107.7/m^2 \pm 11.5$, and gives $\hat{N} = 320$ and an estimated PRB of 0.4. Those data illustrate a second point relating to the "fit" of the FS model to the data. The stopping rule, based on the mean integrated squared error (MISE), indicates that a 1-term FS is adequate for the data. However, information in Figs. 42c and 42d suggest otherwise. The fit of the cosine function versus a histogram of the observed data (Fig. 42c) suggests a poor fit in the fifth and sixth intervals. That poor fit is further substantiated by the chi-square test shown in Fig. 42d. Note that if intervals 1 and 2, 3 and 4, 5 and 6, 7 and 8, and 9 and 10 were combined, a χ^2 value of 0.26 with 3 degrees of freedom would result, indicating a very good fit. Two concepts are important to an understanding of this example. The first is that judging the "fit" by examining histograms of grouped data is difficult because of the arbitrary nature of the class intervals. A different grouping of the same data could produce somewhat different results. Regrouping of data often

DATA TYPE		UNGROUPED DATA		
SAMPLE SIZE	96			
LINE LENGTH	16.29	METERS		
WIDTH	TRUNCATED DATA, WIDTH =	.5080E-01 METERS		
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	16.89	2.320	13.7	12.34 21.43
F(0)	36.57	2.320	6.3	32.02 41.12
D	107.7	11.48	10.7	82.93 132.5

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS ($\hat{v}ar(n) = 67.58$). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 13 DEGREES OF FREEDOM IS 2.160

SQUARED COEFFICIENT OF VARIATION FOR N = .7334E-02

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .4025E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 64.57

FIG. 42b. Summary of results of the Fourier series estimator for ungrouped distance data from Survey O.

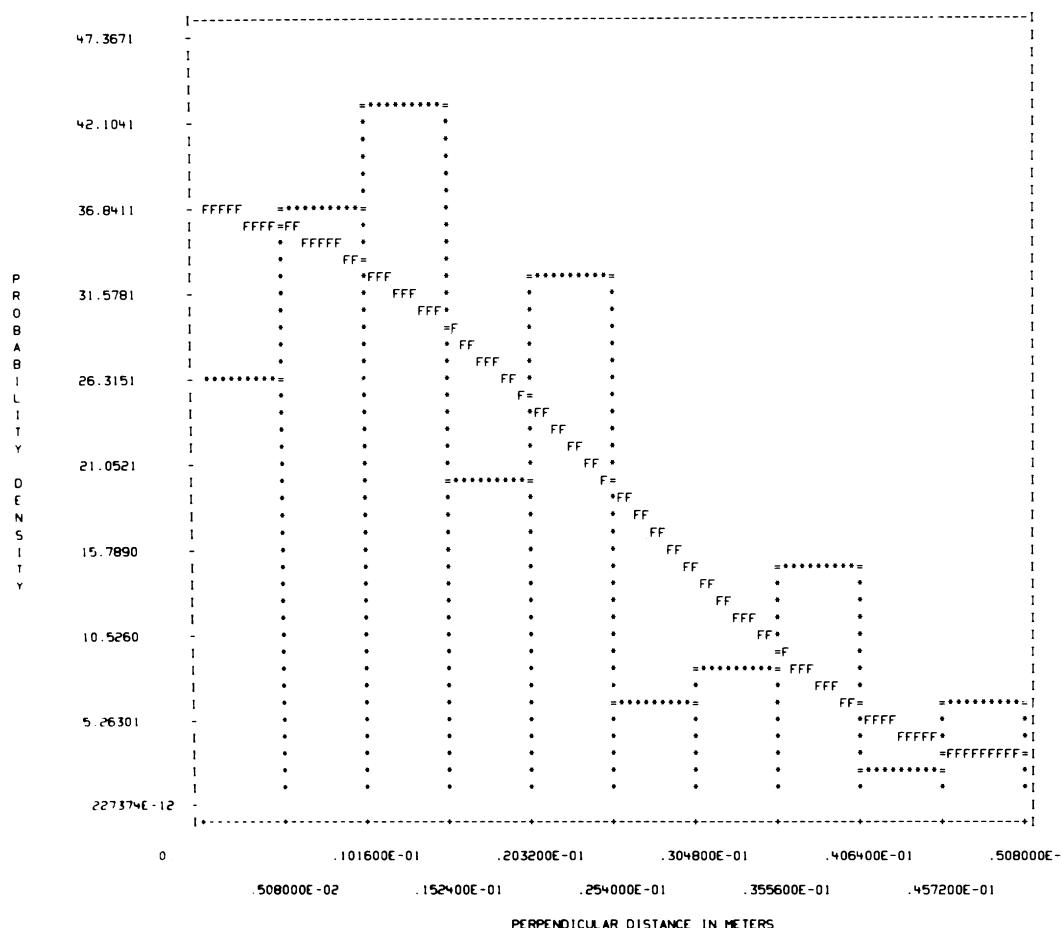


FIG. 42c. The fit of a 1-term cosine series to a histogram of the data from Survey O.

can be used properly (e.g., to gain robustness to some anomaly in the data) and, in general, decisions concerning fit can be roughly made. A second point, discussed by Schweder (1977), concerns the data at the larger distances: i.e., near w . It is the quantity and quality (i.e., accuracy) of the data near the transect line that are crucial. The data near w are of less importance and the fit, quality, and even the quantity of the data near w are of secondary importance. After all, it is $f(0)$ that is of interest, not $f(\bar{x})$ and certainly not $f(w)$. The dated concept of "effective strip width" has clouded this important fact. In summary, fit is perhaps best judged by the stopping rule based on un-

grouped data, which is an integral part of the FS estimation scheme.

Example 5, Measurement Errors

Sometimes, systematic errors occur in measurements of perpendicular and sighting distances or sighting angles. It is important to realize that the accuracy of measurements near the transect line is crucial. Errors in measurement near w are undesirable but have far less effect on the estimate of density (Schweder 1977) because it is $f(0)$ that must be estimated, not $f(w)$. This problem is illustrated and a partial solution is suggested

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	.508E-02	13	17.70
2	.508E-02	.102E-01	18	16.91
3	.102E-01	.152E-01	21	15.40
4	.152E-01	.203E-01	10	13.32
5	.203E-01	.254E-01	16	10.88
6	.254E-01	.305E-01	3	8.32
7	.305E-01	.356E-01	4	5.88
8	.356E-01	.406E-01	7	3.80
9	.406E-01	.457E-01	1	2.29
10	.457E-01	.508E-01	3	1.50

TOTAL CHI-SQUARE VALUE = 15.51 DEGREES OF FREEDOM = 8
PROBABILITY OF A GREATER CHI-SQUARE VALUE = .04990466

FIG. 42d. Chi-square goodness of fit tests of the 1-term cosine series model.

by using the data from Survey 13 of Robinette et al. (1974).

The conduct of Survey 13 included measuring (pacing) perpendicular and sighting distances and later computing the sighting angle. Analysis of the data suggest that considerable measurement error occurred for blocks near the transect line. Perpendicular distances for about 17 percent of the 205 blocks detected were recorded as being exactly on

the transect line (i.e., zero distance). Virtually all methods will perform poorly on those data (cf. Robinette et al. 1974:87). A histogram of the perpendicular distances is shown in Fig. 43a. It seems that distances to blocks near the transect line were "rounded" to zero distance. The problem can also be seen from the sample histogram of the sines of the sighting angles (Fig. 43b). A disproportionate number of blocks were recorded as being detected at 0 and 90° angles. Consequently, the Hayne (1949) estimator performed poorly (45.6% relative bias, Robinette et al. 1974:87). The modified Hayne estimator performed somewhat better, with an estimated relative bias of 30 percent.

The Fourier series approach based on the ungrouped perpendicular distance data required 6 terms to achieve an "adequate" representation of the data (Fig. 43c). The density estimate was 21.01 ± 2.7 blocks/ha, or $\hat{N} = 158.8 \pm 20.4$ blocks. The true $N = 96$; thus the estimated PRB

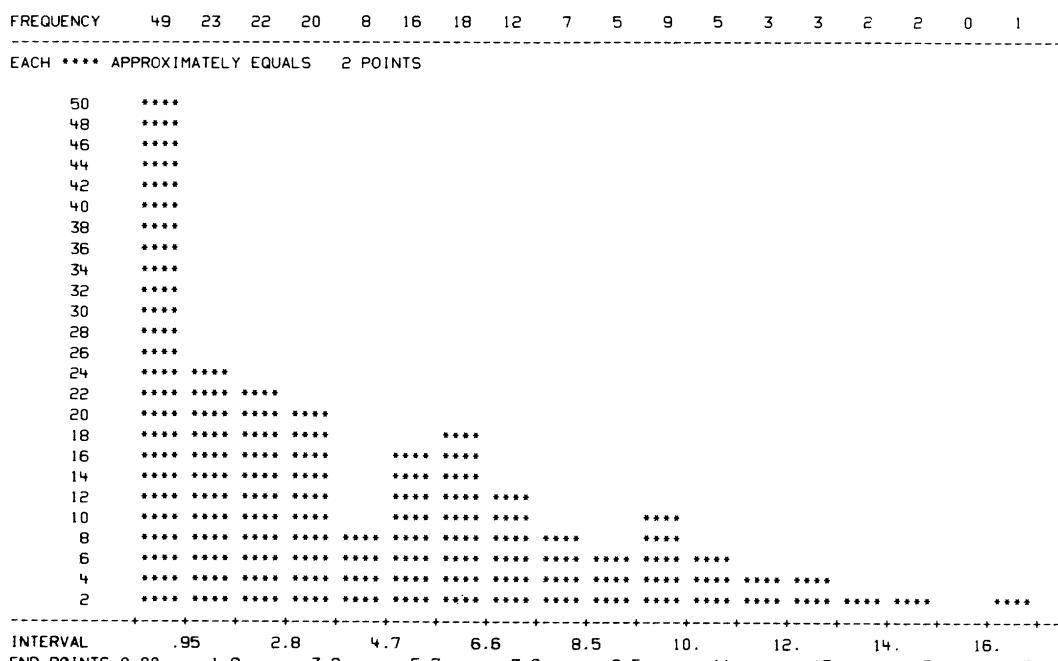


FIG. 43a. Sample histogram of perpendicular distances from Survey 13 (Robinette et al. 1974).

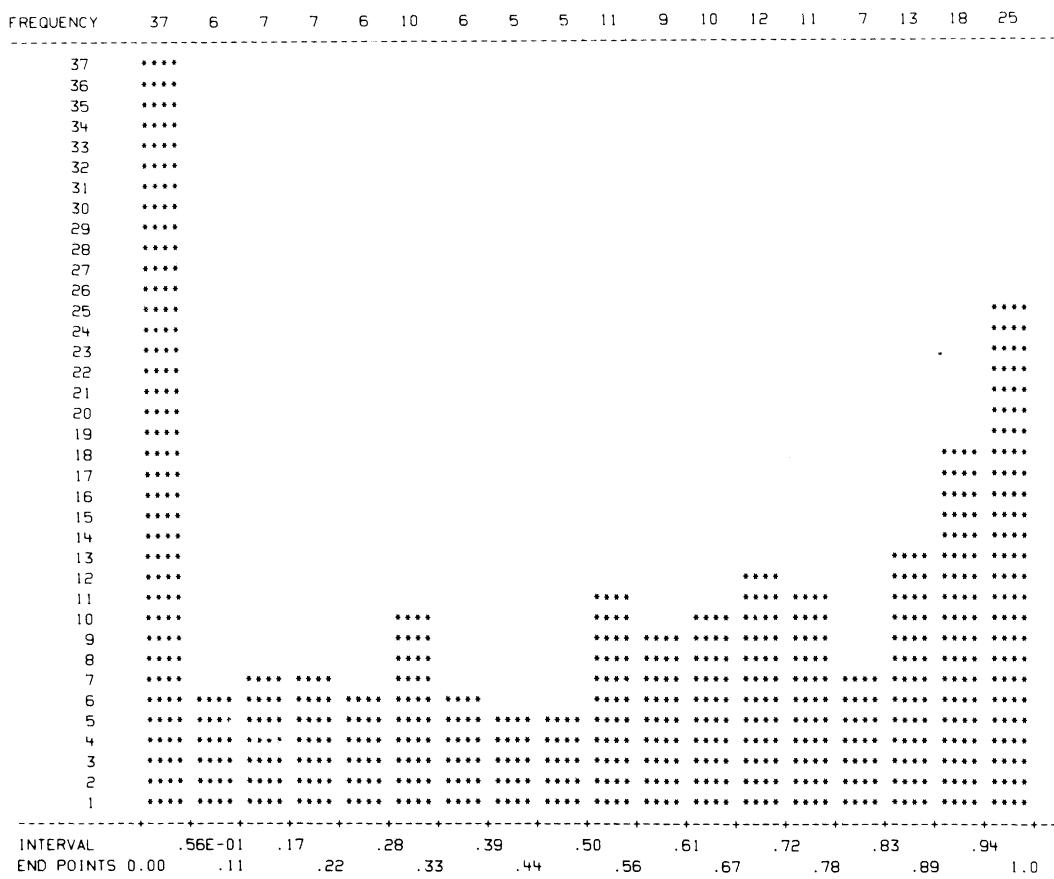


FIG. 43b. Sample histogram of the sine of the sighting angle for Survey 13 (Robinette et al. 1974).

of the Fourier series estimator is 65, even with 6 terms in the FS model. That is a very poor estimate because of the large measurement and rounding errors, particularly near the transect line.

This example illustrates the importance of examining the data by means of histograms, order statistics, or other techniques before accepting an analysis. In this example, grouping the raw data will reduce the effect of the systematic errors of measurement. However, a problem arises because group sizes will be arbitrary and the estimate of density will depend on the size and number of the groups. This problem is fairly minor unless only 2 or 3 groups are used.

We grouped the data from Survey 13 into 6 intervals of equal width to illus-

trate the procedure for achieving robustness to systematic rounding errors. A 2-term Fourier series was selected as the best model for those grouped data (based on the log-likelihood ratio test and the goodness of fit test). The estimate of D was 13.774 ± 1.221 objects/ha; this gives $\hat{N} = 104.1$ and an estimated PRB of only 8.4. In fact, Fourier series estimators based on 1, 3, and 4 terms for those grouped data also provided estimates with far less bias than the estimate based on the ungrouped data. In addition, $\hat{\sigma}_v(\hat{D})$ was 12.9 percent for the ungrouped data and only 8.9 percent for the grouped data (primarily because only 2 cosine terms were adequate for the grouped data).

On the whole, it is believed grouping the data will often achieve a degree of

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 205
 LINE LENGTH 10.52 KILOMETERS
 WIDTH UNTRUNCATED DATA, UNBOUNDED WIDTH
 ***NOTE - THE DATA ARE UNTRUNCATED BUT A WIDTH MUST BE CHOSEN FOR THE FOURIER SERIES ESTIMATOR.
 THE LAST ORDER STATISTIC(LARGEST MEASUREMENT) WAS USED FOR THE WIDTH = 17.07

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	.6973E-01	.4110E-02	5.9	.6168E-01	.7779E-01
A(2)	.2464E-01	.5890E-02	23.9	.1310E-01	.3618E-01
A(3)	.1758E-01	.6038E-02	34.3	.5744E-02	.2941E-01
A(4)	.1397E-01	.5891E-02	42.2	.2928E-02	.2552E-01
A(5)	.1615E-01	.6121E-02	37.9	.4151E-02	.2814E-01
A(6)	.1505E-01	.6187E-02	41.1	.2922E-02	.2718E-01
F(0)	.2157	.2304E-01	10.7	.1705	.2609
D	21.01	2.682	12.8	15.76	26.27

DENSITY(D) UNITS ARE NUMBERS/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2	3	4	5	6
1	1.000	.688	.205	.203	.112	.108
2	.688	1.000	.634	.280	.176	.115
3	.205	.634	1.000	.621	.208	.148
4	.203	.280	.621	1.000	.575	.305
5	.112	.176	.208	.575	1.000	.634
6	.108	.115	.148	.305	.634	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 205. AS EXPLAINED EARLIER IN THE OUTPUT THIS ASSUMPTION WAS MADE TO GET A VARIANCE FOR THE DENSITY ESTIMATE. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .4878E-02

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1141E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 29.95

FIG. 43c. Summary of the Fourier series estimation procedure for the ungrouped perpendicular distance data from Survey 13 (Robinette et al. 1974).

robustness to certain types of measurement errors. Of course, if possible, it is best to avoid or minimize rounding errors when the measurements are made in the field.

Example 6, Data Truncation

In general, it is recommended that data be taken in the field without regard to transect width. A few objects may be detected at extreme distances; those outliers contribute little to the estimates of $f(0)$ and D but often create problems for various estimation schemes, either parametric or nonparametric. The data from Survey 11 of Robinette et al. (1974) are used to illustrate data for which truncation (the establishment of a value for w^*) may be beneficial.

On completion of Survey 11, 352 objects had been detected. The mean perpendicular distance was 19.24 feet (5.86 m) and the sample standard deviation of the data was 16.84 feet (5.13 m). The order statistics are shown in Fig. 44a and a sample histogram is plotted in Fig. 44b. About 97.5 percent of the data are at distances less than 52 feet (15.85 m); however, the longest distance is more than double that length at 119 feet (36.27 m). This is about 6 standard deviations from the mean. The outliers contribute little to the estimation of the important parameters and frequently cause difficulties such as lack of fit for parametric models (and correspondingly biased parameter estimates) and necessitate adding terms in the FS approach.

Using all of the data (i.e., $w^* = 119$

PROGRAM TRANSECT -- LINE TRANSECT DATA ANALYSIS PROGRAM DEVELOPED BY THE UTAH COOPERATIVE WILDLIFE
RESEARCH UNIT. METHODS AND TESTS FROM BURNHAM ET AL (1980). PROGRAM VERSION I, 4/8/79.
ANALYSIS OF EXAMPLES FOR MONOGRAPH

PAGE = 120

* PERPENDICULAR DISTANCE DISTRIBUTIONS *

SURVEY 11 IN FEET

DATA TYPE UNGROUPED DATA
SAMPLE SIZE 352
LINE LENGTH 12.00 KILOMETERS
WIDTH UNTRUNCATED DATA, UNBOUNDED WIDTH

ILLUSTRATION OF THE DISTRIBUTIONS OF OBSERVED PERPENDICULAR DISTANCES PROVIDES INFORMATION ABOUT THE DETECTION FUNCTION $G(x)$ WHICH DICTATES THE TYPE OF ESTIMATOR TO BE USED. HOWEVER, A GRAPHICAL REPRESENTATION ALONE SHOULD NOT BE THE BASIS FOR A DECISION AND MORE RELIANCE SHOULD BE PLACED ON STATISTICAL TESTS. THE ILLUSTRATIONS CAN BE USED TO PICK OUT ANOMALIES IN THE DATA SUCH AS ERRORS OR EXTREME OUTLIERS. CAUTION MUST BE EXERCISED IN INTERPRETING HISTOGRAMS WITH UNEQUAL INTERVALS BECAUSE IT WILL NOT BE IMMEDIATELY OBVIOUS.

FOR THE PURPOSE OF ILLUSTRATING THE HISTOGRAM AND/OR CDF OF UNTRUNCATED DATA IT IS NECESSARY TO IGNORE THAT THERE IS NO FINAL END POINT (UNBOUNDED WIDTH). IN THE CASE OF HISTOGRAMS THE LAST END POINT GIVEN ON THE FIGURES IS EITHER THE LAST ORDER STATISTIC OR A USER DEFINED END POINT. FOR CDF PLOTS THE LAST ORDER STATISTIC IS USED AS THE FINAL END POINT. IN REALITY THE LAST INTERVAL EXTENDS TO INFINITY.

ORDERED PERPENDICULAR DISTANCES

0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.	0.	0.	0.	.610
.610	.610	.610	.610	.610	.610	.610	.610	.610	.610
.610	.610	.610	.610	.610	.914	1.22	1.52	1.52	1.52
1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52
1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52
1.52	1.83	1.83	1.83	1.83	1.83	1.83	2.13	2.13	2.13
2.44	2.44	2.44	2.44	2.44	2.44	2.44	2.44	2.44	2.44
2.44	2.44	2.44	2.44	2.44	2.44	2.44	2.44	2.44	2.44
2.44	2.44	2.74	2.74	2.74	2.74	2.74	3.05	3.05	3.05
3.05	3.05	3.35	3.35	3.35	3.35	3.35	3.35	3.35	3.35
3.35	3.35	3.35	3.35	3.35	3.35	3.66	3.66	3.66	3.66
3.96	3.96	3.96	3.96	3.96	3.96	3.96	3.96	3.96	3.96
4.27	4.27	4.27	4.27	4.27	4.27	4.27	4.27	4.27	4.27
4.57	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88
4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88
5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18
5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18
5.79	5.79	5.79	5.79	5.79	5.79	5.79	5.79	5.79	5.79
5.79	5.79	5.79	5.79	5.79	6.10	6.10	6.10	6.40	6.40
6.40	6.71	6.71	6.71	7.01	7.01	7.01	7.01	7.01	7.01
7.01	7.01	7.01	7.32	7.32	7.32	7.32	7.32	7.32	7.32
7.32	7.32	7.32	7.32	7.32	7.32	7.32	7.62	7.62	7.62
7.62	7.92	7.92	7.92	7.92	7.92	7.92	7.92	7.92	7.92
7.92	7.92	8.23	8.23	8.23	8.23	8.23	8.53	8.53	8.53
8.53	8.84	8.84	8.84	8.84	8.84	9.14	9.14	9.45	9.75
9.75	9.75	9.75	9.75	9.75	9.75	9.75	9.75	9.75	9.75
9.75	10.1	10.1	10.4	10.4	10.4	10.7	10.7	10.7	10.7
11.0	11.0	11.6	11.6	11.6	11.6	11.9	11.9	11.9	11.9
12.2	12.2	12.2	12.2	12.2	12.5	12.5	12.5	12.8	12.8
12.8	13.1	13.1	13.1	13.1	13.4	13.7	13.7	13.7	14.0
14.6	14.6	14.9	14.9	14.9	14.9	14.9	15.5	15.5	15.5
15.8	15.8	15.8	16.2	17.7	18.3	18.9	21.9	27.1	27.1
29.0	36.3								

FIG. 44a. Ordered perpendicular distances from Survey 11 of Robinette et al. (1974). Note the few observations at great distances from the transect line, suggesting truncation may be appropriate. Also, note that 11 percent of the objects were recorded as being on the center line.

feet), the stopping rule indicated that a 6-term Fourier series was required and resulted in $\hat{D} = 17.2 \pm 1.54$ blocks/ha with a \hat{cv} for \hat{D} of 9 percent, and a PRB of 32.1. In contrast, 2.5 percent of the out-

liers ($w^* = 53$ feet) were truncated and it was found that a 2-term Fourier series was adequate. There, $\hat{D} = 16.88$ with a smaller standard error of 1.41 and a \hat{cv} of 8 percent; the estimated PRB is similar

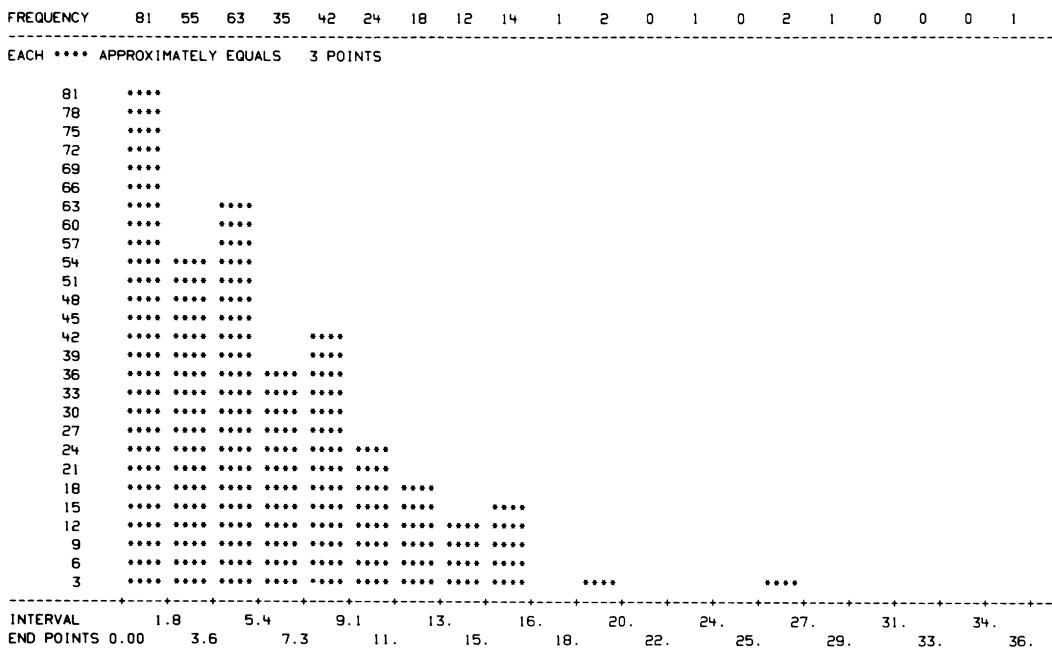


FIG. 44b. Sample histogram of the perpendicular distance data from Survey 11 of Robinette et al. (1974). Note the outlier points beyond about 16.1 m (53 feet).

at 29.9. The bias is caused by the fact that 11 percent of the data were recorded as being exactly on the center line of the transect. Such rounding near the line should be avoided in the field (see *Example 5*). In this example, some decrease in the standard error was realized and the mean square error decreased more ($MSE = 19.68$ vs. 17.37).

Truncating 1–3 percent of the data is recommended if obvious outliers are present in the ungrouped data. Truncation (the establishment of a value for w^*)

is thus somewhat arbitrary but has little effect on the estimate of density. In other words, the FS estimate of D is not sensitive to the selection of w^* , particularly if w^* is chosen to truncate less than, say, 5–10 percent of the data. Truncation obviously affects the number of objects n and will affect $\hat{f}(0)$. However, it is $(nf(0))/(2L)$ that is the estimator of density. The value of n decreases and $f(0)$ increases as the data are truncated. That changes \hat{D} little, as in the above example. What is important is the fact that minor truncation often results in increased precision of \hat{D} . In general, one would always prefer, say, a model with 1 or 2 terms to one with 4–6 terms. It is believed that most estimation approaches would benefit from slight truncation of the data to delete obvious outliers. Unfortunately, all the parametric approaches for ungrouped data in the literature are valid only for untruncated data. Those methods can easily be modified for truncated data; however, iterative numerical methods are necessary to get the ML estimates.

TABLE 11.—SUMMARY OF STATISTICS AND PARAMETER ESTIMATES FROM INDIVIDUAL SURVEY LINES FOR SURVEY K, L = 12.19 m. THE ESTIMATES WERE OBTAINED USING THE FOURIER SERIES MODEL

Line	ℓ_i (km)	n_i	$\hat{f}(0)$	\hat{D}_i/m^2	$\hat{s}_e(\hat{D})$
1	0.002	11	40.67	91.7	31.3
2	0.002	13	20.44	54.5	35.5
3	0.002	20	40.76	167.1	42.0
4	0.002	15	54.07	166.3	63.8
5	0.002	14	36.26	104.1	32.3

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 73
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .5080E-01 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	17.58	2.564	14.6	12.55 22.60
F(0)	37.26	2.564	6.9	32.24 42.29
D	111.6	13.82	12.4	73.21 149.9

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 56.5). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 4 DEGREES OF FREEDOM IS 2.776

SQUARED COEFFICIENT OF VARIATION FOR N = .1060E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .4734E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 69.13

FIG. 45a. Summary of the Fourier series estimation procedure for the ungrouped perpendicular distance data from Survey K.

The estimation theory has been extended for several of the parametric approaches to enable correct analysis of truncated data; those procedures are available as an option in program TRANSECT.

Example 7, Variance Estimates Using Replicate Lines

The data from our table top Survey K are used to illustrate the estimation of D and $\text{var}(\hat{D})$, using replicate lines. Survey K is identical to Survey D except that the transect lines ran the length of the artificial population. The population parameters are N = 319 and D = 107.3/m². Five line transects, each 2.438 m long, were run and a total of 73 objects were detected, n = 73. The relevant information for each individual line is presented in Table 11.

A pooled estimate of density can be computed from Eq. (1.21) by using individual estimates of D_i from the FS estimation approach:

$$\hat{D} = \frac{1}{L} \sum_{i=1}^R \ell_i \hat{D}_i$$

where R = 5 in this example. Using that

estimator and the information in Table 11, $\hat{D} = 116.7$. The sampling variance of the estimate can be computed from the individual estimates of density, D_i (from Eq. (1.22)),

$$\text{var}(\hat{D}) = \frac{1}{L(R-1)} \sum_{i=1}^R \ell_i (\hat{D}_i - \hat{D})^2$$

and

$$\hat{s}e(\hat{D}) = \sqrt{\text{var}(\hat{D})}.$$

That approach yields $\text{var}(\hat{D}) = 482.6$ and $\hat{s}e(\hat{D}) = 22.0$. A 95 percent confidence interval can be established as $\hat{D} \pm t_{0.05} \hat{s}e(\hat{D})$ where $t_{0.05}$ is the value from the t-distribution with R - 1 degrees of freedom. In that example, $t_{0.05} = 2.776$ and the 95 percent confidence interval is 55.7 to 177.7.

Although the estimate of density is quite good, the estimated sampling variance may be an overestimate caused by the very small sample sizes for individual lines (e.g., n₁ = 11). Such a procedure is severely limited to those few surveys where the number of objects seen on replicate lines is fairly large (perhaps at least in the 20 to 30 range).

In general, the best approach is to es-

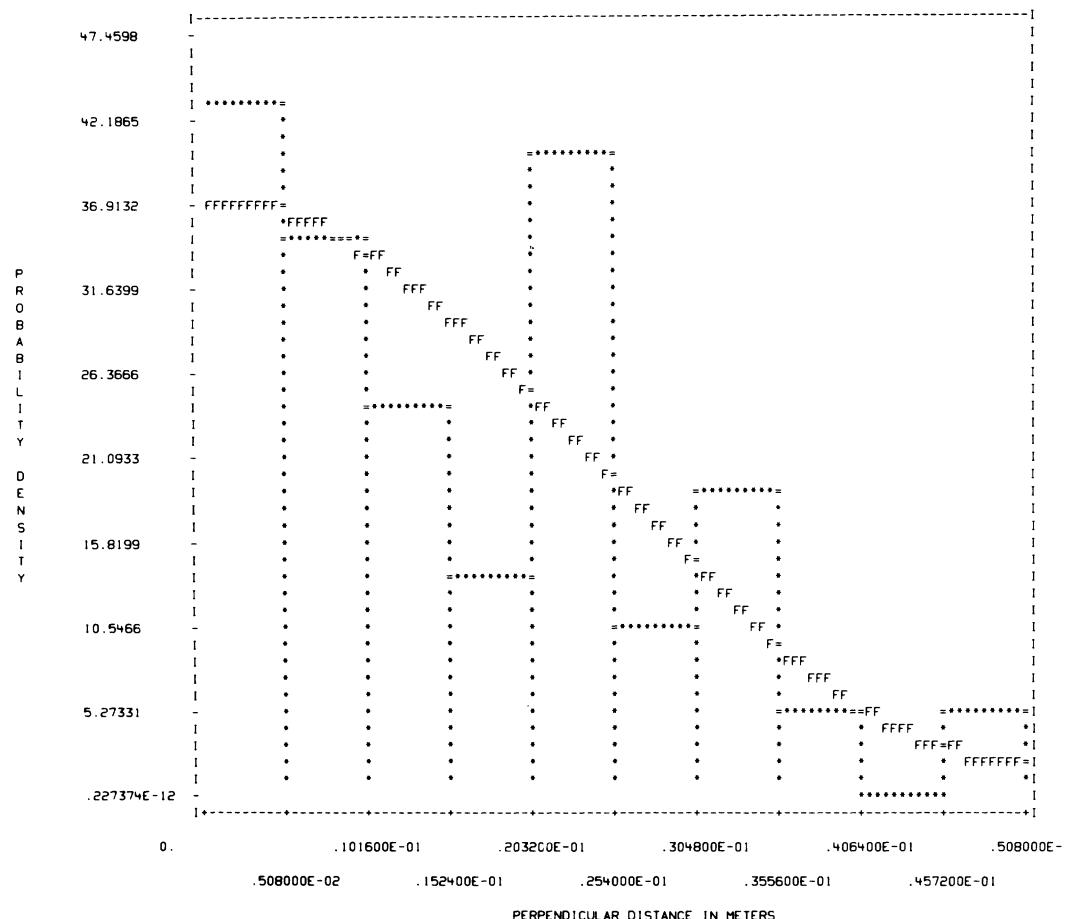


FIG. 45b. The 1-term Fourier series plotted against a sample histogram of the ungrouped data for Survey K.

timate D by using the FS method based on all the perpendicular distance data. The $\text{var}(\hat{D})$ can be estimated best by computing $\hat{\text{var}}(\hat{f}(0))$ using all the data and the variance equation from the FS procedure (Eq. 2.6). The $\hat{\text{var}}(n)$ can be estimated from the data from replicate lines (Eq. 1.23 or 1.24), and $\hat{\text{var}}(\hat{D})$ can then be computed from Eq. (1.17) as

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

or, equivalently,

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

The results of this approach are shown in

Figs. 45a-45c: $\hat{D} = 111.56 \text{ objects/m}^2$ and $\hat{s}_e(\hat{D}) = 13.8$. Again, using the t -distribution with $R - 1$ degrees of freedom, the 95 percent confidence interval is 73.21-149.9, a marked improvement over the first approach. As in several previous examples, $\hat{\text{var}}(n)$ is larger than $\hat{\text{var}}(\hat{f}(0))$; $\hat{cv}(\hat{f}(0)) = 6.9$ percent versus $\hat{cv}(n) = 10.3$ percent.

A third approach that appears attractive is to obtain $\hat{\text{var}}(\hat{D})$ by using the jackknife procedure (see PART 1). To pursue that approach, D is estimated first, using all the data except those objects found on line 1 and denote this $\hat{D}_{(1)}$ to signify that line 1 was excluded. Next, $\hat{D}_{(2)}$ using all the data except those for line 2 is esti-

CELL 1	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	.508E-02	16	13.71
2	.508E-02	.102E-01	13	13.08
3	.102E-01	.152E-01	9	11.89
4	.152E-01	.203E-01	5	10.25
5	.203E-01	.254E-01	15	8.32
6	.254E-01	.305E-01	4	6.28
7	.305E-01	.356E-01	7	4.35
8	.356E-01	.406E-01	2	2.71
9	.406E-01	.457E-01	0	1.52
10	.457E-01	.508E-01	2	0.89

TOTAL CHI-SQUARE VALUE = 14.68 DEGREES OF FREEDOM = 8

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .06568462

FIG. 45c. Simple goodness of fit test of the Fourier series based on arbitrary grouping. The test indicates a satisfactory fit.

mated, and so on. Those estimators $\hat{D}_{(i)}$ for 5 lines of Survey K are 116.5, 117.8, 97.7, 112.3, and 113.4, respectively.

The jackknife pseudovalues are then computed as

$$\hat{D}^{(i)} = R\hat{D} - (R - 1)\hat{D}_{(i)}$$

for the case of equal line lengths, where \hat{D} is the density estimate computed from all the data. The pseudovalues are 91.8, 86.6, 167.7, 108.6, and 104.2, respectively. Finally, the estimate of the sampling variance of \hat{D} for equal ℓ_i from Eq. (1.22) is

$$\hat{\text{var}}(\hat{D}) = \frac{1}{R(R - 1)} \sum_{i=1}^r (\hat{D}^{(i)} - \hat{D})^2.$$

This yields $\hat{\text{var}}(\hat{D}) = 211.4$ and $\hat{\text{se}}(\hat{D}) = 14.5$. This estimate of $\text{se}(\hat{D})$ compares well with the previous approach. Of course, a 95 percent confidence interval is based on the appropriate value from the t -distribution with $R - 1$ degrees of freedom.

Three approaches to estimation of D and $\text{se}(\hat{D})$ are summarized below:

- (1) $\hat{\text{var}}(\hat{D})$ from replicate lines, $\hat{D} = 116.7$ and $\hat{\text{se}}(\hat{D}) = 22.0$;
- (2) $\hat{\text{var}}(n)$ from replicate lines and $\hat{\text{var}}(\hat{f}(0))$ from using all the data, $D = 111.6$ and $\hat{\text{se}}(\hat{D}) = 13.8$; and
- (3) $\hat{\text{var}}(\hat{D})$ from the jackknife approach, $\hat{\text{se}}(\hat{D}) = 14.5$.

The last 2 approaches are favored, but more work on the problem is clearly

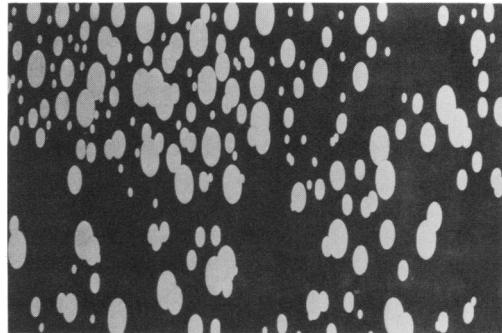


FIG. 46. An artificial population of points and their "elliptical flushing envelopes." $N=318$ and $D=107$ points/m². Ten lines were run across the width of the board in a systematic design with a random first start. Only a portion of the population is shown.

needed. Finally, it is noted that \hat{D} itself can be computed in several ways by using the jackknife approach. At this time, computing \hat{D} by using all of the data and the FS approach is favored.

Example 8, Noncircular Flushing Envelopes

A second artificial population was created similar to the one described in *Example 1* and used for several of the preceding examples. In that second population, 318 points were placed non-randomly on a 1.2192×2.4384 -m plywood board with an area of 2.9729 m² and a density of 107.0 /m². The basis for line transect sampling of that population was an "elliptical flushing envelope" (Fig. 46). The objects of interest were assumed to be at the center of each ellipse. Ellipses were of 4 sizes; the length of the major axes were 3.30, 7.11, 10.16, and 13.46 cm. Following the assumption required for the generalized Hayne estimator, the ratio of major to minor axis of each ellipse was constant ($c = 2.12$). The survey was conducted in the same way as that described in *Example 1*.

The data from Survey A on that population were used to illustrate estimation of density, using 2 approaches. Survey A consisted of 10 lines running across the width of the board in a systematic design

PD	SD	MESANG	COMANG	SINE	
.1102E-01	.3792E-01	17.63	17.84	.291	MEANS
.8054E-02	.1510E-01	14.31	13.94	.216	STANDARD DEVIATIONS
.1162E-02	.2179E-02	2.07	2.01	.031	STANDARD ERRORS OF THE MEANS

CORRELATIONS(LOWER OFF DIAGONAL), AND THEIR TWO TAILED SIGNIFICANCE LEVELS(UPPER OFF DIAGONAL)

PD	SD	MESANG	COMANG	SINE	
*****	-.0003	0.0000	0.0000	0.0000	PD
.3489	*****	.6049	.6702	.6863	SD
.8039	-.0761	*****	0.0000	0.0000	MESANG
.8452	-.0628	.9885	*****	0.0000	COMANG
.8197	-.0596	.9966	.9861	*****	SINE

FIG. 47a. Summary statistics of data from Survey A of a population of points with elliptical flushing envelopes. The mnemonic abbreviations are Perpendicular Distance (PD), Sighting Distance (SD), MEasured ANGLE (MESANG), COMputed ANGLE (COMANG), and COMputed SINE (COMSINE) of the angle.

with a random first start. As before, each line was 1.2192 m long for a total line length of $L = 12.192$ m; the lines were parallel and about 20.3 cm apart. On the survey, 48 objects were detected; and the results of an analysis of these data are given in Figs. 47a-47g.

A variety of basic summary statistics are given in Fig. 47a. The measured and computed sighting angles are highly correlated (see Fig. 47b), so there is reason to believe the data were taken accurately. However, the average sighting angle is only 17.63° . A test was made of the null hypothesis that the observed average sighting angle was 32.7° as

$$z = (\hat{\theta} - 32.7)/(21.56/\sqrt{48}).$$

The computed test value was $z = -4.84$ and is highly significant, thus invalidating both the Hayne and modified Hayne estimators. Our approach to the analysis of the data emphasizes the use of the ungrouped perpendicular distance data (Fig. 47c) and the FS estimator. Alternatively, the generalized Hayne estimator is used because Survey A follows the basis for that estimation procedure exactly.

In that example, the Fourier series estimator performed very well (Fig. 47d).

Density is estimated at 108.5 ± 19.9 objects/m² and the estimated PRB is only 1. The 95 percent confidence interval is 63.5-153.6. The $\text{se}(\hat{D})$ was estimated by using all the data to obtain $\hat{\text{var}}(\hat{f}(0))$ and by using the number observed on each replicate line to compute $\hat{\text{var}}(n)$. The n_i for the 10 lines are 8, 8, 2, 3, 4, 3, 8, 3, 4, and 5, respectively, for n_1, n_2, \dots, n_{10} . Thus, the $\hat{\text{var}}(n)$ (from Eq. 1.24) is 55.11 and $\hat{\text{se}}(n)$ is 7.42. Information in Figs. 47e and 47f gives insight concerning the fit of the Fourier series model to the data.

The results of applying the generalized Hayne estimator to sighting angle and distance data are given in Fig. 47g. The density is estimated at 149.2 ± 36.2 objects/m² with an estimated PRB of 39.4. The 95 percent confidence interval is 67.3 to 231.1, which includes the true density of 107. Most discouraging is the fact that the $\hat{\text{se}}(\hat{D})$ is nearly double that for the FS estimator.

In summary, it is believed that the FS procedure performs well as a general procedure and its bias is small for a wide variety of underlying detection functions. The sampling variance of the FS estimator is small and it is simple to compute for ungrouped data. The properties of *model robustness, pooling robustness,*

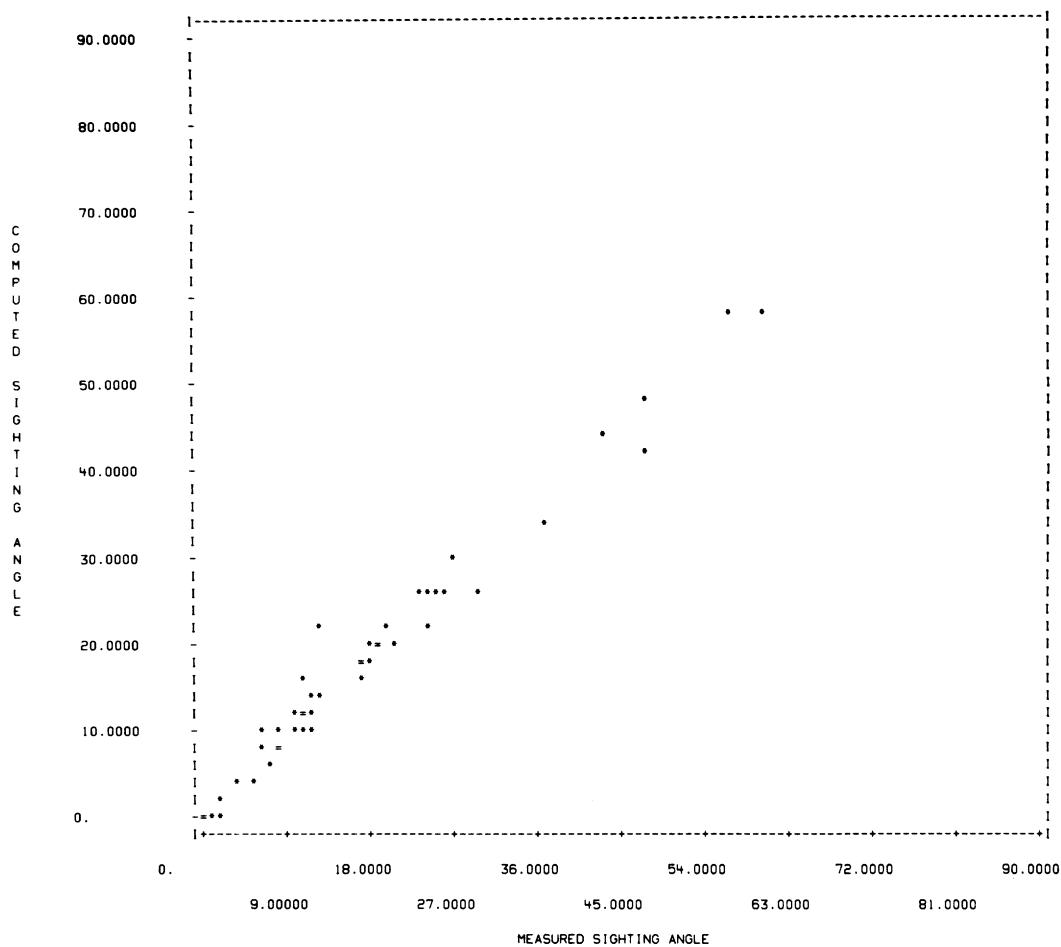


FIG. 47b. Plot of measured and computed sighting angles for Survey A. The correlation coefficient is high ($r=0.99$) and indicates accurate measurement and record keeping of the distance and angle data.

shape criterion, and high *estimator efficiency* are satisfied by the FS estimator.

It might be helpful to contrast the FS approach to perpendicular distance data with the Hayne type approaches to sighting distance and angle data. First, the sighting distance and angle data are quite sensitive to even small departures from critical assumptions. Second, the models on which the estimators are based necessitate an additional assumption over and above those required for the analysis of perpendicular distance data. Third, the underlying models are quite idealized and abstract in relation to real field studies, quite unlike the models for perpendicular distance data. The concepts of cir-

pular or elliptical flushing radii are gross idealizations of the real world. Regardless of the shape of the flushing envelope

FREQUENCY	11	4	6	11	3	4	4	3	0	2
11	****			****						
10	****			****						
9	****			****						
8	****			****						
7	****			****						
6	****		****						
5	****		****						
4		
3		
2
1

INTERVAL .32E-02 .95E-02 .16E-01 .22E-01 .29E-01
END POINTS 0.00 .64E-02 .13E-01 .19E-01 .25E-01 .32E-01

FIG. 47c. Sample histogram of the perpendicular distance data for Survey A. The 48 observations were grouped into 10 equal intervals of width 0.32 cm (0.125 inches).

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 48
 LINE LENGTH 12.19 METERS
 WIDTH TRUNCATED DATA, WIDTH = .3175E-01 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	23.64	5.442	23.0	12.98	34.31
F(0)	55.14	5.442	9.9	44.47	65.81
D	108.5	19.91	18.3	63.50	153.6

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 55.1). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262

SQUARED COEFFICIENT OF VARIATION FOR N = .2392E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .9741E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 71.06

FIG. 47d. Summary of the Fourier series estimation procedure for Survey A.

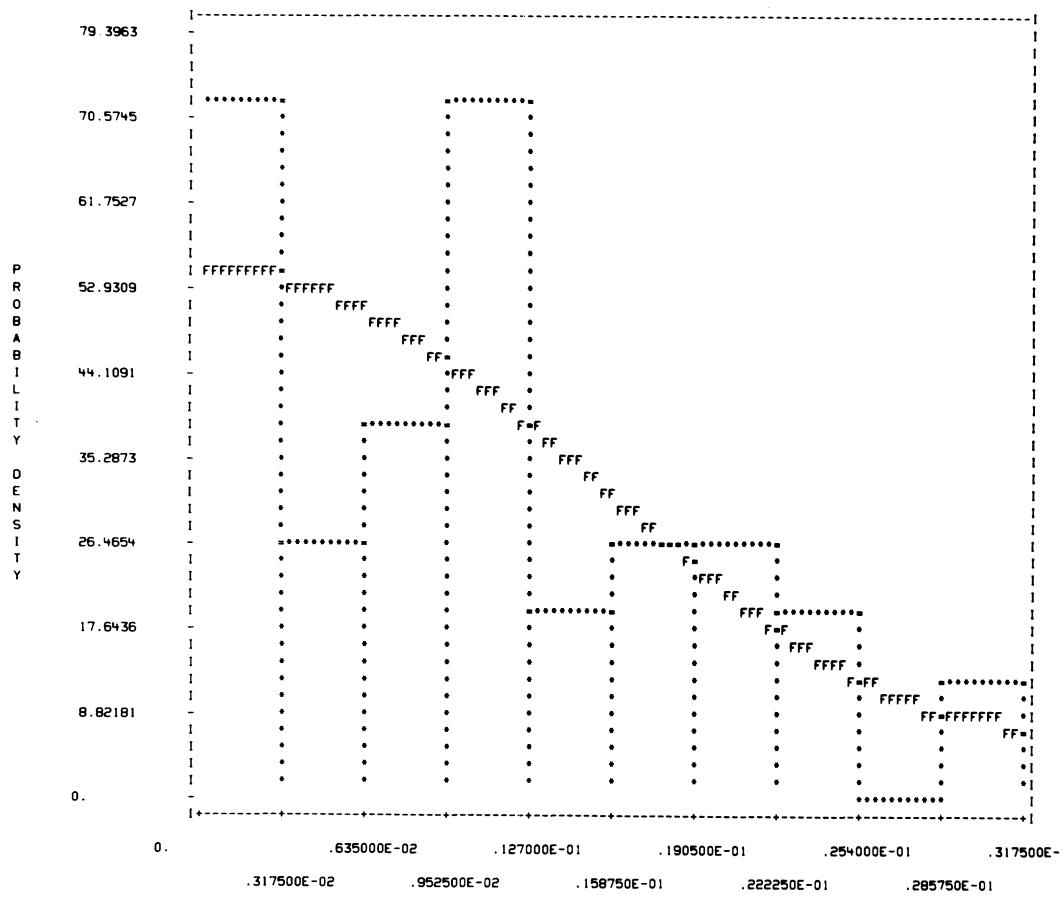


FIG. 47e. The 1-term Fourier series plotted against a sample histogram of the data for Survey A. The 1-term Fourier series was indicated from the stopping rule based on the minimum mean squared error.

CELL I	CUT POINTS	OBSERVED VALUES	EXPECTED VALUES	CHI-SQUARE VALUES
1	0.00	.318E-02	11	.845
2	.318E-02	.635E-02	4	8.00
3	.635E-02	.953E-02	6	.244
4	.953E-02	.127E-01	11	6.43
5	.127E-01	.159E-01	3	5.36
6	.159E-01	.191E-01	4	4.24
7	.191E-01	.222E-01	4	.134E-01
8	.222E-01	.254E-01	3	.217
9	.254E-01	.286E-01	0	1.60
10	.286E-01	.318E-01	2	1.26

TOTAL CHI-SQUARE VALUE = .9.891 DEGREES OF FREEDOM = 8

PROBABILITY OF A GREATER CHI-SQUARE VALUE = .27273814

FIG. 47f. Simple goodness of fit tests of the Fourier series based on grouping the data into 10 intervals of equal size. The test indicates a good fit of the 1-term Fourier series model to the data ($P=0.27$).

lope, it is clear that this concept is a poor one, as evidenced by the fact that animals commonly flush, or are seen, after the observer has passed the animal. Fourth, the estimators are quite sensitive to very small sighting distances because a particular value of $1/r_i$ will dominate the estimate if r_i is small. Again converting the sighting distance and angle data to per-

pendicular distances and employing the FS procedure is recommended, or, at least, some other estimation procedure based on perpendicular distances.

Example 9, Other Estimators

Hemingway's (1971) data were used in PART 2 to illustrate the FS estimator (see Fig. 28). In this example, several other estimators available as options in program TRANSECT are illustrated.

Pollock's (1978) exponential power series estimator is shown in Figs. 48a and 48b. Here, the exact maximum likelihood estimates are obtained numerically, using the ungrouped data. That procedure is *model robust*, tends to be *pooling robust* and essentially meets the *shape criterion*; however, it is somewhat inefficient compared to the Fourier series (efficiency is 29% in this example).

The exponential quadratic estimator is shown in Figs. 48c and 48d and is very similar to the exponential power series.

DATA TYPE UNGROUPED DATA
SAMPLE SIZE 48
LINE LENGTH 12.19 METERS
WIDTH TRUNCATED DATA, WIDTH = .3175E-01 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	2.424	.3912	16.1	1.658	3.191
F(0)	80.39	14.67	18.2	51.64	109.1
D	158.3	37.85	23.9	72.63	243.9

DENSITY(D) UNITS ARE NUMBERS/SQUARE METER

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

THE CONFIDENCE INTERVALS FOR THE COEFFICIENTS A(1) AND F(0) WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE 1.96. THE VARIANCE OF N WAS ESTIMATED USING REPLICATE LINE LENGTHS (VAR(N) = 55.1). THE CONFIDENCE INTERVAL FOR DENSITY WAS CONSTRUCTED WITH A T DISTRIBUTION WITH THE DEGREES OF FREEDOM EQUAL TO THE NUMBER OF LINE LENGTHS - 1. THE T-VALUE WITH 9 DEGREES OF FREEDOM IS 2.262.

SQUARED COEFFICIENT OF VARIATION FOR N = .2392E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .3329E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 41.81

WHEN THE PARAMETER A(1)(I.E.,C) IS ESTIMATED IN A SIGHTING DISTANCE MODEL THERE IS A CORRESPONDING INCREASE IN THE VARIANCE. THERE IS A TRADE-OFF BETWEEN ROBUSTNESS AND VARIANCE. THE VARIANCE ATTRIBUTABLE TO THE ESTIMATION OF THE PARAMETER IS GIVEN BELOW.

SQUARED COEFFICIENT OF VARIATION OF A(1)(I.E.,C) = .2604E-01
SQUARED COEFFICIENT OF VARIATION OF THE MEAN SD RECIPROCAL (1/SD) = .7245E-02

PERCENT OF SAMPLING VARIANCE OF F(0) ATTRIBUTABLE TO ESTIMATING A(1) 78.23

PERCENT OF SAMPLING VARIANCE OF D ATTRIBUTABLE TO ESTIMATING A(1) 45.52

FIG. 47g. Summary of the generalized Hayne estimator for the data from Survey A.

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 73
 LINE LENGTH 60.00 KILOMETERS
 WIDTH TRUNCATED DATA, WIDTH = 400.0 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	148.8	37.21	25.0	75.90	221.8
A(2)	1.414	.4454	31.5	.5415	2.287
F(0)	.7447E-02	.1593E-02	21.4	.4324E-02	.1057E-01
D	.4530E-01	.1105E-01	24.4	.2365E-01	.6695E-01

DENSITY(D) UNITS ARE NUMBERS/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2
1	1.000	.894
2	.894	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 73. AS EXPLAINED EARLIER IN THE OUTPUT THIS ASSUMPTION WAS MADE TO GET A VARIANCE FOR THE DENSITY ESTIMATE. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1370E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .4577E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 23.03

FIG. 48a. Summary of the exponential power series (Pollock 1978) for Hemingway's (1971) data.

That estimator has the same general properties, including low estimation efficiency (24%). Note that in both models the estimators of the 2 parameters are highly correlated (i.e., 0.894 for the exponential power series and -0.942 for the exponential quadratic).

The results of the Gates et al. (1968) procedure appear in Figs. 48e-48f. That procedure is based on the negative exponential density. It is not *model robust* or *pooling robust* and does not meet the *shape criterion*. It is quite efficient if the underlying model is, in fact, negative exponential. The usual chi-square goodness of fit test fails to reject the model (Fig. 48f). However, Epstein's (1960) test gives a sound rejection of the negative exponential. Therefore, one would reject this estimation method for these data. It has been found that the chi-square test is not powerful in assessing most models. That deficiency is not particularly important for estimation schemes that are *model robust* and *pooling robust* and have a stopping rule to dictate the number of terms. However, for models such as the

negative exponential, a more powerful test is crucial. Unfortunately, Epstein's (1960) test is applicable only for ungrouped, untruncated data.

The half-normal estimator is shown in Figs. 48g and 48h and is the last option of TRANSECT illustrated. It meets the *shape criterion* but is not *model robust* or *pooling robust* and seems to be of only limited use. In this example, a good fit is suggested if the last 2 intervals are combined.

The results of the various analyses are summarized in Table 12. In this example, the Fourier series and half-normal estimators compare favorably; the point estimates and estimated standard errors are similar. The 2 generalized parametric estimators give higher density estimates with less precision but, in general, are no more biased than the other methods. None of the 5 estimation methods are rejected by the simple chi-square test; however, the estimates of density vary substantially from one method to the others. It is nearly impossible to perform tests of the data to determine the "prop-

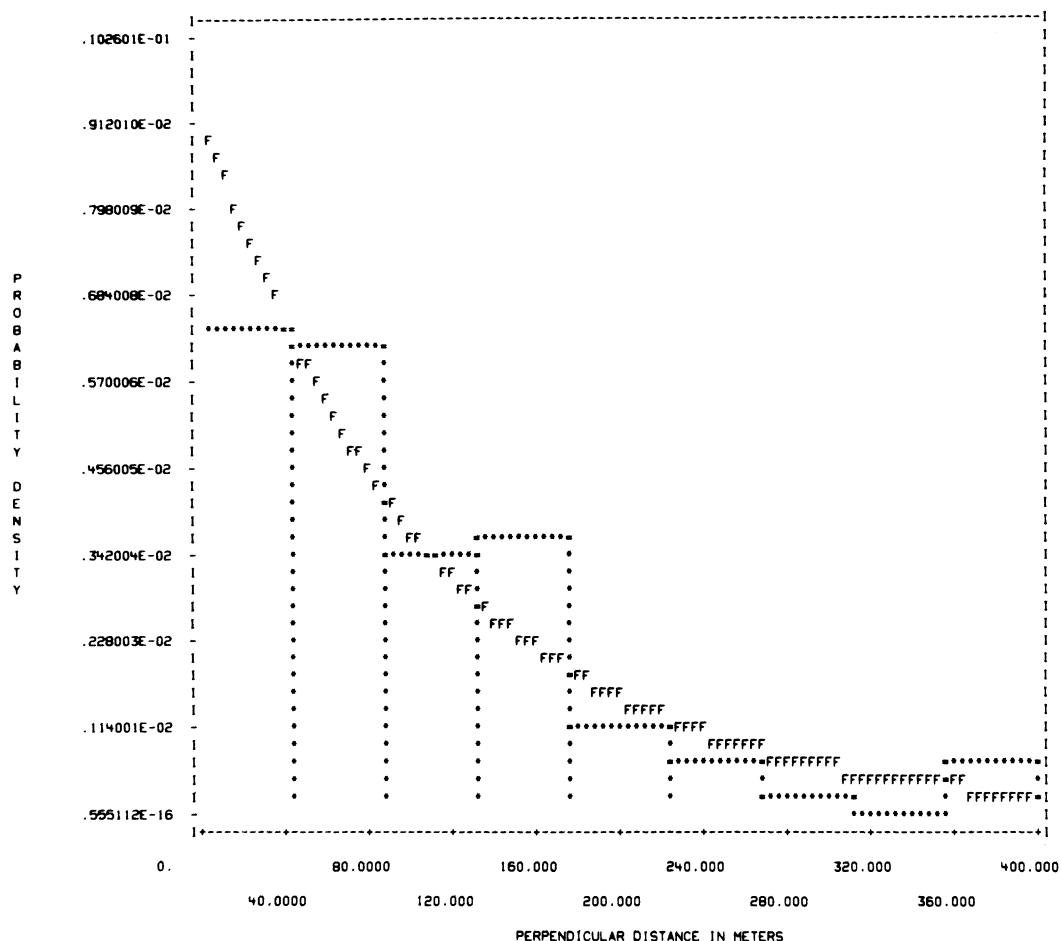


FIG. 48b. Fit of the exponential power series on Hemingway's (1971) data. The perpendicular distance data have been grouped arbitrarily into 9 equal intervals for purposes of illustration. The estimation was done using the ungrouped data. A good fit is indicated ($\chi^2=7.88$, 6 df, $P=0.25$).

er" model with sample sizes in the range typically encountered. It is for this reason that modeling $g(x)$ should be based on criteria independent of specific data sets (e.g., *model robustness, pooling robust-*

ness, shape criterion, and estimator efficiency).

A number of other features are available in program TRANSECT (see Laake et al. 1979), including exact ML estima-

TABLE 12.—ANALYSIS OF HEMINGWAY'S (1971) DATA BY 5 DIFFERENT ESTIMATORS. DENSITY IS GIVEN IN OBJECTS PER HECTARE (SEE ALSO TABLE 8)

Estimator/model	Density \bar{D}	Standard error $\hat{s}_e(\bar{D})$	Coefficient of variation	Model fit
Fourier series	39.6	6.0	0.15	Yes
Exponential power series	45.3	11.1	0.24	Yes
Exponential quadratic	48.7	12.1	0.25	Yes
Negative exponential	56.7	10.0	0.18	No
Half-normal	37.4	5.4	0.15	Yes

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 73
 LINE LENGTH 60.00 KILOMETERS
 WIDTH TRUNCATED DATA, WIDTH = 400.0 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL	
A(1)	.5390E-02	.4264E-02	79.1	-.2968E-02	.1375E-01
A(2)	.1255E-04	.1399E-04	111.5	-.1488E-04	.3998E-04
F(0)	.8006E-02	.1756E-02	21.9	.4564E-02	.1145E-01
D	.4870E-01	.1211E-01	24.9	.2497E-01	.7243E-01

DENSITY(D) UNITS ARE NUMBERS/HECTARE

SAMPLING CORRELATION OF ESTIMATED PARAMETERS

	1	2
1	1.000	-.942
2	-.942	1.000

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = .73. AS EXPLAINED EARLIER IN THE OUTPUT THIS ASSUMPTION WAS MADE TO GET A VARIANCE FOR THE DENSITY ESTIMATE. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1370E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .4810E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 22.16

FIG. 48c. Summary of the exponential quadratic for Hemingway's (1971) data.

tion of these models for the case of grouped data.

ANALYSIS AND INFERENCE PROCEDURES FOR MOBILE POPULATIONS

One of the basic assumptions of line transect theory is that the population is immobile before detection (see PART 1, *Assumptions*). That assumption is explored, and analysis procedures for mobile populations are discussed. This is the only section that deals with populations whose members may move from their original position before being detected. The notation in this section is at variance with other parts of the monograph, including APPENDIX A.

Three possible situations with mobile populations are recognized:

- (1) the animals are virtually at rest before detection,
- (2) the animals move randomly with respect to the observer before detection, or
- (3) the animals move in response to the observer before detection, being either attracted or repelled.

Case (1) stipulates that there is virtually no movement (e.g., a tortoise) and the usual estimation theory will hold. Schweder (1977) has shown that random movement of the animals that is slow relative to the observer (so as to avoid counting animals more than once) is not a practical problem for the usual estimation theory. Case (2), of substantial random movement, has been examined by Yapp (1956), Skellam (1958), and Schweder (1977). The estimator developed by Yapp (1956) and Skellam (1958) requires an estimate of the average velocity of both the animals and observer and it makes some unreasonable assumptions. Thus, the estimator of Yapp (1956) and Skellam (1958) is of little use. Nonrandom movement (Case 3) in response to the observer before detection probably is the most realistic situation. For most species, the movement will be away from the observer. Occasionally, species will move toward the observer (e.g., some species of dolphins); however, that is rare. In this section, only "evasive" animal movement is treated.

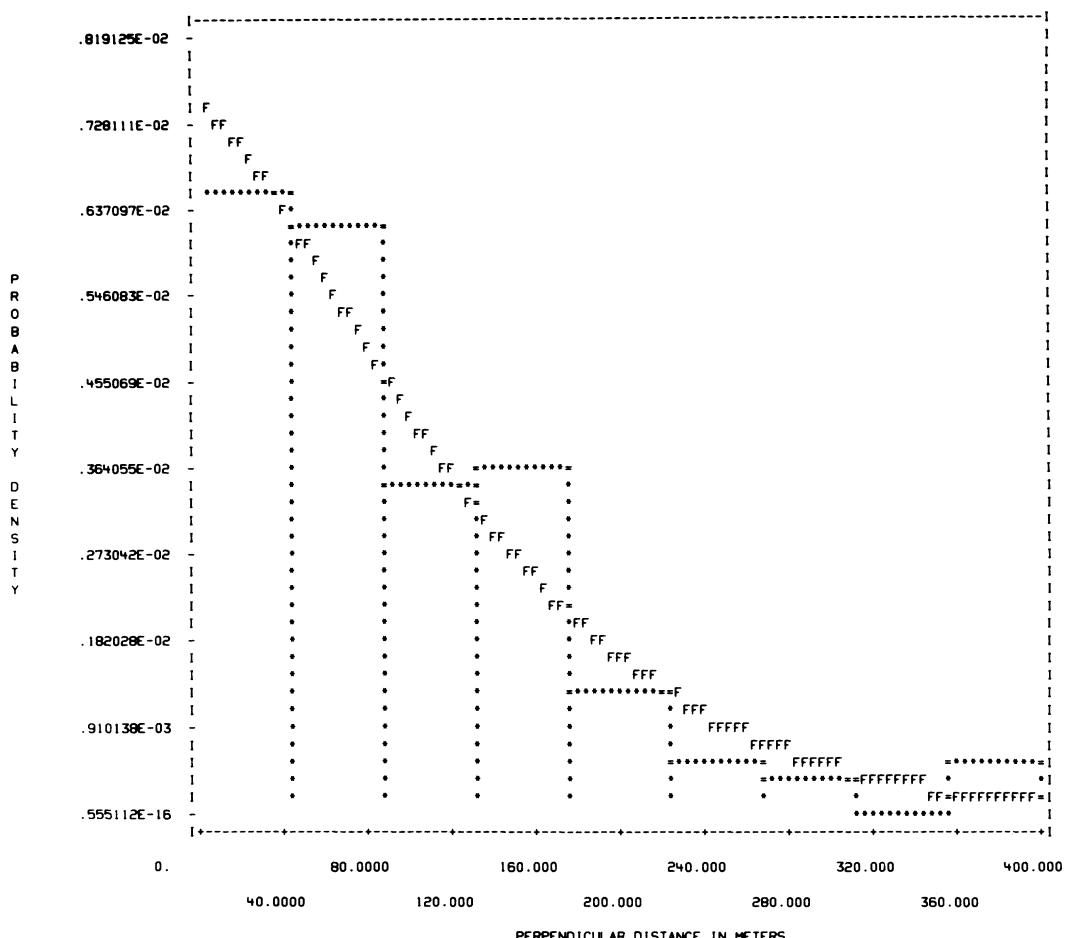


FIG. 48d. Fit of the exponential quadratic on Hemingway's (1971) data. The perpendicular distance data have been grouped arbitrarily into 9 equal intervals for purposes of illustration. The estimation was done using the ungrouped data. A good fit is indicated ($\chi^2=8.32$, 6 df, $P=0.25$).

Animal movement away from the line has been recognized as a potential problem by Eberhardt (1968) and Emlen (1971). The response of the animal to move away from the line only causes problems if the movement is unobserved. If all moving animals were detected and their original location determined, there would not be a problem. However, if one of the following occurs:

- (1) animals move to a new position and are not detected,
- (2) animals move to a new position and are then detected, or

(3) animals are detected in the process of moving,

the unobserved movement will cause the estimate of density to be biased negatively (i.e., it will tend to be low). Only the change in the perpendicular distance affects the estimation of density; any movement parallel to the line is irrelevant unless the animal leaves the sampled area. In the estimation theory presented so far it was assumed, that because of random line placement, the distribution of all perpendicular distances (not just observed ones) is uniform. The movement

DATA TYPE UNGROUPED DATA
 SAMPLE SIZE 73
 LINE LENGTH 60.00 KILOMETERS
 WIDTH TRUNCATED DATA, WIDTH = 400.0 METERS

PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF. OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.9080E-02	.1337E-02	14.7	.6460E-02 .1170E-01
F(0)	.9327E-02	.1239E-02	13.3	.6901E-02 .1175E-01
D	.5674E-01	.1004E-01	17.7	.3706E-01 .7642E-01

DENSITY(D) UNITS ARE NUMBERS/HECTARE

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

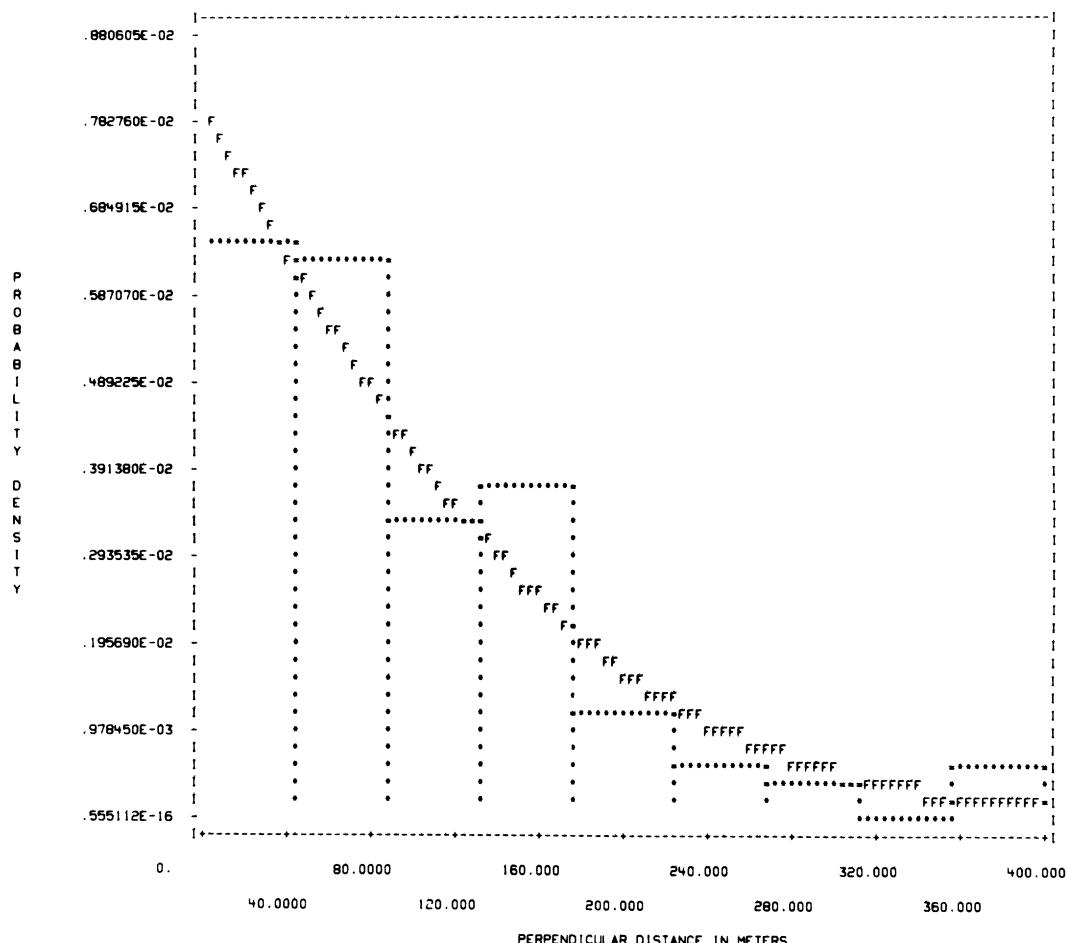
ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 73. AS EXPLAINED EARLIER IN THE OUTPUT THIS ASSUMPTION WAS MADE TO GET A VARIANCE FOR THE DENSITY ESTIMATE. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1370E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .1762E-01

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 43.75

FIG. 48e. Summary of the negative exponential estimator (Gates et al. 1968) for Hemingway's (1971) data.



DATA TYPE	UNGROUPED DATA			
SAMPLE SIZE	73			
LINE LENGTH	60.00 KILOMETERS			
WIDTH	TRUNCATED DATA, WIDTH = 400.0 METERS			
PARAMETER	POINT ESTIMATE	STANDARD ERROR	PERCENT COEFF OF VARIATION	95 PERCENT CONFIDENCE INTERVAL
A(1)	.2957E-04	.5198E-05	17.6	.1938E-04 .3976E-04
F(0)	.6149E-02	.5288E-03	8.6	.5112E-02 .7185E-02
D	.3740E-01	.5433E-02	14.5	.2676E-01 .4805E-01

DENSITY(D) UNITS ARE NUMBERS/HECTARE

NOTES ON VARIANCE CALCULATIONS AND CONFIDENCE INTERVALS

ALL OF THE CONFIDENCE INTERVALS WERE CONSTRUCTED BY ASSUMING ASYMPTOTIC NORMALITY AND USING THE Z-VALUE OF 1.96. IT HAS BEEN ASSUMED THAT THE NUMBER OF OBSERVATIONS N IS A POISSON RANDOM VARIABLE. THUS THE VARIANCE WAS ESTIMATED AS VAR(N) = N = 73. AS EXPLAINED EARLIER IN THE OUTPUT THIS ASSUMPTION WAS MADE TO GET A VARIANCE FOR THE DENSITY ESTIMATE. THIS ASSUMPTION IS CRITICAL IF THE VARIANCE OF N IS THE PREDOMINANT PORTION OF THE VARIANCE OF DENSITY.

SQUARED COEFFICIENT OF VARIATION FOR N = .1370E-01

SQUARED COEFFICIENT OF VARIATION FOR F(0) = .7398E-02

PERCENT OF THE VARIATION OF DENSITY ATTRIBUTABLE TO THE SAMPLING VARIANCE OF N = 64.94

FIG. 48g. Summary of the half-normal estimator for Hemingway's (1971) data.

of animals away from the line creates a nonuniform distribution of animals with respect to the line. Some possible forms for such distributions are illustrated in Fig. 49 in contrast to the uniform distribution. The unobserved movement usually will cause the histogram of the observed perpendicular distances to appear as illustrated in Fig. 50 if the sample size is large (say greater than 150).

The general estimation theory is changed under those conditions. The average probability of observing an animal in the strip is

$$P_w = \int_0^w g(x)h(x|w)dx,$$

where $h(x|w)$ is the pdf of perpendicular distances. The distribution $h(x|w)$ is uniform under the usual estimation theory and the average probability is

$$P_w = \frac{1}{W} \int_0^w g(x)dx.$$

Assuming no movement, the essential

quantity $f(0)$ is related to P_w in the following manner,

$$f(0) = \frac{1}{wP_w} = \frac{1}{\int_0^w g(x)dx}$$

However, with movement the value of $f(0)$ is

$$f(0) = \frac{h(0|w)}{\int_0^w g(x)h(x|w)dx}$$

and hence it is not equal to $1/wP_w$. Rather the quantity which needs to be estimated is

$$f_m(0) = \frac{1}{w \int_0^w g(x)h(x|w)dx} = \frac{1}{wP_w}.$$

The value of $f_m(0)$ (the subscript m being used to denote movement) will almost always be greater than the value of $f(0)$. Therefore, estimators that fit the data more closely (i.e., estimate $f(0)$) will be more biased. This equation is discussed further in APPENDIX D and in Laake (unpublished thesis).

←

FIG. 48f. Fit of the negative exponential estimator on Hemingway's (1971) data. The perpendicular distance data have been grouped arbitrarily into 9 equal intervals for purposes of illustration. The estimation was done using the ungrouped data. A good fit is indicated ($\chi^2=8.01$, 7 df, $P=0.33$), however, a much more powerful test is possible in this case.

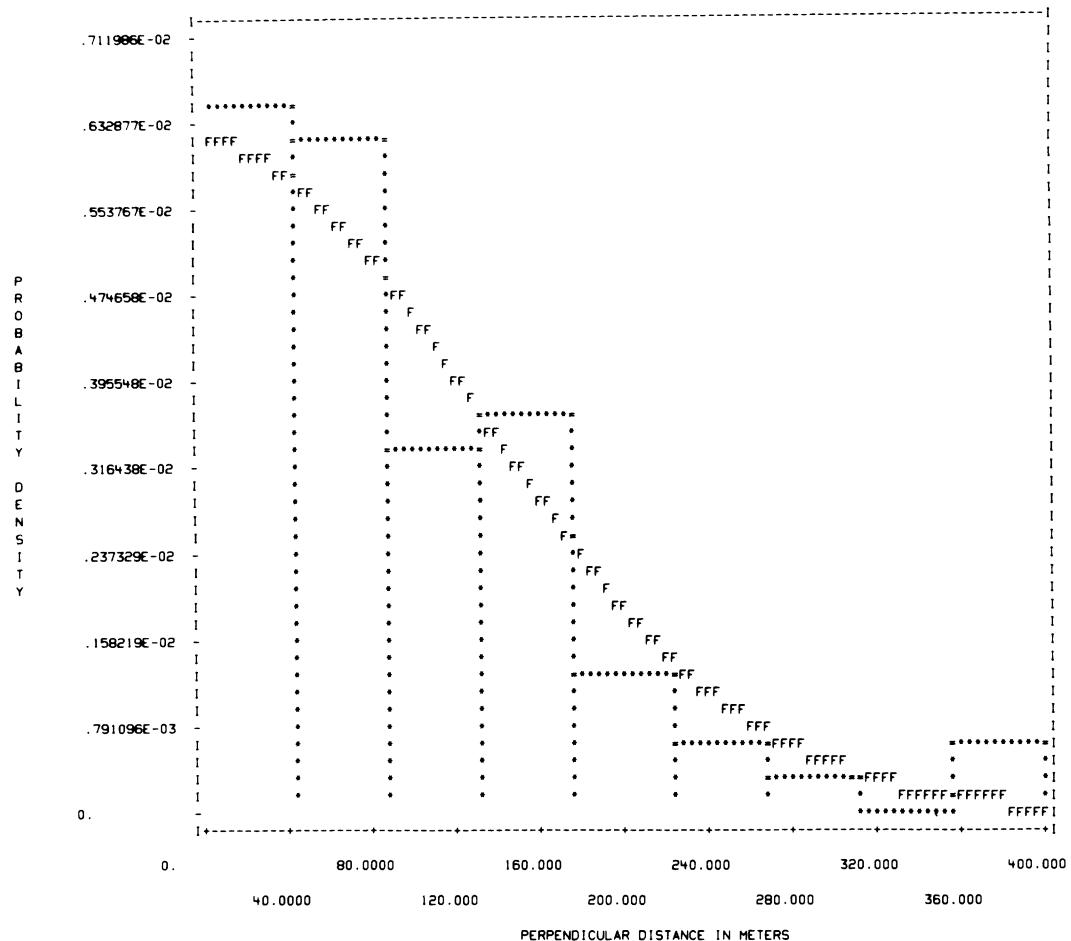


FIG. 48h. Fit for the half-normal estimator on Hemingway's (1971) data. The perpendicular distance data have been grouped arbitrarily into 9 equal intervals for purposes of illustration. The estimation was done using the ungrouped data. A good fit is indicated if the last 2 intervals are pooled ($\chi^2=2.95$, 5 df, $P=0.71$).

Recently, Smith (1979) studied the problem of estimation if animals move before being detected. Although his mathematical approach was more complex than ours, he derived the same basic result as above. He also showed that under certain conditions the general formula for estimation with animal movement is

$$D = \frac{E(n)f(0)}{2Lk}$$

(see APPENDIX D for further interpretation of the constant k). However, it is unlikely that k will ever be known and it cannot be estimated from the data. Also,

any assumption about k would be animal, habitat, and observer specific. The assumption would not be testable without further information on the original positions of the observed animals or the actual positions of all the animals. If that information were available, an assumption would not be necessary. For those reasons, the method proposed by Smith (1979) is of little practical value.

It is unforeseen that a method of deriving a reliable estimator of density will be developed for animal movement away from the line without making critical and untestable assumptions. Therefore, an examination of the robustness of existing

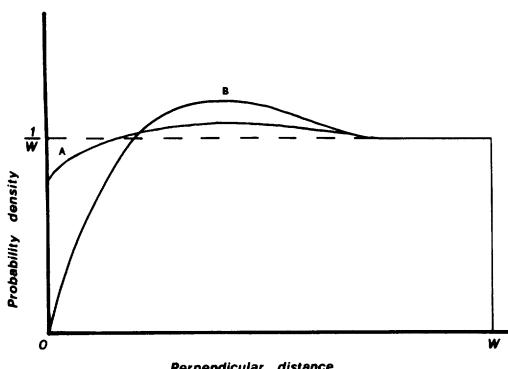


FIG. 49. Hypothetical forms for the pdf of the true (unobserved) distances, for animals moving away from the line. Movement in A is far less than that represented in B.

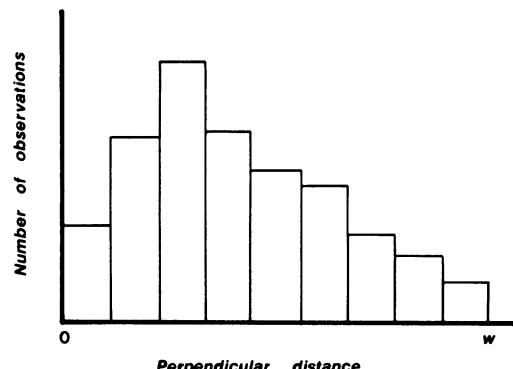


FIG. 50. Histogram of the observed perpendicular distances for a hypothetical population of animals moving some distance away from the line before being observed.

estimators to undetected movement is highly desirable.

Estimators Robust to Movement

Burnham and Anderson (1976) suggested that isotonic regression should perform well with movement. Since then, work suggested that the exponential polynomial (i.e., $g(x) = \exp(-(ax + bx^2))$) may also be robust to movement due to its monotone nature (Table 13). Therefore, those 2 estimators and the Fourier series were tested through computer simulation to examine their robustness with respect to animal movement away from the line. Simple parametric estimators were not examined because they do not perform well even for cases with no movement. It is believed that other generalized parametric estimators (e.g., exponential power series) will perform in a similar manner as the exponential polynomial.

Estimators based on ungrouped perpendicular distances were studied in an effort to find one that is particularly robust to animal movement. The robustness of an estimator can be determined by relaxing an assumption in an experimental situation and measuring the affect on the estimator in terms of bias and sampling variance. Ideally, the robustness of estimators to animal movement should

be evaluated using an animal population of known size. However, there would be many potential field problems, and the results would be applicable only to a particular species and habitat. Therefore, an alternative approach was used. Line transect sampling for animal populations that move away from the line was simulated on a digital computer. Using those simulated data, the biases were compared among several estimators.

The model used for simulating the animal movement problem was described by Laake (unpublished thesis), and all results presented here are from that source. Essentially, a stochastic model of animal movement was constructed that provided various forms of $h(x|w)$ that

TABLE 13.—SUMMARY OF ESTIMATORS THOUGHT TO BE ROBUST TO SMALL OR MODERATE MOVEMENT OF ANIMALS BEFORE DETECTION. ALL ESTIMATORS ARE MONOTONICALLY DECREASING PROVIDED THE PARAMETERS OF THE EXPONENTIAL POLYNOMIAL ARE APPROPRIATELY CONSTRAINED

Estimator	Functional form
Exponential polynomial	$g(x) = \exp[-(ax + bx^2)]$
Exponential power series	$g(x) = \exp[-(\lambda/x)^p]$
Reversed logistic	$g(x) = \frac{(1+b)\exp(-ax)}{1+b\exp(-ax)}$
Isotonic regression	None
One-term Fourier series	None

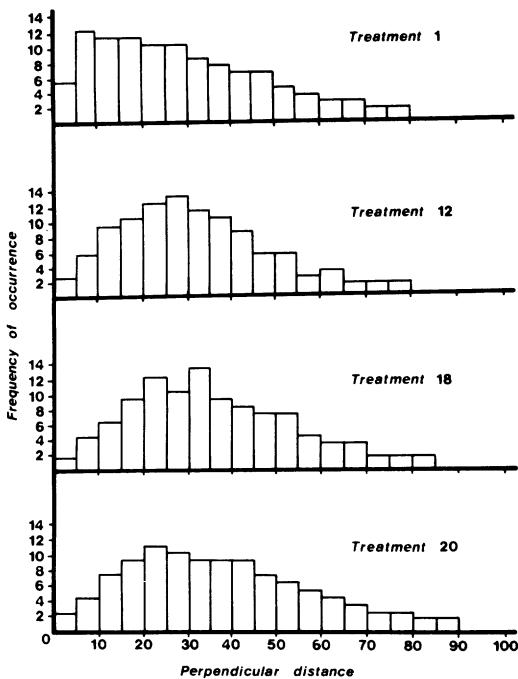


FIG. 51. Histograms of observed perpendicular distances for 4 treatments (different amounts of movement) from Experimental Situation I. The histograms are averages of the 25 replications for $n=100$.

represent "treatments." The model was based on the assumptions that an animal is more likely to move farther the closer it is to the line of travel of the observer. Then, 2 forms for detection functions were used to provide the observed data set: (1) half-normal and (2) negative exponential.

Four experimental situations were constructed on the basis of the detection function and the amount of movement away from the center line; note, the functions $h(x|w)$ (treatments) represent that movement. Various amounts of movement and an array of sample sizes were examined within the 4 experimental situations. The results from 2 of those experimental situations are given here: (1) Experimental Situation I used a half-normal detection function and (2) Experimental Situation IV used a negative exponential detection function. Twenty-five replications were created for each

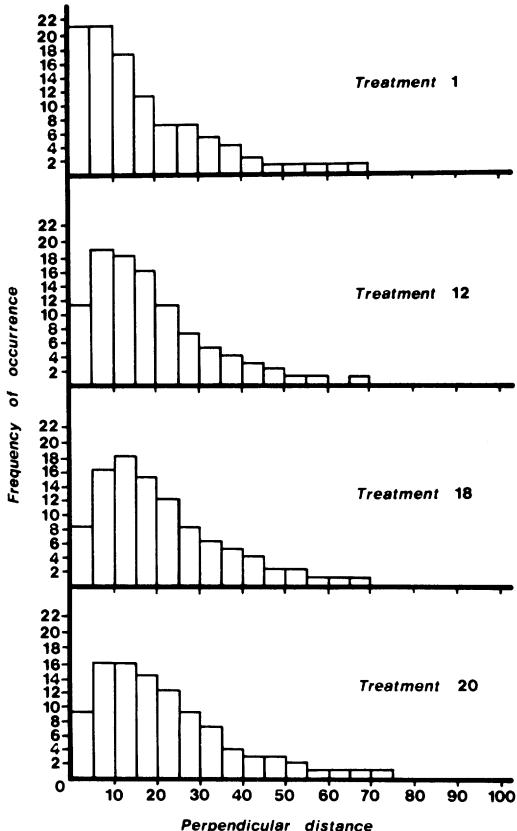


FIG. 52. Histograms of observed perpendicular distances for 4 treatments (different amounts of movement) from Experimental Situation IV. The histograms are averages of the 25 replications for $n=100$.

treatment within the experimental situation. Figs. 51 and 52 present the averaged histograms of observed perpendicular distances for 25 replications for 4 treatments.

Each of the 3 estimators was applied to the 25 individual data sets to make estimates of $f(0)$. They were compared to $f_m(0)$ which would provide an unbiased estimate of density and a percentage relative bias was calculated. The average percentage relative biases for the 25 replications are summarized in Tables 14 and 15. Examination of those tables reveals that the exponential polynomial and isotonic regression estimators are substantially more robust to movement than the Fourier series. Therefore, the

TABLE 14.—EXPERIMENTAL SITUATION I. AVERAGE PERCENTAGE RELATIVE BIAS FOR 25 REPLICATIONS

Movement ratio ¹ E(X)/E(Y)	Treatment ²	n = 50			n = 100		
		Exponential polynomial	Isotonic regression	Fourier series	Exponential polynomial	Isotonic regression	Fourier series
0.05	1	-9.7	-3.0	-21.1	-8.1	-10.9	-19.8
	2	-5.1	-6.7	-29.6	-7.9	-16.0	-25.3
	3	-8.5	-9.5	-25.0	-10.5	-16.7	-24.6
	4	-8.4	-12.1	-34.4	-9.4	-16.4	-43.1
	5	-6.8	-10.7	-38.3	-7.6	-13.5	-34.5
	6	-7.9	-11.6	-30.8	-10.0	-16.6	-41.6
	7	-10.8	-13.7	-27.3	-9.8	-16.7	-33.7
	8	-4.6	-6.3	-17.7	-8.1	-14.9	-26.2
	9	-9.7	-12.5	-24.2	-9.3	-16.5	-27.4
	10	-5.1	-5.9	-16.6	-7.1	-12.5	-29.7
0.10	11	-12.1	-15.8	-23.0	-14.3	-23.1	-28.9
	12	-17.7	-23.9	-60.8	-17.6	-27.7	-75.3
	13	-21.3	-27.0	-68.2	-19.8	-28.1	-62.0
	14	-19.1	-25.6	-61.7	-18.4	-27.1	-76.3
	15	-16.3	-23.9	-51.9	-17.9	-27.4	-70.9
	16	-16.0	-23.3	-56.9	-15.5	-24.1	-72.3
0.20	17	-28.2	-34.2	-69.2	-27.9	-35.8	-70.6
	18	-36.2	-42.7	-89.6	-34.5	-42.8	-90.4
	19	-32.1	-38.1	-81.9	-31.4	-40.3	-87.9
0.40	{ 20	-48.1	-52.8	-87.4	-48.5	-54.1	-87.9

¹ An index to the relative severity of movement; e.g., 0.05 indicates relatively little movement while 0.40 indicates extreme movement.² Various combinations of the movement function and the detection function.

exponential polynomial is recommended in cases where some, but not much, movement is suspected as the sample was taken.

Monotonically Decreasing Estimators

The exponential polynomial and isotonic regression estimators are fundamentally different from the Fourier series estimator. The former estimators assume that the detection function is monotonically decreasing and they fit the data based on that assumption; the Fourier series, however, is not constrained to be monotonically decreasing. The exponential polynomial estimator is constrained to meet the *shape criterion* and monotonicity by constraining the parameters to be greater than or equal to zero. The isotonic regression estimator is constructed to provide a monotonically decreasing step function as a representation of the underlying function. The first step in the function ranges from 0 to $X_{(i)}$ and the value of the function for the first step

is $i/(nX_{(i)})$. The value of i is chosen to maximize the value of the function. If that were continued for the rest of the sample, it would provide a monotonically decreasing step function. Hayes (unpublished thesis) suggested a grouped analog of isotonic regression (see PART 4) as a modification to Emlen's (1971) method. He suggested averaging the numbers of observations in the intervals near the origin and using the maximum average to estimate what is essentially $f(0)$. This is exactly what the isotonic regression estimator is doing with continuous data.

On theoretical grounds, the exponential polynomial is a preferred estimator in comparison to isotonic regression. Isotonic regression has several practical problems. It cannot handle any zero values for perpendicular distances because division by zero is undefined. Also, simulation studies by Hayes (unpublished thesis) and in this monograph have shown when there is no movement (i.e., all assumptions met) isotonic regression may be biased by an order of magnitude.

TABLE 15.—EXPERIMENTAL SITUATION IV. AVERAGE PERCENTAGE RELATIVE BIAS FOR 25 REPLICATIONS

Movement ratio ¹ E(X)/E(Y)	Treatment ²	n = 50			n = 100		
		Exponential polynomial	Isotonic regression	Fourier series	Exponential polynomial	Isotonic regression	Fourier series
0.05	1	-28.3	-19.3	-37.6	-29.1	-32.3	-37.2
	2	-38.1	-38.1	-44.1	-37.3	-41.8	-45.1
	3	-36.0	-34.1	-42.2	-32.2	-37.9	-41.6
	4	-49.8	-48.7	-53.3	-49.9	-50.7	-52.6
	5	-49.8	-47.4	-53.0	-47.0	-49.3	-51.1
	6	-44.5	-45.9	-49.5	-45.7	-48.1	-50.7
	7	-51.1	-50.5	-54.1	-50.5	-52.6	-55.3
	8	-48.0	-48.8	-52.6	-49.7	-49.9	-53.3
	9	-45.7	-47.5	-51.8	-47.0	-49.4	-51.5
	10	-46.5	-45.4	-50.2	-42.1	-46.6	-48.3
0.10	11	-41.7	-41.6	-47.7	-41.2	-45.1	-48.8
	12	-54.5	-53.8	-57.5	-55.1	-56.7	-59.0
	13	-52.2	-49.0	-54.7	-51.5	-53.3	-55.7
	14	-58.3	-59.5	-62.5	-60.8	-61.8	-64.2
	15	-56.9	-54.2	-60.9	-56.4	-57.3	-58.7
	16	-54.6	-54.7	-57.9	-54.8	-56.4	-59.5
0.20	17	-57.4	-54.2	-60.7	-56.4	-58.0	-59.3
	18	-66.1	-64.9	-71.9	-66.6	-67.3	-69.3
	19	-62.7	-63.0	-65.7	-63.5	-64.9	-66.1
0.40	{ 20	-68.6	-67.7	-72.7	-70.1	-70.4	-72.4

¹ An index to the relative severity of movement; e.g., 0.05 indicates relatively little movement while 0.40 indicates extreme movement.² Various combinations of the movement function and the detection function.

Such a bias occurs because the function is often maximized by the first order statistic, and as sample size n gets large, the first order statistic approaches zero so that the estimate of $f(0)$ approaches infinity. The grouped analog of isotonic regression as suggested by Hayes (unpublished thesis) avoids the problem but it requires arbitrary grouping of the data.

The exponential polynomial estimator is only one of many possible estimators that can be constrained to be monotonically decreasing. The exponential power series (Pollock 1978) is another monotonically decreasing estimator. It was tested in some initial investigations and performed as well as the exponential polynomial. However, it has severe numerical problems and requires an excessive amount of computing time. Pollock (pers. comm.) suggested that if λ is constrained to be greater than 1, the exponential power series estimator may be much more satisfactory. The reversed logistic (Eberhardt 1978a) and other generalized parametric estimators may also perform as

well as the exponential polynomial in providing an estimator robust to slight amounts of movement.

The assumption that the function is monotonically decreasing is a distinct advantage when animals move away from the line. As illustrated by Figs. 51 and 52, the underlying pdf of observed perpendicular distances (represented by the histogram) may dip at the origin. Therefore, any estimator that is not constrained will estimate $f(0)$ more closely. However, $f(0)$ is negatively biased in relationship to the quantity $f_m(0)$, that provides an unbiased estimate if w is large. Estimators like those based on the exponential polynomial and isotonic regression which are constrained will provide less biased results. This is clearly illustrated by Figs. 53 and 54. Those figures compare the fit to the data provided by the exponential polynomial and the Fourier series. Fig. 53 corresponds to Experimental Situation I in which a half normal detection function was used. In that case, there is a large discrepancy in the estimates. Fig.

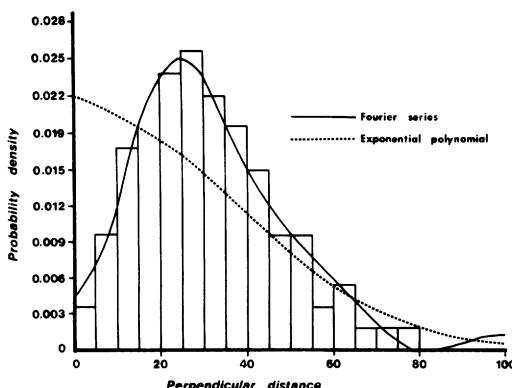


FIG. 53. The Fourier series and exponential polynomial fit to the average histogram of Treatment 12 of Experimental Situation I.

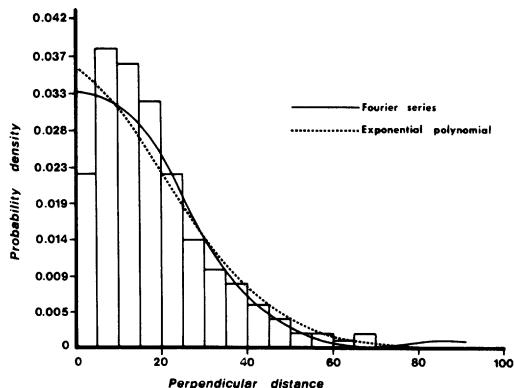


FIG. 54. The Fourier series and exponential polynomial fit to the average histogram of treatment of Experimental Situation IV.

54 corresponds to Experimental Situation IV in which a negative exponential detection function was used. The underlying form of $f(x)$ is quite different in comparison to Fig. 53 even though they represent the same treatment of movement. The 2 estimates shown in Fig. 54 are not very different because the shape of the pdf of observed perpendicular distances does not deviate greatly from a monotonically decreasing function. Finally, it is worth mentioning that the Fourier series estimator performs better under movement if constrained to only 1 term. That enforces the monotonicity, still it is not as good as the exponential polynomial.

Criticalness of Movement Assumption

The simulation study does not address the criticalness of the assumption of no movement. It would require knowledge about the validity of the model and the amount of movement occurring in the real world. Therefore, the question of how much bias can be expected is left unanswered. However, an example will be given to put the simulation study into perspective. Consider an investigator studying rabbits, where the detection function might be approximately half-normal. Assume the investigator will be observing rabbits to a width of 100 m. If

30 percent of the animals were moving and the overall average amount of movement was 1.3 m, it could correspond to Treatment 3 of Experimental Situation I. On the basis of the computer simulations, a relative negative bias of 6–9 percent would be expected if the exponential polynomial estimator was used. However, if 30 percent of the animals moved an average of 12 m before detection, the PRB is approximately 32.

The information needed to determine the bias to expect could be gathered possibly from field experiments with an animal population of known size. That would not be a satisfactory approach for some animals such as porpoise; but for others like rabbits it may be a viable one. Eberhardt (1978a) has suggested using radiotelemetry to monitor animal movement; however, it may suffer from the lack of resolution in the measurements. If animals were contained in a large pen and an observation tower was available, information could be gathered about the response of the animal to the observer and the amount of unobserved movement that occurs. However, information is also needed about the detection function for the observer. It was illustrated previously that the detection can have a significant affect on the bias present in the estimate. This is rather disturbing because the estimate of density is no longer

independent of the observer and habitat when animals move away from the line.

It is important to realize that no information about the amount of movement and relative bias can be gained from the histogram of the observed perpendicular distances. The functions $h(x|w)$ and $g(x)$ are completely confounded in the resulting distribution of observed perpendicular distances. This can be illustrated with a simple example using grouped data for 2 intervals of perpendicular distance. The animals have moved away from the line so there are 5 animals in the first interval and 10 in the second. If the detection function were a half-normal, the probabilities could be 0.9 and 0.8 for detecting the animals in the 2 intervals. Then it would be expected that 4.5 and 8.0 animals would be seen in the intervals on the average. However, if the probabilities were 0.9 and 0.5, which could represent a negative exponential, then 4.5 and 5.0 animals would be seen on average in the 2 intervals. Thus, it would appear that more movement was occurring in the instance where the detection function was a half-normal. The histograms in Figs. 53 and 54 illustrate this point. The estimate for the exponential polynomial is actually less biased for Fig. 53 rather than Fig. 54. It is also possible that a histogram could appear monotonically decreasing and yet the estimate would be biased. That occurred in Treatment 1 of Experimental Situation IV as illustrated in Fig. 52. On the other hand, it is also possible that a histogram could appear as if movement was occurring when it really was not. This could be caused by random variation if the sample size was small (e.g., see APPENDIX B); or, would result if the underlying detection function was not monotonically decreasing with $g(0)$ less than unity. However, in both cases, the use of a monotonically decreasing estimator can be expected to provide better results.

Movement of animals from their original location before detection represents a potentially serious problem in line transect sampling. If the sample size is rela-

tively large (say $n > 150$), then undetected movement may be revealed in the histogram of the perpendicular distance data (see Fig. 50). If the sample size is small, then little can be said (see APPENDIX B). Regardless of sample size, the amount or importance of undetected movement cannot be assessed from the histogram (i.e., $h(x|w)$ and $g(x)$ or confounded). If undetected movement is relatively minor, then the use of an estimator based on a monotonically decreasing function will minimize bias in \hat{D} and probably be satisfactory. If undetected movement is moderate or severe then substantial bias can result.

Recommendations

Two approaches are recommended to reduce the bias incurred from animal movement away from the transect line: (1) use a monotonically decreasing estimator like the exponential polynomial, and (2) develop field procedures which minimize the amount of unobserved movement occurring. The first approach is self-explanatory but the second is somewhat more difficult to implement.

Field procedures that maximize the chance of seeing an animal before or shortly after it moves will minimize the amount of movement and will reduce the bias directly. Those procedures will depend on the behavior of the animal; therefore, specific recommendations cannot be given here. However, several general suggestions will be given as examples. If animals respond to the observer at distances greater than the maximum possible observation distance, a faster pace of travel may reduce the amount of movement that occurs before the animals are observed. This seems to work for aerial surveys of porpoise schools. A similar suggestion might involve searching further ahead and concentrating on the center line rather than laterally. This would allow the observer to detect animals immediately after they respond and move. If 2 observers were available, one could search close to the line and the other

TABLE 16.—SUMMARY OF CAPABILITIES AVAILABLE IN PROGRAM TRANSECT

Estimator	Ungrouped data	Grouped data	Untruncated data	Truncated data
Fourier series	X	X	X ¹	X
Exponential polynomial	X	X	X	X
Exponential power series	X	X	X ¹	X
Half-normal	X	X	X	X
Negative exponential	X	X	X	X
Hayne	X		X	
Modified Hayne	X		X	
Generalized Hayne	X		X	

¹ The largest observation is used for the width of the transect.

could search at distances farther down the line to prevent unobserved movement. Experience in the field with a particular animal will permit better judgements to be made about specific field procedures that will minimize movement.

COMPREHENSIVE COMPUTER PROGRAM

We developed a comprehensive computer algorithm, TRANSECT, to allow biologists and statisticians to analyze line transect data easily, accurately, and thoroughly. Program documentation is given by Laake et al. (1979). A careful analysis of line transect data requires many tedious calculations (e.g., estimates of parameters and tests of hypotheses), especially for grouped or truncated data. In addition, various plots and graphs are important aids in the analysis. For those reasons, we discourage analysis of line transect data by hand. The program is relatively inexpensive and easy to use and should be a convenient analytic tool.

The basic capabilities of program TRANSECT are summarized in Table 16. The FS is the default estimator; and the other estimators are available as options. The estimation theory used in TRANSECT is given in PART 3. Exact ML estimators are derived for all the parametric models; no approximations are used. In each case, the parameter estimates are computed and printed, as are the corresponding estimates of sampling standard errors and correlations. In general, \hat{a}_i , $\hat{f}(0)$, and \hat{D} are printed with their

standard error, coefficients of variation, and confidence intervals.

In addition, simple options enable the display of a series of cdf plots, histograms, scatter diagrams, and other summaries. Finally, a comprehensive series of tests are computed to assess crucial assumptions and goodness of fit. The figures in ILLUSTRATIVE EXAMPLES give insight into the output from TRANSECT. Further options and use of the program are explained in Laake et al. (1979).

The program is written in ANSI FORTRAN IV, the most common scientific programming language. A magnetic tape with the FORTRAN code and all the examples treated under ILLUSTRATIVE EXAMPLES is available from SHARE Program Library Agency, P.O. Box 12076, Research Triangle Park, N.C. 27709 at a cost of approximately \$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.

The program consists of a main routine and 55 subroutines. Comment statements document major functions of the program, which is approximately 7,200 FORTRAN statements. The program requires about 150K of core storage without an overlay structure on a CDC 6600 but could be reduced considerably by using overlays. The program accepts line transect data on up to 100 individual lines and a total of 1,600 objects detected. Those limitations can be relaxed easily by modifying the program.

The cost of compiling the program is in

the \$10-\$20 range, depending on the computer and local charging schemes. Analysis of an "average" data set using several of the options available ranges from \$1 to \$5. The program will run on most medium or large computers with little or no modification.

DISCUSSION AND SUMMARY

Discussion of Assumptions

The usefulness of transect sampling in a given problem depends on whether or not the basic assumptions of the method can be met. Below, the 4 assumptions and the effect of their failure are reiterated.

The most critical assumption is that if an object is on the line of travel, it will be seen with probability 1 (i.e., $g(0) = 1$). Failure of that assumption means $g(0) < 1$, and objects are missed. It could result from poor field methods, or attempting the method on an unsuitable species. If \hat{D} is an unbiased estimator under the assumption ($g(0) = 1$), failure of the assumption biases \hat{D} simply as, $E(\hat{D}) = Dg(0)$. If, for example, there is only a 70 percent probability of detecting an object on the line, that alone causes a 30 percent relative bias in the estimator. Failure of the assumption cannot be detected by any form of data analysis, consequently, it must be evaluated based on knowledge of the intended application alone.

The second most important assumption is that objects (animals) do not move in response to the observer before being detected. Evasive movement can have a devastating effect on reliability of the line transect method. For example, if 30 percent of the animals moved so as to be undetectable (but the remaining 70% satisfied the assumption), one could have $E(\hat{D}) = 0.7D$. In a sense, evasive movement violates the assumption of $g(0) = 1$ (certainly its effect on the estimator bias is very similar). Extreme movement may sometimes be evident from examination of a histogram of the data, however, moderate movement cannot be identified re-

liably that way (because all assumptions can be right yet, by chance, the first cell or 2 of such a histogram may have a lower count than the third cell, see APPENDIX B). It is concluded that the assumption of no evasive movement (or attraction) must be evaluated based on knowledge of the intended application.

Movement that would result in an animal being counted more than once will cause a positive bias on \hat{D} . That could happen if a species flushes and flees ahead of the observer, only to be flushed again further down the line. The investigator must assess that possible problem.

The last 2 assumptions are not as critical as the first 2. In particular, the assumption of accurate recording of data (distances and angles) is under at least partial control of the investigator. This is one area of field methodology where much innovation and improvement is possible. Systematic errors in recording data can cause biased estimates. The critical data to record accurately are short perpendicular distances; yet that is where we have observed frequent rounding errors. That is, objects near, but not on the line have their perpendicular distances recorded as zero. Such a systematic error tends to cause a positive bias in estimated density. (Grouping of data in the analysis can reduce some of the bias.) Errors in distances of objects far from the line are much less critical.

If the data suffer small, random errors in measurement, the estimate of population abundance will not be biased, but its sampling variance will be increased (i.e., \hat{D} will be less precise). That might happen with animals that flush, and it is difficult to know their original location precisely.

Statistical independence of detection is only used in deriving the sampling variance of \hat{D} , given by Eq. (1.17). That assumption can fail if detection of one individual influences (increasing or decreasing) the probability of detecting other individuals close by. (One bird fleeing may alarm others and cause a chain reaction.) If the assumption fails, the sam-

pling variance of \hat{D} obtained by Eq. (1.17) is likely to be an underestimate of true variance. The entire problem is avoided if the sampling variance of \hat{D} is derived by replicate lines, that is from Eq. (1.22).

Other authors often invoke several more assumptions (e.g., random distribution of objects, population response behavior, being constant during the study). Those assumptions are not needed provided the study is properly designed and robust data analysis procedures are used. Essentially, random line placement replaces random object distribution, and use of the Fourier series estimators eliminates the need for assuming an unchanging detection function during the study.

Summary of Important Points

This section summarizes some of the key points in PARTS 1 and 2.

1. In line transect sampling, the survey must be conducted in such a way as to: (a) assure that all objects on the center line are detected, (b) determine accurately the location of the (straight) center line once an object is detected, and (c) measure accurately at least the perpendicular distance from the center line to the object detected.
 2. Field procedure for line transect sampling are summarized in the section entitled, FIELD SAMPLING PROCEDURES AND STUDY DESIGN.
 3. It is important to realize that line transect sampling allows objects away from the center line to go undetected. The fact that objects away from the center line go undetected or the reasons why they were undetected are not important.
 4. In general, it is recommended that estimates of density be based on ungrouped perpendicular distances. However, estimation from grouped data is sometimes advantageous and this theory is developed.
 5. The general form of the estimator of density for perpendicular distance data (grouped or ungrouped) is
- $$\hat{D} = \frac{n\hat{f}(0)}{2L}.$$
6. The use of the Fourier series estimator is recommended to obtain $\hat{f}(0)$ as an omnibus procedure. It is *model robust, pooling robust*, has the *shape criterion*, and has high *estimator efficiency* for sample sizes typically encountered.
 7. Our simulation studies and investigations of robustness and estimation efficiency indicate that the Fourier series method is generally superior to other methods.
 8. Estimates of density with the Fourier series, sampling variances and covariances, and standard errors can readily be computed by using 7 simple equations (2.1–2.7) if the data are not grouped.
 9. A comprehensive computer program, TRANSECT, is available as a powerful tool in the analysis of line transect data (Laake et al. 1979). TRANSECT is inexpensive, accurate, and easy to use. We recommend its use rather than attempting an analysis by hand.
 10. Program TRANSECT enables a thorough analysis of data under the Fourier series estimation procedure for both grouped and ungrouped data. A number of other estimation schemes are also available as options in the program. In addition, a series of data plots, histograms, and other options are available through program TRANSECT.
 11. Many examples of the rigorous analysis of line transect data are given under ILLUSTRATIVE EXAMPLES and readers are encouraged to review that material carefully.
 12. Using sighting distances and sighting angles as a basis for estimating density is not recommended.
 13. If members of the population under study commonly move substantially before being detected, substantial bias will result. Line transect sampling should not be used in those cases. If movement is less severe, an estimator based on a monotonically decreasing model is preferred (e.g., exponential quadratic).

PART 3

STATISTICAL ESTIMATION THEORY FOR ROBUST METHODS

This part of the monograph presents a summary of the statistical theory underlying the methods for estimating $f(0)$ that have been investigated during our research. Not all of those methods were subsequently incorporated into program TRANSECT or otherwise recommended for use. This material is oriented to the more quantitative reader, especially those who will pursue additional mathematical research on line transects. We assume the reader is familiar with basic mathematical statistics.

Broadly speaking, 2 types of models have been examined for $f(x)$: (1) models that are proper probability density functions, and (2) functions that are used to approximate $f(x)$, but need not be pdfs because they may be negative, although they do integrate to 1. Models in the second category that were considered are all linear models of the type

$$f(x) = \frac{1}{w} + \sum_{j=1}^m a_j \phi_j\left(\frac{x}{w}\right), \quad 0 < x < w$$

where the $\phi_j(\cdot)$ are known functions integrating to zero on $(0, w)$.

Maximum likelihood (ML) theory is recommended for inference with models that are true pdfs. One would use ML inference with other types of models (e.g., Fourier series), except that problems can arise because those models can take on negative values. ML estimation was tried for the Fourier series (FS) model with ungrouped data and found to not always work (one often encounters logarithms of negative numbers). Therefore, it is recommended that a "method of expectations" approach be used with linear models that need not be pdfs when data are ungrouped. Of the possible models in that category only the FS has been developed fully. For grouped data, our experience has been that the cell probabilities will be positive, hence ML inference is possible; we used it and recommend

it, even for models such as the FS with grouped data.

In what follows, the theory of numerical ML estimation used in program TRANSECT is summarized. Once the general theory is understood, treatment of any model that is a proper pdf is easily stated. Next, the theory for the FS estimator followed by a consideration of other linear models is presented. Then, application of the general theory to the specific parametric models examined is discussed. Finally, the specifics of the theory for the modified and generalized Hayne estimators are discussed.

NUMERICAL MAXIMUM LIKELIHOOD ESTIMATION

Ungrouped Perpendicular Distance Data

Let $f(x, \theta)$, $0 < x < w$ be a pdf of known form based on an unknown parameter θ which may be vector valued, i.e., $\theta' = (\theta_1, \dots, \theta_r)$, $r \geq 1$. It is assumed there is a well-defined parameter space. Also, w may be infinite or finite (if finite, $f(x, \theta)$ is often the truncated version of some pdf on 0 to ∞). The general theory of ML estimation is standard, and may be found in any text on mathematical statistics (e.g., Rao 1973).

Let x_1, \dots, x_n be a random sample from $f(x, \theta)$. Then the likelihood function is $\mathcal{L}(\theta) = \prod_{i=1}^n f(x_i, \theta)$. For the types of well-behaved models one normally uses, the ML estimator $\hat{\theta}$ can be found as the solution to the equations

$$g_j(\theta) = \frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta_j} = 0 \quad j = 1, \dots, r.$$

Those equations often require numerical solution, particularly for any truncated distribution ($w < \infty$), and for many 2 parameter models.

The asymptotic variance-covariance matrix of $\hat{\theta}$ is $I^{-1}(\theta)/n$, where the elements of this Fisher information matrix are given by

$$I_{jh}(\theta) = -E\left[\frac{\partial^2 \ln(f(x, \theta))}{(\partial \theta_j)(\partial \theta_h)}\right]$$

$$j, h = 1, \dots, r.$$

Analytical evaluation of $I(\theta)$ involves integration over the sample space, and is not always possible. An empirical estimator must then be used, namely, the Hessian matrix with elements

$$H_{jh}(\hat{\theta}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln(f(x_i, \hat{\theta}))}{(\partial \theta_j)(\partial \theta_h)}$$

$$h = 1, \dots, r.$$

In what follows, it is assumed that the Hessian matrix will be used if the information matrix is not known analytically.

The ML estimate $\hat{f}(0) = f(0, \hat{\theta})$, and its asymptotic sampling variance is

$$\text{var}(\hat{f}(0)) = \delta' I^{-1}(\theta) \delta,$$

$$\delta' = \left(\frac{\partial f(0, \theta)}{\partial \theta_1}, \dots, \frac{\partial f(0, \theta)}{\partial \theta_r} \right). \quad (3.1)$$

For some models, the ML estimators and their sampling variances exist in closed form (e.g., the untruncated negative exponential and half-normal). However, ML inference for most other models will require numerical procedures. A hybridization of the Newton-Raphson and Marquardt procedures generally works well if starting values are at all reasonable (for information on those 2 separate techniques see Aoki (1971) and Marquardt (1963)). The numerical method used in TRANSECT is given below.

Let $\hat{\theta}_0$ be a starting value for that iterative technique. For the i^{th} iteration, the updated estimate is

$$\hat{\theta}_{i+1} = \hat{\theta}_i + (nI(\hat{\theta}_i))^{-1}g(\hat{\theta}_i). \quad (3.2)$$

If θ is vector valued, then $g(\theta) = (g_1(\theta), \dots, g_r(\theta))'$. At each step of the iteration, 2 conditions must be satisfied:

(a) $\hat{\theta}_{i+1}$ is in the parameter space and (b) $\ln \mathcal{L}(\hat{\theta}_{i+1}) > \ln \mathcal{L}(\hat{\theta}_i)$. If (a) and (b) are both

true, one continues the iteration using the Newton-Raphson method, Eq. (3.2). The iterations terminate when the relative change in each parameter is small. If θ is a scalar ($r = 1$), the condition is

$$\frac{|\hat{\theta}_{i+1} - \hat{\theta}_i|}{|\hat{\theta}_i|} < \varepsilon, \quad (3.3)$$

otherwise, the condition is applied to each component of θ . A plausible value of ε is 0.00001. In practice, it is recommended that the convergence criterion be adjusted for sample size. Thus one uses $\varepsilon = \varepsilon^*/n$, for say $\varepsilon^* = 0.001$.

If one or both of conditions (a) and (b) are not satisfied, $\hat{\theta}_{i+1}$ is not an acceptable next step. Then, a modified Marquardt procedure is suggested wherein $\hat{\theta}_{i+1}$ is found as

$$\hat{\theta}_{i+1,k} = \hat{\theta}_i + (nI(\hat{\theta}_i) + \lambda^k A)^{-1}g(\hat{\theta}_i),$$

where λ is some small number like 2 or 4 (4 works well) and A is a diagonal matrix of just the diagonal elements of $nI(\hat{\theta}_i)$. One takes the sequence $\hat{\theta}_{i+1,k}$, $k = 1, 2, \dots$, until conditions (a) and (b) are both first met. The resulting $\hat{\theta}_{i+1,k}$ is taken as the updated estimate, $\hat{\theta}_{i+1}$.

Theoretically, this procedure should always work when there is a unique solution to the problem. In practice it may fail; one then should try a different starting value. If a very ill-conditioned model is used for $f(x, \theta)$, a more robust numerical maximization procedure may be needed (the exponential power series model for shape parameter $p < 1$ is such a problem model).

When that procedure has converged to the ML estimate $\hat{\theta}$, the estimated information matrix is also immediately available. Using $\hat{\theta}$ and $I(\hat{\theta})$ in Eq. (3.1), one obtains the estimated sampling variance of $\hat{f}(0)$.

Grouped Perpendicular Distance Data

Let the perpendicular distance data be grouped into k intervals defined by cut points, $0 \equiv c_0 < c_1 < c_2 < \dots < c_k \equiv w$.

There still exist an underlying, continuous detection function and corresponding pdf $f(x)$. Each detected object will fall into one of the k mutually exclusive intervals. Under the assumption that each detection is an independent event, the frequency counts by interval are multinomial random variables:

$$\Pr\{n_1, \dots, n_k\} = \frac{n!}{(n_1)! \dots (n_k)!} \cdot \prod_{i=1}^k (p_i)^{n_i}, \quad (3.4)$$

$$p_i = \int_{c_{i-1}}^{c_i} f(x) dx.$$

If $f(x)$ is a known function of an unknown parameter θ , the cell probabilities are known functions of θ , say $p_i(\theta)$, $i = 1, \dots, k$.

Equations for obtaining exact ML estimates and their large sample variance covariance matrix are well known for the general multinomial distribution (e.g., Rao 1957, 1973). Assuming suitable regularity conditions, the ML estimators are the solution to the likelihood equations

$$g_j(\theta) = \frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta} = \sum_{i=1}^k \frac{n_i}{p_i} \left(\frac{\partial p_i}{\partial \theta_j} \right) = 0$$

$$j = 1, \dots, r.$$

The Fisher information matrix $I(\theta)$ has element j,h as

$$I_{jh}(\theta) = \sum_{i=1}^k \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \theta_j} \right) \left(\frac{\partial p_i}{\partial \theta_h} \right)$$

$$j, h = 1, \dots, r.$$

The solution of the likelihood equations usually requires iterative numerical methods. Fisher's method of scoring was used in program TRANSECT (see Rao 1973), with the modification of going to a Marquardt procedure when necessary. The numerical procedure used to find $\hat{\theta}$ for ungrouped data is identical to the procedure used with grouped data, except that the analytic formula for the information matrix is known. Consequently, the procedure is not repeated here. When convergence is achieved, the estimate $\hat{\theta}$

and its estimated variance-covariance matrix are available. Then, $\hat{f}(0)$ is readily obtained and, based on Eq. (3.1), $\hat{\text{var}}(\hat{f}(0))$ can be obtained.

It is worth noting that the cell probabilities can be written as $p_i = F(c_i) - F(c_{i-1})$, where $F(c_i)$ is the cdf corresponding to $f(x, \theta)$. The various partials of the p_i thus requires only partials of $F(x)$. For example,

$$\frac{\partial p_i(\theta)}{\partial \theta_j} = \frac{\partial F(c_i)}{\partial \theta_j} - \frac{\partial F(c_{i-1})}{\partial \theta_j}.$$

Implementation of numerical ML estimation for the grouped data case is seen, therefore, to require only equations (or methods) to compute 2 types of quantities: $F(x)$ and $\partial F(x)/\partial \theta_j$, $j = 1, \dots, r$, and these are needed only at the cut point values.

FOURIER SERIES

Ungrouped Data

The theory behind the FS estimator has been published by Crain et al. (1978); hence only the major aspects of the theory will be given. The FS estimator is based on the concept of a FS expansion (Churchill 1963, Tarter and Kronmal 1976) of a pdf over the finite interval 0 to w . The even extension $\phi(x)$ of $f(x)$ from $[0, w]$ to $[-w, w]$ is defined as

$$\phi(x) = \begin{cases} f(x) & 0 \leq x \leq w \\ f(-x) & -w \leq x \leq 0 \end{cases}$$

(see Churchill 1963).

Assume that $\phi(x)$ may be expanded in a Fourier series over $[-w, w]$. Then,

$$\phi(x) = \frac{1}{w} + \sum_{j=1}^{\infty} \left\{ a_j \cos \left(\frac{j\pi x}{w} \right) + b_j \sin \left(\frac{j\pi x}{w} \right) \right\}$$

$$-w \leq x \leq w,$$

where the coefficients a_j and b_j are given by

$$a_j = \frac{1}{w} \int_{-w}^w \phi(x) \cos \left(\frac{j\pi x}{w} \right) dx$$

$$= \frac{2}{w} \int_0^w f(x) \cos\left(\frac{j\pi x}{w}\right) dx,$$

$$b_j = \frac{1}{w} \int_{-w}^w \phi(x) \sin\left(\frac{j\pi x}{w}\right) dx = 0.$$

j = 1, 2, 3, . . . (3.6)

We have used the facts that the sine functions are odd, the cosine functions are even, and $\phi(x)$ is even by construction. It follows that

$$f(x) = \frac{1}{w} + \sum_{j=1}^{\infty} a_j \cos\left(\frac{j\pi x}{w}\right) \quad 0 \leq x \leq w.$$

Finally, because $\cos(0) = 1$,

$$f(0) = \frac{1}{w} + \sum_{j=1}^{\infty} a_j \doteq \frac{1}{w} + \sum_{j=1}^m a_j, \quad (3.5)$$

where m is a truncation point to be selected. It can be shown that a sufficient condition for the representation $f(0) = 1/w + \sum_{j=1}^{\infty} a_j$ is that $f(x)$ be continuous (from the right) at $x = 0$ and monotonically nonincreasing over $[0, w]$. Because $f(x) = g(x)/\mu$ and $g(x)$ are monotonically nonincreasing by definition, the only assumption needed for the validity of the FS method is continuity of $f(x)$ at $x = 0$.

It should be remarked at this point that $f(x)$ could have been expanded in some other type of orthogonal series. However, in many applications the detection function $g(x)$ and hence $f(x)$ may resemble a low order cosine term, especially when x is near zero. Also, $f'(0) = 0$ and $f''(0) < 0$ seem to be reasonable assumptions to impose on line transect sampling, and those conditions are met by cosine terms (also see Eberhardt 1978a:11).

The FS model is not a true pdf because it is possible for $f(x)$ to be negative, especially for x near w . That makes it impossible to use ML estimation routinely with the FS model. A method of expectations approach may be based on the observation that

$$a_j = \frac{2}{w} E\left(\cos\left(\frac{j\pi x}{w}\right)\right).$$

Therefore, an unbiased estimator for a_j is

$$\hat{a}_j = \frac{2}{w} \frac{1}{n} \sum_{i=1}^n \cos\left(\frac{j\pi x_i}{w}\right)$$

The estimator of Eq. (3.6) has been shown by Tolstov (1962:50–53) to be optimal in the sense of giving a minimum mean integrated squared error (MISE) estimator of $f(x)$. Because of the orthogonality of the FS, Eq. (3.6) holds no matter what value of m is chosen.

Equations for the sampling variances and covariances of the \hat{a}_j may be derived using a trigonometric identity. Based on Eq. (3.6) one can derive

$$\text{cov}(\hat{a}_j, \hat{a}_h) = \frac{1}{n} \left[\left(\frac{2}{w} \right)^2 \cdot \int_0^w \cos\left(\frac{j\pi x}{w}\right) \cdot \cos\left(\frac{h\pi x}{w}\right) \cdot f(x) dx - (a_j a_h) \right].$$

By using the identity (Vance 1954),

$$\cos(\theta_1) \cos(\theta_2) = \frac{1}{2} [\cos(\theta_1 + \theta_2) + \cos(\theta_1 - \theta_2)]$$

the cosine products in this above covariance formula can be reduced to simple cosine terms that are easily related to the \hat{a}_j . Thus,

$$\text{cov}(\hat{a}_j, \hat{a}_h) = \frac{1}{n} \left[\frac{1}{w} (a_{j+h} + a_{j-h}) - (a_j a_h) \right] \quad j \geq h. \quad (3.7)$$

Defining $a_0 = 2/w$, Eq. (3.7) is valid with $j = h$ and thus provides the sampling variance of \hat{a}_j . For simplicity of notation, let $\hat{\Sigma}/n$ represent the m by m variance-covariance matrix of $\hat{a}_1, \dots, \hat{a}_m$.

Unbiased estimators of the sampling variances and covariances of the Fourier coefficients are obtained from Eq. (3.7) by replacing the a 's by their estimators and replacing $1/n$ by $1/(n-1)$. Let $\hat{\Sigma}/(n-1)$ represent the unbiased estimator of the sampling covariance-variance matrix of $(\hat{a}_1, \dots, \hat{a}_k)$. That estimator of sampling variances and covariances is identical to the empirical estimator obtained by defining, for example, $y_{ki} = \frac{2}{w}$

$\cos(k\pi x_i/w)$ and computing

$$\hat{\text{var}}(\hat{a}_k) = \frac{\sum_{i=1}^n (y_{ki} - \bar{y})^2}{n(n-1)}.$$

One can extend this by defining

$$\hat{f}_i(0) = \frac{1}{w} + \sum_{k=1}^m \frac{2}{w} \cos\left(\frac{k\pi x_i}{w}\right),$$

and computing the sampling variance of $\hat{f}(0)$ (the average of the $\hat{f}_i(0)$) in the usual way.

A computable estimator of $\hat{f}(0)$ is defined as $\hat{f}(0) = \frac{1}{w} + \sum_{j=1}^m \hat{a}_j$, where m is a truncation point to be chosen. Although our estimate $\hat{f}(0)$ is biased, Rosenblatt (1956) has shown that essentially all density estimators are biased. Our numerical work indicates that the bias can be made very small with a truncation point of only modest size ($m \leq 6$). Kronmal and Tarter (1968) derived an objective rule for selecting m based on minimizing the MISE. The MISE is defined as

$$\text{MISE} = E \left[\int_0^w (\hat{f}(x) - f(x))^2 dx \right],$$

where

$$\hat{f}(x) = \frac{1}{w} + \sum_{j=1}^m \hat{a}_j \cos\left(\frac{j\pi x}{w}\right).$$

Due to the orthogonality of the series, MISE becomes

$$E \left[\sum_{j=1}^m (\hat{a}_j - a_j)^2 \frac{w}{2} + \sum_{j=m+1}^{\infty} (a_j)^2 \frac{w}{2} \right].$$

Assuming that MISE has a unique minimum as a function of m (see Kronmal and Tarter 1968),

$$\text{MISE}_{m+1} - \text{MISE}_m = \frac{w}{2} \left[\text{var}(\hat{a}_{m+1}) - (a_{m+1})^2 \right].$$

We find it appropriate to keep adding terms to $\hat{f}_m(x)$ until the first m such that $\text{MISE}_{m+1} - \text{MISE}_m \geq 0$. That is, select the smallest value of m such that $\text{var}(\hat{a}_{m+1}) \geq (a_{m+1})^2$. This is the stopping

rule suggested by Kronmal and Tarter (1968), it gives the following inequality

$$\frac{1}{w} \left(\frac{wa_{2m+2} + 2}{n+1} \right)^{1/2} \geq |a_{m+1}|.$$

If $f(x)$ has continuous derivatives, up to order 3, and bounded over the interval $(0, w)$, then,

$$a_m = \frac{2w}{m^2 \pi^2} [f'(w)(-1)^m - f'(0)] \\ + \frac{2w^2}{m^3 \pi^3} \int_0^w f'''(x) \sin\left(\frac{m\pi x}{w}\right) dx,$$

hence a_m is of order $1/m^2$. Therefore, for a given value of m , one can expect a_{2m+2} to be much smaller than a_{m+1} . Consequently, we suggest computing the stopping rule by using a_{2m+2} as zero rather than estimating it. That leads to stopping with the first value of m such that

$$\frac{1}{w} \left(\frac{2}{n+1} \right)^{1/2} \geq |\hat{a}_{m+1}|.$$

That stopping rule provides a reasonable procedure for selection of m (and is used in TRANSECT). More research on other stopping rules probably will yield an improvement.

Although the FS estimator is robust, it is not necessarily unbiased. In fact the selection rule is deliberately a trade-off of bias versus precision. One consequence of that possible bias is that the nominal 95 percent confidence intervals do not always have actual coverage of 95 percent. Let p_m = the probability of selecting m terms of the FS model; m has been constrained not to exceed 6 in program TRANSECT. Let C_m be the coverage of the nominal 95 percent confidence interval on $f(0)$ given the m -term FS estimator is used. Then the actual coverage using our selection rule is $\sum_{m=1}^6 p_m C_m$. The values of p_m are not readily computed, but the values of C_m are computable provided the true Fourier series coefficients are known.

For even moderate sized samples (say 50 or more), the estimators \hat{a}_i and $\hat{f}_m(0)$

(based on m -terms) have nearly normal distributions so the coverages C_m are affected only by the ratio of bias to standard error of $\hat{f}_m(0)$. Given this ratio, C_m is easily found (see e.g., Cochran 1963:12–15). The FS estimator has small relative bias when the true detection function is the half-normal. However, even in that case the nominal 95 percent confidence interval on the true $f(0)$ is likely to have true coverage of 90, 85, or even 80 percent (much of this is because $m = 1$ is often selected and C_1 is much less than 95%).

There is relatively little that can be done about the situation. It is a consequence of using a minimum mean integrated squared error criterion of estimator selection. Moreover, it is the coverage on \hat{D} that is really of interest, and there have been no investigations of that coverage yet. Dr. G. C. White (pers. comm.) looked at coverage of $f(0)$ using simulation and demonstrated it was often less than the nominal 95 percent. He suggested an improvement in coverage will occur by using our stopping rule to select m , but then take $m+1$ terms in the FS estimator.

Confidence interval coverage is a subject that needs more research, including comparison of a variety of estimators with regard to their properties, especially confidence interval coverage under a broad, but realistic, range of detection functions. The sampling variance of n enters into the variance of \hat{D} ; we have observed that often 50 percent or more of the variance of \hat{D} is attributable to n , thus extrapolating from coverage on $f(0)$ to coverage on D is not really possible. Until the coverage of the confidence interval on D is investigated, it is unknown to us whether there is any reason for concern. We emphasize that our goal has been to devise a robust estimator, hence, the necessary trade-off of bias versus precision may tend to preclude achieving exact 95 percent confidence interval coverage.

Grouped Data

The general theory of analysis of grouped line transect data was presented

above. To apply that theory to any model, such as the FS model for fixed m , requires an expression for the cdf $F(x)$. Integrating the FS model yields

$$F(x) = \frac{x}{w} + \sum_{j=1}^m a_j \left[\frac{w}{j\pi} \sin\left(\frac{j\pi x}{w}\right) \right].$$

Evaluating $F(x)$ at c_{h-1} and c_h leads to $p_h = F(c_h) - F(c_{h-1})$:

$$\begin{aligned} p_h &= \frac{c_h - c_{h-1}}{w} \\ &+ \sum_{j=1}^m a_j \left(\frac{w}{j\pi} \right) \left[\sin\left(\frac{j\pi c_h}{w}\right) \right. \\ &\quad \left. - \sin\left(\frac{j\pi c_{h-1}}{w}\right) \right]. \end{aligned}$$

The required partials of p_h with respect to the a_j are trivial.

A linear estimator of the a_j could have been devised that would not be affected by the possibility of negative cell probabilities (which could happen in theory because $f(x)$ can be negative). However, from experience with the FS model for grouped data, the p_h are never negative if reasonable cut points are used. Program TRANSECT numerically obtains the ML estimators of the a_j using the general theory for grouped data given above.

GENERALIZED LINEAR MODELS

An alternative linear model to the FS is a simple polynomial, $f(x) = a_0 + \Sigma a_j x^j$. We investigated and programmed this model for grouped data, but found it performed no better than the FS model, and is more difficult to use. Rather than give the basic theory of just the polynomial model, some general results applicable for any linear model are provided. The polynomial case follows from these results (as does much of the FS case).

Let

$$f(x) = \frac{1}{w} + \sum_{j=1}^{\infty} a_j \phi_j \left(\frac{x}{w} \right), \quad 0 < x < w$$

where the $\phi_j(\cdot)$ are known functions satisfying

$$\int_0^w \phi_j \left(\frac{x}{w} \right) dx = 0.$$

This guarantees that $\int_0^w f(x)dx = 1$, but allows $f(x) < 0$ to be possible.

An expression of $f(x)$ in an orthogonal series would have

$$\int_0^w \phi_j\left(\frac{x}{w}\right)\phi_h\left(\frac{x}{w}\right)dx = 0.$$

Such orthogonality is convenient but not necessary. The FS model has $\phi_j\left(\frac{x}{w}\right) = \cos\left(\frac{j\pi x}{w}\right)$ and is an orthogonal series. The simple polynomial model has

$$\phi_j\left(\frac{x}{w}\right) = \left(\left(\frac{x}{w}\right)^j - \frac{1}{j+1}\right),$$

and is not an orthogonal series. Representing $f(x)$ as an orthogonal polynomial series is possible but not simple.

For a given m , we approximate $f(x)$ by

$$f(x) = \frac{1}{w} + \sum_{j=1}^m a_j \phi_j\left(\frac{x}{w}\right).$$

The problems to be solved in this approach are to

- (1) estimate a_1, \dots, a_m ,
- (2) derive the properties of $\hat{a}_1, \dots, \hat{a}_m$, especially bias and sampling variances and covariances,
- (3) derive properties of $\hat{f}_m(x)$ (globally) and of $\hat{f}_m(0)$; for example bias, sampling variance and MISE, and
- (4) derive a stopping rule to choose m : we suggest the MISE criterion for ungrouped data and likelihood ratio tests for grouped data.

The analysis of grouped data is straightforward when using ML theory with such linear models. The real concern is with the analysis of ungrouped data and for that we suggest using the method of expectation to estimate the parameters. The following quantities are defined

$$c_{jh} = \int_0^w \phi_j\left(\frac{x}{w}\right)\phi_h\left(\frac{x}{w}\right)dx$$

$$j, h = 1, \dots, m$$

$$b_h = E\left(\phi_h\left(\frac{x}{w}\right)\right) = \sum_{j=1}^m a_j c_{jh}$$

$$h = 1, \dots, m.$$

Let the m by m matrix C be $[c_{jh}]$ and the vector $b = (b_1, \dots, b_m)'$. The vector of parameters is $a = (a_1, \dots, a_m)'$. The equation $b = Ca$ has the unique solution $a = C^{-1}b$; this assumes that C^{-1} exists, which it will for any reasonable choice of functions $\phi_j(\cdot)$.

Given the above developments, an unbiased estimator of a can readily be developed; n vectors of transformed data are defined as

$$b(i) = \begin{bmatrix} \phi_1\left(\frac{x_i}{w}\right) \\ \vdots \\ \phi_m\left(\frac{x_i}{w}\right) \end{bmatrix} \quad i = 1, \dots, n$$

(x_1, \dots, x_n are the perpendicular distance data). Under reasonable conditions $\hat{b} = \frac{1}{n} \sum_{i=1}^n b(i)$ is a multivariate normal random variable, $MVN(\hat{b}, \frac{1}{n} \hat{\Sigma})$.

An unbiased, strongly consistent estimator of $\hat{\Sigma}$ is

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (b(i) - \hat{b})(b(i) - \hat{b})'.$$

The variance covariance matrix of \hat{b} is $\frac{1}{n} \hat{\Sigma}$.

It follows that $\hat{a} = C^{-1}\hat{b}$ is unbiased for a and the variance-covariance matrix of \hat{a} is $(C^{-1} \hat{\Sigma} C^{-1})/n$. Next,

$$\hat{f}(0) = \frac{1}{w} + \sum_{j=1}^m \hat{a}_j \phi_j(0).$$

The expectation, hence bias of $\hat{f}(0)$ can be derived easily. The sampling variance of $\hat{f}(0)$ is

$$\text{var}(\hat{f}(0)) = \frac{1}{n} (\phi' C^{-1} \hat{\Sigma} C^{-1} \phi)$$

where $\phi' = (\phi_1(0), \dots, \phi_m(0))$.

The only difficult problem is to determine a stopping rule. We explored the equation for the MISE of $\hat{f}(x)$; unfortunately, it is very complicated unless the model is one of an orthogonal series

(such as the FS model). If a good stopping rule could be found for the polynomial model, that model could be useful in line transects, but we would still prefer the FS model on the basis that it satisfies the shape criterion and is an easy model to use (if the first term of the polynomial model is dropped it satisfies the *shape criterion*; i.e., set $a_1 \equiv 0$). Additional research and development on polynomial models seems worthwhile because such models can satisfy all the criteria that we recommend line transect models for $f(x)$ should meet.

There are methods related to the above ideas (essentially, linear models or methods based on series expansions) that are worth considering. For example the orthogonal series work of Brunk (1976, 1978) and methods based on splines (see e.g., Boneva et al. 1971). Dr. D. L. Burdick (pers. comm.) has developed a method based on splines; Gates (1979a) has also developed some ideas in that area.

PARAMETRIC MODELS

Numerical Methods for Parametric Models

The first purpose of this section is to present the general numerical methods used to get the exact ML estimators and their variance-covariance matrices for parametric models. The methods are almost entirely numerical; no approximations were used in the course of getting the ML estimators implemented in program TRANSECT, nor were less efficient moment type estimators used. The second purpose of this section is to present, and to comment on some of the parametric models examined (figures in APPENDIX C show plots of detection functions for many of those models).

Many models, including 2 generalized parametric models investigated in detail, can be conveniently expressed as

$$f(x) = \frac{1}{\mu(\theta)} e^{-h(x,\theta)} \quad 0 < x < w$$

$$\mu(\theta) = \int_0^w e^{-h(x,\theta)} dx,$$

where w can be infinite, and θ may be a vector valued parameter $\theta = (\theta_1, \dots, \theta_r)'$, $r \geq 1$. It is assumed that $h(x,\theta)$ has all first and second partial derivatives with respect to all components of θ , there is a well-defined parameter space, and that $h(x,\theta) \geq 0$ for all x and θ .

Numerical integration is sometimes required to obtain ML estimates in either the grouped or ungrouped case, and this integration is much simpler to do if w is finite (although it may be large with respect to the observed distance data). Consequently, we did not always try to implement the nontruncated cases ($w = \infty$).

Ungrouped Data

The required computations for ungrouped data are complex. Yet, it is desirable to describe clearly the process of obtaining the theoretical information matrix because those methods have been used to compute theoretical asymptotic efficiencies under the exponential power series and exponential quadratic models. Also, those equations are used by TRANSECT in estimating the information matrix.

At most, $(r+1)(r+2)/2$ quantities must be evaluated, where r is the number of parameters in the model. They are

$$\mu(\theta) = \int_0^w \exp(-h(x,\theta)) dx$$

$$Q_j = \int_0^w \left(\frac{\partial h}{\partial \theta_j} \right) \exp(-h(x,\theta)) dx \\ j = 1, \dots, r$$

$$R_{ij} = \int_0^w \left(\frac{\partial h}{\partial \theta_i} \right) \left(\frac{\partial h}{\partial \theta_j} \right) \exp(-h(x,\theta)) dx \\ i, j = 1, \dots, r.$$

The log-likelihood equations are

$$g_j(\theta) = - \sum_{i=1}^n \frac{\partial h(x_i, \theta)}{\partial \theta_j} + n \frac{Q_j}{\mu(\theta)} \\ j = 1, \dots, r.$$

The elements of the information matrix are

$$I_{ij} = \frac{R_{ij}}{\mu(\theta)} - \left(\frac{Q_i}{\mu(\theta)} \right) \left(\frac{Q_j}{\mu(\theta)} \right)$$

$$i, j = 1, \dots, r.$$

The estimator of $f(0)$ is $\hat{f}(0) = 1/\mu(\hat{\theta})$, with asymptotic sampling variance

$$\text{var}(\hat{f}(0)) = (f(0))^4 Q' (nI(\theta))^{-1} Q,$$

where Q is the vector (Q_1, \dots, Q_r) .

The required integrals usually are very complex or even impossible to express analytically. After expending much effort trying to implement those integrals analytically for some models, we decided it is better just to use numerical integration. Good numerical integration procedures exist, so that even completely numerical solutions for ML estimators (numerical integration at each step of an iterative process) take only a few seconds of CPU time on a moderate to large sized computer.

Grouped Data

Estimation of θ from grouped data by ML methods requires only the capability to compute $F(c_i)$ and $\partial F(c_i)/\partial \theta_j$. It was found that it is often much easier (if not essential) to compute these quantities by numerical integration rather than attempting analytical methods. The first step is to compute

$$P_i = \int_{c_{i-1}}^{c_i} \exp(-h(x, \theta)) dx \quad i = 1, \dots, k.$$

Then compute $\mu(\theta) = P_1 + P_2 + \dots + P_k$ and normalize the P_i to get $p_i(\theta) = P_i/\mu(\theta)$.

Generally, the function $h(x, \theta)$ will be simple enough so that its partials with respect to the θ_j can be determined. Let these be $(\partial h(x, \theta)/\partial \theta_j) \equiv \partial h/\partial \theta_j$. Numerically compute

$$\frac{\partial P_i}{\partial \theta_j} = - \int_{c_{i-1}}^{c_i} \left(\frac{\partial h}{\partial \theta_j} \right) \exp(-h(x, \theta)) dx,$$

then,

$$\frac{\partial \mu(\theta)}{\partial \theta_j} = \sum_{i=1}^k \frac{\partial P_i}{\partial \theta_j},$$

and finally,

$$\frac{\partial p_i}{\partial \theta_j} = \left[\frac{1}{\mu(\theta)} \left(\frac{\partial P_i}{\partial \theta_j} \right) - \left(\frac{\partial \mu(\theta)}{\partial \theta_j} \right) p_i(\theta) \right].$$

Given those computed cell probabilities and their partials, the information matrix of $\hat{\theta}$ can be computed. To solve for $\hat{\theta}$ actually requires those integrals to be computed at each step of the iterative process. Nonetheless, with an efficient overall computer algorithm $\hat{\theta}$ can be found with (at most) a few seconds of CPU time.

Simple Parametric Models

Any 1-parameter parametric pdf is considered to be a simple parametric model for $f(x)$. The most prominent of those are the negative exponential and the half-normal models. Developments of such models have generally been only for ungrouped, untruncated data.

Negative Exponential Model

Gates et al. (1968) thoroughly developed the negative exponential model for ungrouped, untruncated data only. The basic model is

$$f(x) = \lambda e^{-\lambda x} \quad 0 < x < \infty.$$

An efficient, closed form estimator of $f(0)$ exists here. However, for the truncated case,

$$f(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda w}} \quad 0 < x < w,$$

and the ML estimator is not closed form. Numerical procedures are easy to apply because the first and second partials of $f(x)$ are expressible in closed form. Relevant equations for ungrouped data are

$$\frac{\partial \ln(f(x))}{\partial \lambda} = \frac{1}{\lambda} - x - \frac{w}{e^{\lambda w} - 1}$$

$$I(\lambda) = \left(\frac{1}{\lambda} \right)^2 - \frac{w^2 e^{\lambda w}}{(e^{\lambda w} - 1)^2}.$$

Equations for grouped and truncated data are

$$F(x) = \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda w}} \quad 0 < x < w$$

$$\frac{\partial F(x)}{\partial \lambda} = F(x) \left[\frac{x}{e^{\lambda x} - 1} - \frac{w}{e^{\lambda w} - 1} \right].$$

For either grouped or ungrouped data,

$$\text{var}(\hat{f}(0)) = \left(\frac{f(0)}{\lambda} \right)^2 \left[1 - \frac{w\lambda}{e^{\lambda w} - 1} \right]^2 \text{var}(\hat{\lambda}).$$

The negative exponential model represents one of the few cases for which simple analytic equations exist for the first and second partial derivatives, thus eliminating the need for numerical integration.

The negative exponential model does not lead to robust estimators; they are neither *model robust* nor *pooling robust*. Also, it does not satisfy the *shape criterion* ($f'(0) = 0$), but is spike shaped at $x = 0$. That model has only one shape, as λ is a scale parameter. Its lack of robustness is due primarily to the inflexibility of shape. Fig. C2 (APPENDIX C) shows the shape of the negative exponential detection function.

Half-Normal Model

Quinn (unpublished thesis) presented results for the half-normal model applied to ungrouped, untruncated data. The basic model is

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

$$0 < x < \infty.$$

Closed form results exist for that untruncated, ungrouped case. For the truncated and grouped versions of that model, equations for $f(x)$, $F(x)$, and their partials with respect to σ are complicated because integrals of $f(x)$ over a finite interval cannot be expressed in closed form.

The analytic equations needed for the ungrouped case (i.e., first and second partials of $f(x)$) were determined, but because they are very complex, resorting to numerical integration rather than using them is recommended. For that half-normal model, $h(x, \theta) = \frac{1}{2} \left(\frac{x}{\sigma} \right)^2$ and the

necessary partial derivative is easily found. For truncated data,

$$\text{var}(\hat{f}(0)) = \left(\frac{\partial \hat{f}(0)}{\partial \sigma} \right)^2 \text{var}(\hat{\sigma}),$$

$$\frac{\partial \hat{f}(0)}{\partial \sigma} = f(0) \left[-\frac{1}{\sigma} + \frac{f(0)w}{\sigma} \cdot \exp \left(-\frac{1}{2} \left(\frac{w}{\sigma} \right)^2 \right) \right].$$

The reader probably would find it quite instructive to determine how complicated the analytic equations for estimation under the half-normal model are for any case except that of ungrouped, untruncated data. All other cases require substantial effort to program. Contrast this situation with the relative ease of implementing the FS model (which has a number of advantages compared to the half-normal model).

For the half-normal detection model (see Fig. C2, APPENDIX C), σ is a scale parameter and thus the model takes on only one shape. The estimator under this model is not *model robust*, although it does satisfy $f'(0) = 0$.

Generalized Parametric Models

As a general rule, the shape of any 1-parameter model of $f(x)$ is relatively inflexible. Thus, use of such models constitutes a restrictive assumption about the shape of the detection curve and therefore leads to nonrobust estimators. To obtain a more general family of curve forms for $f(x)$, generalized parametric models with 2 or more parameters are considered.

Two generalized parametric models were investigated intensively and several others were examined more casually. At early stages of our research, it was thought that the answer to robust line transect estimation would be to use some generalized parametric model, specifically the exponential power series (Pollock 1978). Our subsequent investigations, however, changed our minds because those models lead to robust but inefficient estimators of $f(0)$. Furthermore, the

data analysis with such models requires extensive numerical computer methods, not needed with the FS estimator.

Exponential Power Series

A model that incorporates the negative exponential and half-normal as special cases is Pollock's (1978) exponential power series. The detection function is

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

and the corresponding pdf in the untruncated case is

$$f(x) = \begin{cases} \exp\left(-\left(\frac{x}{\lambda}\right)^p\right) & 0 < x < \infty \\ \frac{\exp\left(-\left(\frac{x}{\lambda}\right)^p\right)}{\lambda\Gamma(1+1/p)} & 0 < \lambda \\ 0 & 0 < p. \end{cases}$$

Pollock (1978) deserves credit for suggesting this model for use in line transects. However, other researchers also considered this model and developed results for it (see e.g., Quinn unpublished thesis, Ramsey 1979, G. P. Patil et al. 1979).

That model can take on a wide variety of shapes (see Fig. C3, APPENDIX C); note that λ is a scale parameter and p is a shape parameter. For $p = 1$ the negative exponential results, whereas for $p = 2$ the half-normal model results. In the limit as $p \rightarrow \infty$, $f(x)$ becomes uniform on $(0, \lambda)$, but when $p \rightarrow 0$, $f(x)$ becomes extremely spike-shaped, that effect is so marked with a histogram of real data, the spike near $x = 0$ is not distinguishable, and over a finite interval $f(x)$ looks almost uniform. For $p > 1$, as λ becomes large, $f(x)$ again will look very flat over an interval near the line. Those redundancies in the shape of $f(x)$ apparently are partly responsible for the very high variances of $\hat{f}(0)$ based on this model. We (and others) have found it very hard to estimate p with any precision (see also Ramsey 1978, G. P. Patil et al. 1979). In fact, we constrained $p \geq 0.5$ in the numerical methods to avoid severe numerical problems with the ML estimator. Pollock intended

that the method be used forcing $p \geq 1$ (K. H. Pollock, pers. comm.). It is easy to determine that $g'(0)$ satisfies

$$g'(0) = \begin{cases} 0 & \text{if } p > 1, \\ -p/\lambda & \text{if } p = 1 \\ -\infty & \text{if } p < 1. \end{cases}$$

We believe that spiked models (such as when $p \leq 1$) are not realistic for line transect studies.

Because of its wide array of shapes, the model does lead to robust estimators in the sense of small bias. However, it generally has a large sampling variance for $\hat{f}(0)$. It also has a very high sampling correlation of $\hat{\lambda}$ and \hat{p} , often exceeding 0.9.

The equations needed for inference procedures with the exponential power series become so complex for anything except ungrouped, untruncated data that a totally numerical (as opposed to an analytical, or part analytical) approach is recommended. This model is a special case of the form $f(x) = \exp(-h(x, \theta))/\mu$, wherein $h(x, \theta) = (x/\lambda)^p$. The partials of $h(x, \theta)$ are readily obtained.

Exponential Quadratic

Another 2-parameter model that includes as special cases the negative exponential and half-normal is

$$f(x) = \frac{1}{\mu} \exp(-ax - bx^2) \quad \begin{matrix} 0 < x < w \\ 0 \leq a \\ 0 \leq b. \end{matrix}$$

Here, μ is just a normalizing constant

$$\mu = \int_0^w \exp(-ax - bx^2) dx,$$

and w may be finite or infinite; the function $h(x, \theta)$ is just $ax + bx^2$. That model is in the exponential family, and is a special case of more general models like $f(x) = \exp\left(-\sum_{i=1}^m a_i x^i\right)/\mu$ (see e.g., Crain 1974). For a general number of terms m , this is called the exponential polynomial model. In practice, we always used it with 2 terms. For $b = 0$, the negative ex-

ponential model results, and for $a = 0$ one has the half-normal model. As with the exponential power series model, the ML estimator for the model must be entirely numerical. Approximations to the ML estimator and other ad hoc estimators can be developed but are not recommended (e.g., Anderson et al. 1978 that appeared before we accepted the necessity for exact numerical ML estimation).

If $f(x)$ (and the underlying detection function) is to be monotonically decreasing on $[0, w]$, constraints must be placed on the parameters. If $w = \infty$, both a and b must be strictly positive. In general, there is an extremum point at $x = -(a/2b)$ and $f(x)$ is monotone on either side of this point. Evaluating $f''(-a/2b)$ gives $-2b$. It is concluded that one can not allow $a < 0$ and $b > 0$ because the model corresponds to a detection function giving values greater than 1 for some x values. However, one could use models with $a > 0$ and $b < 0$ in the range $0 < x < a/|2b|$, except that those forms are more spiked at the origin than the negative exponential model ($b = 0$), and thus their value is questioned. For practical purposes, this family of curves should have $0 \leq a$ and $0 \leq b$, in which case all curve forms fall between that of the negative exponential and the half-normal (see Fig. C4, APPENDIX C for some plots of this detection function).

Estimation with the model is robust, although estimates of $f(0)$ have a large sampling variance; a greater loss of precision occurs when generalizing to this 2-parameter model than had been expected. The shape criterion is not met except for $a = 0$ ($f'(0) = -a$); however, some members of this curve family are quite flat near $x = 0$, so the *shape criterion* is often nearly satisfied.

A Mixture Model

Another generalized family of pdfs including both the negative exponential and half-normal cases is the mixture of the two:

$$f(x) = (p)\exp(-\lambda x) + (1-p)\exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)/\mu,$$

for $0 \leq p \leq 1$ and μ being the required normalizing constant (expressible in closed form only if $0 < x < \infty$). All curve shapes in that family fall between that of the negative exponential and half-normal. That range of shapes can be seen in several of the figures in APPENDIX C, (see, e.g., Fig. C4).

We see no theoretical advantage to that model because it requires 3 parameters to cover a subset of the curve shapes in the 2-parameter models discussed above and it is slightly more difficult to deal with numerically. We did not investigate it as a basis for estimation. Quinn (unpublished thesis) used the model as a basis for simulation evaluation of line transect models; it is convenient in that regard, at least for the untruncated case.

Reversed Logistic Model

Eberhardt (1968, 1978a) recommended the reversed logistic (2-parameter) model, and used it as a basis for simulation of line transect sampling. For estimation purposes that model has been implemented with grouped data, by G. P. Patil et al. (1979) and in the present study. The detection function for the model is

$$g(x) = \frac{(1+b)e^{-ax}}{1+be^{-ax}} \quad 0 < a \\ 0 < b,$$

and the pdf and cdf, respectively, for the untruncated case are

$$f(x) = \frac{ae^{-ax}}{(1+be^{-ax})\ln(1+b)/b} \\ F(x) = 1 - \frac{\ln(1+be^{-ax})}{\ln(1+b)}.$$

The range of curve shapes in that family is more restricted than in the 3 models given above (see APPENDIX C). At one extreme, $g(x)$ approaches the negative exponential model as $b \rightarrow 0$. Because it is a limiting case, the reversed logistic mod-

el is always less convex and less spiked at the origin than the negative exponential with detection function $\exp(-ax)$. Note that a is a scale parameter, whereas b is a shape parameter. We have $g'(0) = -a/(1+b)$, so $g'(0) = 0$ is only approached as a limiting case. However, for small values of that ratio, the condition $g'(0) = 0$ is almost met. Moreover, the model will have a distinct shoulder near $x = 0$ whenever that ratio is small. Note, also in this regard that $g(x)$ has an inflection point at $x = \ln(b)/a$ for $b < 1$. Thus, if b is at all large there is a distinct shoulder near $x = 0$. Fig. C5 (APPENDIX C) shows some plots of the reversed logistic detection function.

The ML estimator does not exist in closed form for any case (truncated or untruncated, grouped or ungrouped), and thus, requires the use of numerical methods. However, the required partial derivatives can be found in closed form. The required partials are easier to find for the analysis of grouped data when $0 < x < \infty$, they are

$$\frac{\partial F(x)}{\partial a} = \frac{bx}{\ln(1+b)(e^{ax}+b)},$$

$$\frac{\partial F(x)}{\partial b} = \frac{1}{\ln(1+b)} \left[\frac{1-F(x)}{1+b} - \frac{1}{e^{ax}+b} \right].$$

For truncated, grouped data one has $F_w(x) = F(x)/F(w)$ from which relation the required partials can be found for this case. An equation for the asymptotic variance of $f(0)$ can be derived, as it requires only the additional knowledge of $\partial f(0)/\partial a$ and $\partial f(0)/\partial b$ (that are tedious to obtain but can be found).

The reversed logistic model may lead to robust estimators, but investigations by G. P. Patil et al. (1979) and in the present study show that $\hat{f}(0)$ has a very large sampling variance under the model. Our limited investigation of that model indicated that it would not lead to a more efficient estimator of $f(0)$ than either the exponential quadratic or power series models; therefore, we did not investigate it fully.

Incomplete Gamma Model

If $H(x)$ represents any cdf, then $g(x) = 1 - H(x)$ defines a detection function model. This simple observation can be used as the basis for generating an unlimited number of parametric models. Sen et al. (1974) used that approach to base a model on the incomplete gamma cdf, however they did not investigate in that paper the model given below. The resultant $f(x)$ is more complex than any of the above models. For the untruncated case,

$$f(x) = \frac{\beta}{\alpha} \left[1 - \int_0^x \frac{\beta^\alpha y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)} dy \right]$$

$$0 < \alpha$$

$$0 < \beta.$$

The corresponding detection function is $g(x) = f(x)\alpha/\beta$.

That family of curves takes on a wide variety of shapes (see Fig. C6, APPENDIX C) that are appropriate for line transect models. For $\alpha > 1$, $g'(0) = 0$ (i.e., the *shape criterion*), whereas for $\alpha \leq 1$ the detection curve is spiked at the origin. The case $\alpha = 1$ gives the negative exponential model, but the half-normal model is not a special case of the family. The curve shape is controlled by α , and β is a scale parameter. That incomplete gamma model has essentially the same range of useful shapes as the exponential quadratic, exponential power series, and reversed logistic models. Thus, there is no apparent advantage in implementing that model and we do not know of anyone who has done so. There are disadvantages with the model: inference based on it is extremely complex, more so than any of the generalized parametric models discussed above. We do not recommend expending the effort to investigate the model.

A Comment

Using $g(x) = 1 - H(x)$, it is apparent there is no end to the parametric models one could conceive; however, 2 parame-

ters suffice to cover the logical range of basic detection function shapes. Thus, for modeling the detection function as a family of curve shapes, any of the 5 parametric models discussed above is basically suitable (as are other models).

In our initial investigations, we were thus led to ask if some such families might be "inadmissible" in the following sense. Let there be r different parametric families of generalized models. Let the true pdf $f(x)$ be arbitrary except that it has a form like those covered by those families. Let $\hat{f}_i(0)$ be the ML estimator of $f(0)$ based on the i^{th} parametric family. If there exists a pair i,j ($i \neq j$) such that $\text{var}(\hat{f}_i(0)) < \text{var}(\hat{f}_j(0))$ for all (reasonable) shapes of $f(x)$, then, the j^{th} family is not admissible as a basis for estimation of $f(0)$. We investigated such a condition only with respect to the exponential quadratic and exponential power series pair of families. Neither is inadmissible with respect to the other. In general, we doubt any reasonable family of curves would be inadmissible.

The significance of this, if true, is that investigation of alternative generalized parametric models will lead not to the elimination of any, but only to a plethora of competing estimators. We have found that such generalized parametric models produce less efficient estimators than the approach using robust linear models (such as the FS model) for the curve shapes we believe reasonable in line transect sampling. It is suggested, therefore, that investigation of such generalized parametric models will be fruitless for producing an estimate competitive with the FS approach. We do not conclude that further research is not needed; rather, that it should concentrate on linear models, improved stopping rules, and improved estimators. Also there are other areas of nonparametric methods such as kernel estimators or spline methods that may prove fruitful. Finally, we do not believe there can be *dramatic* improvements in estimator properties beyond those of the Fourier series.

METHODS FOR SIGHTING DISTANCE AND SIGHTING ANGLE DATA

This section presents the essence of the specific theory underlying the Hayne, modified Hayne, and generalized Hayne estimators. We assume the reader has already read *Estimation Based on Sighting Distances and Angles* in PART 1, which gives the more abstract theory leading to those Hayne type estimators.

For all 3 estimators, one assumes that the conditional pdf of x (perpendicular distance) given sighting distance r , is expressible as

$$f(x|r) = \frac{1}{r} c(x/r).$$

Hence, it is assumed that r is a scale parameter in $f(x|r)$. It is convenient to define $y = (x/r) = \sin(\theta)$. Letting $h(r)$ represent the pdf of r , the joint distribution of x and r as $f(x,r) = h(r) \frac{1}{r} c(x/r)$. Making the transformation to y and r the joint pdf of y and r is seen to factor,

$$f(y,r) = \frac{1}{r} h(r) c(y) \quad 0 < y < 1, \quad 0 < r.$$

Therefore, y and r (or θ and r) are independent random variables under that scale parameter assumption on r in $f(x|r)$.

From *Estimation Based on Sighting Distances and Angles*, PART 1,

$$f(0) = E\left(\frac{1}{r}\right) c(0),$$

$$\text{hence, } \hat{D} = \frac{n\hat{c}(0)}{2L} \bar{r}_h,$$

$$\bar{r}_h = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i}.$$

The parameter $c(0)$ and its estimator depend on the assumed model for $c(y)$. The variance of $\hat{f}(0) = \hat{c}(0)\bar{r}_h$ can be obtained in general terms. Using independence, an exact representation is (using $\hat{c} = \hat{c}(0)$)

$$\text{var}(\hat{f}(0)) = (f(0))^2 [(cv(\hat{c}))^2 + (cv(\bar{r}_h))^2]$$

$$+ (cv(\hat{c})cv(\bar{r}_h))^2 \Big].$$

That, however, is not the formula on which we have based an estimator. Because it is negligibly small, we dropped the term $(cv(\hat{c})cv(\bar{r}_h))^2$ for simplicity and used

$$\text{var}(\hat{f}(0)) = (f(0))^2 \Big[(cv(\hat{c}))^2 + (cv(\bar{r}_h))^2 \Big]. \quad (3.8)$$

The variance of \hat{c} varies with the model. The variance of \bar{r}_h is estimated empirically.

The approximation of Eq. (3.8) is quite good for any "reasonable" data. By this is meant any case in which both coefficients of variation are less than 0.5; they generally will be unless the sample size is very small. In fact, if both coefficients of variation are less than 0.25 the cross-product term contributes relatively little. For example, if $cv(\hat{c}) = cv(\bar{r}_h) = 0.25$, the true $cv(\hat{f}(0)) = 0.3590$, and the approximation is 0.3536. By the time one computes $\text{var}(\hat{D})$, the effect of this approximation becomes totally negligible.

The Hayne Estimator

For the Hayne estimator (Hayne 1949), one assumes that y is a uniform random variable on $[0,1]$. Thus, $c(y) \equiv 1$, and $c(0) = 1$ is known, not estimated. Applying Eq. (3.8),

$$\text{var}(\hat{f}(0)) = (f(0))^2 ((cv(\bar{r}_h))^2).$$

The only other theory underlying the Hayne estimator presented in the section *The Hayne Estimator* (PART 2) discussed here is the test of the assumption that $E(\theta) = 32.7^\circ$. In the past, the test was based on estimating the sampling variance of θ from the data (i.e., from $\theta_1, \dots, \theta_n$). However, the hypothesis really being tested is that y is uniform $[0,1]$. That completely specifies the distribution of $\theta = \arcsin(y)$; thus $\text{var}(\theta)$ can be found theoretically. Making the indicated transformation, we have the pdf of θ as $f(\theta) = \cos(\theta)$, $0 < \theta < \pi/2$ (θ in radians).

It is a simple matter to calculate

$$\text{var}(\theta) = \pi - 3 \quad (\theta \text{ in radians squared})$$

or

$$\text{var}(\theta) = 464.843 \quad (\theta \text{ in degrees squared}).$$

Therefore, assuming one will be working with θ degrees, $\text{var}(\theta) = 464.843/n$, and for large samples

$$z = \frac{\sqrt{n}(\bar{\theta} - 32.7)}{21.56}$$

is approximately a standard normal variable under the null hypothesis.

Modified Hayne Estimator

Burnham and Anderson (1976) introduced this estimator. They assumed that an alternative to Hayne's model might be θ uniformly distributed on the interval $[0, \pi/2]$. That implies an average sighting angle of 45° . Angles often lie in the range 32.7° to 45° , suggesting the approximation to $c(y)$ as a linear combination of those 2 cases. When θ is a uniform random variable, the distribution of $c(y)$ is readily found to be

$$c(y) = \frac{2}{\pi} \frac{1}{(1 - y^2)^{1/2}}.$$

A generalized model for $c(y)$ is now

$$c(y) = p + (1 - p) \frac{2}{\pi} \frac{1}{(1 - y^2)^{1/2}} \quad 0 < y < 1, \quad (3.9)$$

and $c(0) = p + 2(1 - p)/\pi$. We could estimate p from the data y_1, \dots, y_n , as by ML estimation (but it would not be closed form). Alternatively, by using a method of expectation we could find an estimator of p as $\hat{p} = (\bar{y} - 2/\pi)(0.5 - 2/\pi)$. This leads to an estimate of c . However Burnham and Anderson (1976) introduced an estimator by interpolating between $c(0) = 1$ and $c(0) = 2/\pi$, based on $\bar{\theta}$. There is no reason to introduce a changed version, so we use that modified Hayne estimator with this mixture model for $c(y)$. We note, however, that the model of Eq. (3.9) could be rigorously used.

Based on interpolation with the observed sighting angle, Burnham and Anderson (1976) suggested $\hat{c} = (1 - \delta) + \delta(2/\pi)$ where $\delta = (\bar{\theta} - 32.7)/(45 - 32.7)$.

Substituting the equation for δ in \hat{c} leads to $\hat{c} = 1.96606 - 0.0295432(\bar{\theta})$, which is what we used in the section *The Modified Hayne Estimator* in PART 2 ($\bar{\theta}$ is in degrees here, not radians).

It is necessary to know $\text{var}(\hat{c})$ to use Eq. (3.8) as a basis for $\hat{\text{var}}(f(0)) = (0.0295432)^2 \text{var}(\theta)/n$, and $\text{var}(\theta)$ can be estimated from the $\theta_1, \dots, \theta_n$. This explains the source of our estimator for $\text{var}(f(0))$ in the section *The Modified Hayne Estimator* (PART 2).

Generalized Hayne Estimator

A logical way to generalize the Hayne model is to assume that animals are sighted whenever the observer crosses an imaginary flushing curve about the animal. Such a curve is meant to be representative of the animals' perception of the approaching observer, and as such, it may be reasonable to assume an elliptic flushing curve with one axis parallel to the line of the observer's travel (Fig. 33). That model has been developed by Burnham (1979); those results are summarized here with respect to $c(y)$.

Imagine an x, z coordinate system superimposed on Fig. 33, with the bird at the zero point. The general formula for the ellipse is

$$\left(\frac{x}{a}\right)^2 + \left(\frac{z}{b}\right)^2 = 1.$$

Using the relationships $r^2 = x^2 + z^2$, $y = x/r$, and solving the elliptic equation for x as a function of y gives

$$x = \frac{yb}{\left[1 + \left(\left(\frac{b}{a}\right)^2 - 1\right)y^2\right]^{1/2}}.$$

Assuming the ratio of ellipse axes is fixed at $c = b/a$,

$$x' = \frac{yc}{\left(1 + (c^2 - 1)y^2\right)^{1/2}},$$

where $x' = x/a$ is uniformly distributed on the interval $(0,1)$ (because of random line placement). The transformation $y \rightarrow x'$ is 1 to 1; it follows that the cumulative distribution function of the random variable y is simply

$$C(y) = \frac{yc}{\left(1 + (c^2 - 1)y^2\right)^{1/2}},$$

for $y \in (0,1)$ and $0 < c$. Differentiating $C(y)$ gives the pdf of y

$$c(y) = \frac{c}{\left(1 + (c^2 - 1)y^2\right)^{3/2}}.$$

It is readily seen that $c(0) = c = b/a$ hence $c(0)$ and c can be used interchangeably.

From Burnham (1979),

$$E(y) = \frac{1}{1 + c}.$$

Thus, an estimator of c is

$$\tilde{c} = \frac{1}{\bar{y}} - 1,$$

however, the estimator \tilde{c} is not the ML estimator. Rather, the ML estimator \hat{c} satisfies the likelihood equation

$$g(c) = \frac{n}{c} - \sum_{i=1}^n \frac{3c(y_i)^2}{1 + (c^2 - 1)(y_i)^2} = 0;$$

the asymptotic sampling variance of \hat{c} is

$$\text{var}(\hat{c}) = \frac{5}{4} \frac{c^2}{n}.$$

That ML estimator must be found by numerical methods. A method that works well (and is used by TRANSECT) is the Newton-Raphson iterative method using \tilde{c} as a starting value. The formula for that method as applied here is

$$\hat{c}_{i+1} = \hat{c}_i + \frac{1.25}{n} (\hat{c}_i)^2 g(\hat{c}_i)$$

with $\hat{c}_0 \equiv \tilde{c}$ and stopping when, say, $|\hat{c}_{i+1} - \hat{c}_i| < (0.001/n)$.

THE RELATIVE EFFICIENCY OF VARIOUS ESTIMATORS OF $f(0)$

Introduction

Early in our investigations we studied the theoretical relative efficiency of estimation of $f(0)$ based on the FS versus the ML estimator for the exponential power series and exponential quadratic models (for ungrouped perpendicular distance

data). At that time we did not know whether any of those 3 models would serve as a general basis for robust estimation. The results of those investigations of efficiency suggested that the FS estimator would be a good omnibus procedure (criteria such as *pooling robustness* were conceived later and confirmed the choice of the FS method). A brief presentation of the investigation and its basic results are given here.

The following material deals with a complex subject; the intended audience is statisticians (or readers well versed in statistical theory). It is assumed that the reader has a basic understanding of the concepts of relative efficiency, asymptotic relative efficiency, and the Fisher information matrix. Those concepts are not excessively difficult, but the methods that had to be used to compute relative efficiencies are quite complex.

Let $f(x, \theta)$ be any parametric model, with $\theta = (\theta_1, \dots, \theta_r)'$ representing $r \geq 1$ parameters. If $\hat{\theta}$ is the ML estimator of θ , then the ML estimator of $f(0)$ is $f(0, \hat{\theta})$ with sampling variance $\text{var}(f(0, \hat{\theta}))$. Asymptotically, the sampling variance of $f(0, \hat{\theta})$ can be determined on the basis of the Fisher information matrix. General equations for such parametric models were given under *Numerical Methods for Parametric Models* and are based on representing

$$f(x, \theta) = \exp(-h(x, \theta)) / \mu(\theta) \quad 0 < x < w,$$

where

$$\mu(\theta) = \int_0^w \exp(-h(x, \theta)) dx,$$

and $f(0, \theta) = 1/\mu(\theta)$. Those equations were used in numerically obtaining the ML estimator in program TRANSECT. Here, they were used to compute numerically the theoretical information matrix of $\hat{\theta}$,

$$Q_j = \int_0^w \left(\frac{\partial h}{\partial \theta_j} \right) \exp(-h(x, \theta)) dx \\ j = 1, \dots, r,$$

$$R_{ij} = \int_0^w \left(\frac{\partial h}{\partial \theta_i} \right) \left(\frac{\partial h}{\partial \theta_j} \right) \exp(-h(x, \theta)) dx$$

$$i, j = 1, \dots, r.$$

The i, j^{th} element of the Fisher information matrix, $I(\theta)$, is

$$I_{ij} = \frac{R_{ij}}{\mu(\theta)} - \left(\frac{Q_i}{\mu(\theta)} \right) \left(\frac{Q_j}{\mu(\theta)} \right).$$

Defining the 1 by r vector $Q = (Q_1, \dots, Q_r)'$, then the asymptotic (large sample) variance of $f(0, \hat{\theta})$ is

$$\text{var}(f(0, \hat{\theta})) = (f(0))^4 Q' (I(\theta))^{-1} Q / n.$$

We only are concerned with 2 particular parametric models, both with 2 parameters ($r = 2$). In both models, setting one parameter to a fixed value, and estimating the remaining one is of interest because it corresponds to some simple 1-parameter models (often the negative exponential or half-normal). Thinking of those 2-parameter models as generalizations of simple models, one is led to ask how much efficiency in estimating $f(0)$ is lost when both parameters are estimated rather than assuming a particular shape and estimating only 1 parameter. It is known that generalizing to 2 parameters will greatly increase robustness and hence reduce bias, but will a substantial loss in efficiency occur in comparison with the simple model?

If $\theta = (\theta_1, \theta_2)'$ and one considers θ_2 as known, the equations given above are all applicable with $r = 1$. In particular, $I(\theta) = I_{11}(\theta)$, and

$$\text{var}(f(0, \hat{\theta}_1 | \theta_2)) = (f(0))^4 (Q_1)^2 / (I_{11}(\theta) n).$$

The relative efficiency (RE) of $\hat{f}(0)$ on the basis of the full 2-parameter model versus that obtained on the basis of estimating only θ_1 given θ_2 is

$$\text{RE} = \frac{\text{var}(f(0, \hat{\theta}_1 | \theta_2))}{\text{var}(f(0, \hat{\theta}_1, \hat{\theta}_2))}.$$

The asymptotic relative efficiency of the ML estimator $f(0, \hat{\theta})$ versus $f(0 | \hat{\theta}_1, \theta_2)$ is, therefore,

$$\text{RE} = \frac{(Q_1)^2}{I_{11}(\theta) (Q' (I(\theta))^{-1} Q)}. \quad (3.10)$$

We computed that sort of efficiency because if it were high for all reasonable

shapes (i.e., values of θ_2) there would be no reason to base estimation on the simple model (i.e., arbitrarily fix θ_2) rather than to use the general model (and gain much greater robustness with only a moderate loss in efficiency).

The Fourier Series

Given any underlying detection function $\exp(-h(x,\theta))$, the theoretical FS coefficients are

$$a_j = \frac{2}{\mu(\theta)w} \int_0^w \cos\left(\frac{j\pi x}{w}\right) \exp(-h(x,\theta)) dx.$$

That calculation and that for $\mu(\theta)$, Q_j , and R_{ij} can be done efficiently if combined into 1 loop of a computer program calling a numerical integration routine. Given the FS coefficients a_1, \dots, a_{2m} , $E(\hat{f}_m(0))$ can be computed, and the sampling variance of $\hat{f}_m(0)$,

$$E(\hat{f}_m(0)) = (1/w) + a_1 + \dots + a_m$$

$$\text{var}(\hat{f}_m(0)) = \frac{1}{n} \left[\sum_{h=1}^m \sum_{j=1}^m \text{cov}(\hat{a}_j, \hat{a}_h) \right],$$

where $\text{cov}(\hat{a}_j, \hat{a}_h)$ is given by Eq. (3.7). From the above, the mean square error (MSE) of $\hat{f}(0)$ can be computed,

$$\begin{aligned} \text{MSE}(\hat{f}_m(0)) &= \text{var}(\hat{f}_m(0)) \\ &\quad + (E(\hat{f}_m(0)) - f(0))^2. \end{aligned} \quad (3.11)$$

Notice that the bias does not depend upon sample size so that the MSE is not proportional to $1/n$.

In a parametric context, one computes asymptotic relative efficiency as a ratio of variances (because asymptotically the estimators are unbiased). Asymptotic relative efficiency is not meaningful here because it will be zero (see Crain et al. 1978 for a discussion of this point). Instead, the relative efficiency of the FS versus the ML estimator of $f(0)$ for the underlying parametric model is defined as

$$\text{RE}(\text{FS}_m) = \frac{\text{var}(f(0,\hat{\theta}))}{\text{MSE}(\hat{f}_m(0))}, \quad (3.12)$$

for any given sample size n .

It is important to understand Eq. (3.12). The value of $\text{RE}(\text{FS}_m)$ will exceed 1 only if the mean square error of the FS estimator based on m terms is smaller than the sampling variance of the estimator represented by $\hat{f}(0) = f(0,\hat{\theta})$ (some estimator based on a parametric model). Then, based on the criterion of selecting an estimator with the smallest mean square error, the FS estimator would be preferred over that based on the given parametric model.

We examined that relative efficiency and other properties of the FS model for the exponential power series and exponential quadratic detection functions for $m = 1, \dots, 6$, $n = 40, 60, 80, 100, 150, 200$, and 400 , and a range of parameter values. The m that gave the minimum MSE of the FS estimator and the m that gave minimum MISE of $\hat{f}_m(x)$ were determined. The percentage relative bias (PRB) as well as the sampling variances of the ML estimator and the relative efficiency in Eq. (3.10) were examined.

That study produced hundreds of pages of output of which only a few key items are presented. In particular, we looked at the theoretical optimum efficiency of the FS estimator by using Eq. (3.11) to find (numerically) the m that minimized $\text{MSE}(\hat{f}_m(0))$. Using that (optimal) m , the FS relative efficiency Eq. (3.12), which is called the theoretical optimal FS relative efficiency was tabulated. Although that efficiency cannot be obtained in practice, it shows how good the FS estimator can be in theory.

Also, the m value from the stopping rule was computed to give the minimum MISE FS model. That m and the one giving the minimum MSE of $\hat{f}_m(0)$ were generally the same or only one unit different except for very spiked detection functions (such as the negative exponential).

Summary results on relative efficiencies and percentage relative bias are presented below for each of the 2 generalized models examined.

TABLE 17.—PERCENTAGE RELATIVE EFFICIENCY (RE, SEE EQ. (3.12) AND TEXT FOR EXPLANATION) OF THE THEORETICAL OPTIMAL FS ESTIMATOR VS. THE ML ESTIMATOR OF $f(0)$ WHEN THE TRUE DETECTION FUNCTION IS THE EXPONENTIAL POWER SERIES, $g(x) = \exp(-(x/\lambda)^p)$, $0 < x < w$. ALSO SHOWN IS THE PERCENTAGE RELATIVE BIAS (PRB) OF THE FS ESTIMATOR. RESULTS ARE GIVEN FOR SAMPLE SIZES 40 AND 100. WITHOUT LOSS OF GENERALITY, THOSE COMPARISONS ARE OVER THE SPACE (p, c) , $c = w/\lambda$

Scale-truncation value, $c = w/\lambda$	Sample size	Shape parameter, p									
		0.75		1.0		1.5		2.0		2.5	
		RE	PRB	RE	PRB	RE	PRB	RE	PRB	RE	PRB
0.5	40	1468	-11.6	941	-5.4	401	0.0	222	1.5	150	1.6
	100	1468	-11.6	941	-5.4	401	0.0	222	1.5	150	1.6
1.0	40	940	-19.2	831	-11.6	544	-2.2	298	2.8	181	5.5
	100	463	-19.2	511	-11.6	527	-2.2	286	2.8	158	5.5
2.0	40	396	-30.8	418	-24.0	335	-14.2	477	-7.7	717	-3.3
	100	248	-23.3	235	-16.1	256	-6.4	174	0.5	135	3.5
3.0	40	333	-30.1	257	-23.2	273	-12.9	441	-5.8	548	-1.1
	100	206	-24.6	181	-16.7	210	-7.0	236	-0.8	152	3.6
4.0	40	304	-29.0	233	-22.0	262	-11.9	403	-4.9	463	-0.1
	100	170	-24.8	147	-17.5	191	-7.6	242	-1.1	170	3.4
5.0	40	226	-33.3	214	-21.6	261	-11.4	392	-4.4	393	0.5
	100	150	-25.1	129	-17.9	186	-7.9	252	-1.4	183	3.2

Models Investigated

With 2-parameter models, one is required to examine results over a 3-dimensional space (θ_1, θ_2, w) , because a value of w must also be specified. It was found in both models that the 3-dimensional space could be reduced to a 2-dimensional space without loss of generality as regards quantities like percentage relative bias of the FS and other relative efficiencies of interest. The mathematical verification of the validity of those "map-

pings" from 3- to 2-dimensional space are not given because they are very tedious (as a matter of interest the computer results corroborated the results).

Exponential Power Series Model

For the exponential power series model, $h(x, \theta) = (x/\lambda)^p$, $0 < \lambda$, and $0 < p$ see Fig. C3, APPENDIX C, for detection function shapes. For investigating the properties of the FS, the 3-dimensional space (λ, p, w) maps into $(1, p, c)$ where $c = (w/\lambda)$. Thus, the scale parameter λ can be set to 1 and one deals only with a shape parameter p and a truncation-scale parameter $c = w/\lambda$. The various properties of interest to us (such as the PRB of the FS model) will be the same for any sets (λ_1, p_1, w_1) and (λ_2, p_2, w_2) such that $p_1 = p_2$ and $(w_1/\lambda_1) = (w_2/\lambda_2)$.

To get a feel for how it is proven consider the following:

$$\begin{aligned} \mu(\lambda, p, w) &= \int_0^w \exp(-(x/\lambda)^p) dx \\ &= \mu(1, p, c), \end{aligned}$$

which follows by transforming the inte-

TABLE 18.—PERCENTAGE RELATIVE EFFICIENCY OF THE ML ESTIMATOR OF $f(0)$ UNDER THE EXPONENTIAL POWER SERIES MODEL WHEN BOTH λ AND p ARE ESTIMATED VERSUS WHEN p IS KNOWN AND ONLY λ IS ESTIMATED

Scale-truncation value $c = w/\lambda$	Shape parameter, p				
	0.75	1.0	1.5	2.0	2.5
0.5	16.0	20.3	26.4	30.5	33.5
1.0	16.3	20.5	26.1	29.7	32.1
2.0	16.7	20.6	25.3	27.9	29.8
3.0	17.0	20.7	25.4	28.6	30.3
4.0	17.2	20.9	26.0	28.7	30.3
5.0	17.4	21.1	26.3	28.7	30.3
Averages	16.8	20.7	25.9	29.0	31.1

TABLE 19.—MARGINAL CHOICES FOR THE PARAMETERS $a^* = aw$ AND $b^* = bw$ OF THE EXPONENTIAL QUADRATIC MODEL. WITHOUT LOSS OF GENERALITY, WE SET $w = 1$. THE 100 PERCENT TRUNCATION LEVEL REPRESENTS THE LIMITING CASE OF $f(x)$ UNIFORM ON $(0,1)$

Truncation level (%)	a^*	b^*
100	0.0	0.0
80	0.22314	0.03190
70	0.35667	0.07423
60	0.51082	0.13750
50	0.69315	0.22748
40	0.91629	0.35415
30	1.20397	0.53706
20	1.60944	0.82120
10	2.30259	1.35280
5	2.99573	1.92074
1	4.60517	3.34332

grand to $z = x/\lambda$. Now the FS coefficients are

$$a_m(\lambda, p, w) = \frac{2}{w\mu} \int_0^w \cos\left(\frac{m\pi x}{w}\right) \exp(-(x/\lambda)^p) dx = a_m(1, p, c).$$

It follows that $f_m(0, \lambda, p, w) = f_m(0, 1, p, c)$. Similar results can be derived regarding the equivalence of efficiencies.

Now one must examine only the var-

ious theoretical properties of interest over the 2-dimensional space (p, c) , which is equivalent to just setting $\lambda = 1$. This was done for a 10 by 10 matrix of those parameter values. First, 10 shape parameters were chosen to cover the range of shapes possible: $p = 0.75(0.25)3.0$. Next, it was determined that $c = 0.5(0.5)5.0$ covered the range of truncation-scale values of interest. The “cross product” of those choices gave 100 parameter pairs. For all of them, we numerically computed the true $f(0)$, the expected value of the FS estimator, $f_m(0)$, its variance, $\text{var}(\hat{f}_m(0))$ and $\text{MSE}(\hat{f}_m(0))$, all for $m = 1, \dots, 6$, the optimal FS estimator, and several relative efficiencies as explained above for a range of sample sizes (as well as all sorts of intermediate results like information matrixes). Key results for a very restricted, but quite representative range of p , c , and n are presented in Tables 17 and 18.

Exponential Quadratic Model

For the exponential quadratic model, $h(x, \theta) = ax + bx^2$ ($a > 0$, $b > 0$), and the 3-dimensional space (a, b, w) maps, without loss of generality for our purposes, into the 2-dimensional space $(a^*, b^*, 1)$

TABLE 20.—PERCENTAGE RELATIVE EFFICIENCY (RE, SEE EQ. (3.12) AND TEXT FOR EXPLANATION) OF THE THEORETICAL OPTIMAL FS ESTIMATOR VERSUS THE ML ESTIMATOR OF $f(0)$ WHEN THE TRUE DETECTION FUNCTION IS THE EXPONENTIAL QUADRATIC, $g(x) = \exp(-ax - bx^2)$, $0 < x < w$. ALSO SHOWN IS THE PERCENTAGE RELATIVE BIAS (PRB) OF THE FS ESTIMATOR. RESULTS ARE GIVEN FOR SAMPLE SIZES 40 AND 100. WITHOUT LOSS OF GENERALITY, THOSE RESULTS ARE OVER THE SPACE $a^* = aw$ AND $b^* = bw^2$

a^* values	Sample size	b* Values									
		0.0		0.07423		0.35415		1.35280		3.34332	
		RE	PRB	RE	PRB	RE	PRB	RE	PRB	RE	PRB
0.0	40	400	0	413	0.5	464	1.9	716	2.3	1318	-4.8
	100	400	0	413	0.5	464	1.9	716	2.3	1318	-4.8
0.35667	40	481	-3.8	499	-3.5	570	-2.9	846	-3.7	658	-10.8
	100	481	-3.8	499	-3.5	570	-2.9	846	-3.7	658	-10.8
0.91629	40	461	-10.6	470	-10.5	496	-10.6	477	-12.5	422	-8.6
	100	461	-10.6	470	-10.5	496	-10.6	477	-12.5	422	-8.6
2.30259	40	189	-18.3	189	-18.3	188	-18.4	184	-18.8	230	-13.8
	100	97	-18.3	97	-18.3	130	-13.0	133	-13.3	128	-13.8
4.60517	40	116	-20.0	116	-20.0	117	-20.0	120	-20.2	126	-20.6
	100	71	-16.5	71	-16.5	72	-16.6	72	-16.7	89	-14.3

TABLE 21.—PERCENTAGE RELATIVE EFFICIENCY OF THE ML ESTIMATOR OF $f(0)$ UNDER THE EXPONENTIAL QUADRATIC MODEL WHEN BOTH a AND b ARE ESTIMATED VS. WHEN a IS CONSIDERED KNOWN (TOP NUMBER FOR EACH VALUE OF a) AND ONLY b IS ESTIMATED AND THE SAME COMPARISON OF PERCENTAGE RELATIVE EFFICIENCY WHEN b IS CONSIDERED KNOWN (BOTTOM NUMBER FOR EACH VALUE OF a) AND ONLY a IS ESTIMATED

a^* values	b^* values					b^* averages
	0.0	0.07423	0.35445	1.35280	3.34332	
0.0	15.6	15.6	15.3	14.3	12.9	14.7
	37.5	37.5	37.7	38.3	40.3	38.3
0.35667	15.6	15.5	15.2	14.1	12.8	14.6
	38.0	38.0	38.2	38.8	40.8	38.8
0.91629	15.3	15.2	14.9	13.8	12.6	14.4
	38.7	38.7	38.9	39.6	44.6	39.5
2.30259	14.1	14.0	13.7	12.8	12.1	13.3
	40.0	40.1	40.3	41.4	43.4	41.0
4.60517	11.8	11.7	11.6	11.4	11.5	11.6
	42.7	42.8	43.2	44.4	45.7	43.8
						Overall averages
						13.7 40.3

where $a^* = aw$ and $b^* = bw^2$. To choose a range of values of a^* and b^* to cover the space of interest, we resorted to a different tactic here (neither is a shape or scale parameter).

For $b^* = 0$, one has the negative exponential model and for $a^* = 0$, the half-normal model. We therefore fixed $w = 1$, then fixed $b^* = 0$ and chose a^* values to achieve 11 different truncation levels for a negative exponential model. Similarly, we chose 11 values of b^* with $a^* = 0$ and $w = 1$. The resultant 121 combinations of a^* and b^* were then examined in the same way as for the exponential power series model (i.e., a range of sample sizes and numerous relative efficiencies examined). Table 19 gives the full set of values of a^* and b^* chosen. Because the relative efficiencies and PRBs vary smoothly over that space, results for only some of these combinations are presented in Tables 20 and 21.

Results and Discussion

From Table 17, one can see the general behavior of the PRB for the FS estimator under the exponential power series mod-

el and the optimal relative efficiency of the FS compared to the ML estimator of $f(0)$. Bear in mind that it is a theoretical optimal efficiency and that it is relative to estimating both λ and p , not just λ with p fixed (some of the latter RE values are given in Crain et al. 1978).

The FS percentage RE and the PRB depend on sample size as expected. In Table 17, for samples of size 40 and 100, the theoretical RE ranges from 200 to 1,500 percent. However, the extreme cases (like 1,000%) correspond to very spiked models which are regarded as both unreasonable and inappropriate. The range of desired and reasonable detection function shapes is approximately in the space spanned by $1.5 < p$ and $0.5 < (w/\lambda) < 4.0$. In that range of parameter values, the PRB is not worse than -15 percent for $n = 40$ nor -8 percent for $n = 100$ (PRB always decreases as n increases). The theoretical optimum RE of the FS estimator for that range of parameter values is 200 to 500 percent. This means that the ML estimators under the exponential power series would have standard errors that range from about 1.4 to 2.2 times the standard errors of the theoretically optimum FS estimator. Al-

though it is realized the actual efficiency might not be that large, it was sufficient to concentrate our attention on the FS method as being quite suitable in practice.

A different type of (asymptotic) RE for the exponential power series model is given in Table 18. It answers the question of how much efficiency is lost if it is thought the true detection curve has a shape that corresponds approximately to a given shape parameter p , but one estimates both p and λ to achieve robustness. With proper interpretation of parameters, Table 18 gives the RE of Eq. (3.10). It is easy to see that a great loss in efficiency occurs if both p and λ (rather than just λ), are estimated. The actual RE of $\hat{f}(0)$ depends almost entirely on the true shape parameter (note that the cases with $c = 5.0$ correspond to a nearly untruncated model). Beyond $p = 2.5$, relatively little change is evident in the tabulated relative efficiencies. Thus, for the exponential power series model with $p > 1$ in which case $g'(0) = 0$ holds, there is no more than about 33 percent RE if that general 2-parameter model is used, instead of fixing (i.e., knowing) p . Yet the FS model is quite efficient where the true $g(x)$ satisfies $g'(0) = 0$. Again, the result motivated our choice of the FS method as an omnibus method.

For the exponential quadratic model, Table 19 gives the selected values of a^* and b^* , while Tables 20 and 21 present the various corresponding relative efficiencies. In that family, the most spiked curves are for $a^* > 0$ and $b^* = 0$ ($a^* = 0$ and $b^* = 0$ for $0 < x < w$ is really the uniform $(0,w)$ case). The percentage RE of the FS estimator is again generally 100 percent or more, but overall it seems lower than for the comparison to the exponential power series. Although the results still oriented us toward the FS, they tended to suggest that in a choice between these 2 parametric models, the exponential quadratic model might have higher efficiency.

A distinct difference is evident in the asymptotic RE's of $\hat{f}(0)$ compared to the

case of known a versus known b (Table 21). In fact, as a rough guideline, it is seen that if one assumed b was known and estimated only a to get $\hat{f}(0) = f(0, \hat{a}|b)$, the RE of $\hat{f}(0) = \hat{f}(0, \hat{a}, b)$ versus $f(0, \hat{a}|b)$ would only be about 14 percent. The converse RE of $\hat{f}(0)$ vs. $f(0, b|a)$ is about 40 percent. In either case, a substantial loss in efficiency is engendered by going to the 2-parameter model to gain robustness and the losses of efficiency are greater than we would have expected. Again, the results of this investigation favored the FS as a better method than this 2-parameter model.

Two recent papers have presented some results related to the above. Ramsey (1979) investigated the exponential power series model and found, as we did, a considerable loss in efficiency when both parameters are estimated rather than just one. G. P. Patil et al. (1979) investigated the negative exponential model for some real data, and then also fit the exponential power series to those data. They observed, as we did, a tremendous loss in estimation efficiency of $f(0)$ with the 2-parameter model. It is also interesting that they ended up suggesting a cosine model (i.e., the FS model) as a promising robust method. They and we were working on this problem independently.

We repeat that the results given here are but a small (but very representative) part of a large numerical investigation of the FS model vis-à-vis 2 general parametric models. We concluded that the FS is likely to be as efficient as ML estimation based on those models (if not more so) and equally robust with real data (and realistic sample sizes, say $n > 40$), where the true detection function is not known and would be only approximated by any model (parametric or otherwise). Partly because of the theoretical nature of the efficiency comparisons, we also conducted a simulation study (COMPUTER SIMULATION RESULTS) of the FS and other estimators. Such simulations, however, are necessarily of limited scope and we put at least as much reliance on the theoret-

ical results, which suggest the good behavior of the FS estimator.

COMPUTER SIMULATION RESULTS

Simulation Study Design

Extensive computer simulation studies were performed to investigate empirically the properties of many of the perpendicular distance estimators presented in this monograph. Simulation provides information about *model robustness* and *estimator efficiency*, which are 2 of our criteria for robust estimation of $f(0)$. *Model robustness* can be determined by examining the bias in the estimate of $f(0)$ for a variety of underlying detection functions. Statistical efficiency can be examined empirically by calculating the standard error and MSE of $\hat{f}(0)$ from replicate samples of a particular detection function. This section describes the manner in which the simulation was performed and gives a qualitative description of the results.

Simulation of a line transect for estimation from perpendicular distances was performed by specifying a form for the underlying detection function $g(x)$ and then deriving the corresponding pdf $f(x)$. A sample of observed perpendicular distances was then generated directly from the pdf $f(x)$ by a variety of random number generation techniques depending on the particular form of the pdf (Maisel and Gnugnoli 1972). That is equivalent to the more intuitive approach of generating a perpendicular distance x_i from the uniform distribution on the interval $[0, w]$ and a random probability P_i from the interval $[0, 1]$; where, if P_i is less than $g(x_i)$, the object is detected and the procedure is continued until n observations have been made.

The observer in the field provides the detection function, which is the pooled result of many different functions. The detection function can vary widely, which implies the need for a model robust estimator, as explained in PART 1 (*The Central Problem*). For the simula-

tion, 9 different detection functions were chosen, using the criterion that they be representative of many different shapes that might occur in the field. The negative exponential was suggested by Gates et al. (1968) as a feasible model so it was used as one of the functions. The negative exponential pdf is

$$f(x) = a \exp(-ax) \quad 0 < x < w.$$

The parameter a was set to 0.065 for the simulation. Quinn (unpublished thesis) and Hemingway (1971) suggested the half-normal as an alternate model. It was used here in the form

$$f(x) = \frac{\exp(-x^2/2\sigma^2)}{\sigma\sqrt{2\pi}} \quad 0 < x < w,$$

with $\sigma = 33.33$. Those 2 functions were each used at 3 truncation values ($w = 100, 65$, and 35), which provided 6 different models. The other 3 models were provided by the modified beta function, which is very similar to the power series model suggested by Eberhardt (1968). The modified beta pdf is

$$f(x) = \frac{a+1}{w} (1-x/w)^a \quad 0 < x < w = 100.$$

We used 3 different values of a (0.5, 1.0 and 2.5) to provide convex, linear, and concave ("spiked") forms to represent the underlying detection functions. It is irrelevant that these functions may or may not exactly represent some true situation (spiked detection functions are probably not realistic models). They were chosen only to enable examination of a broad range of possible detection function shapes. The graphs of those underlying detection functions are given in APPENDIX C (Figs. C1 and C2).

From each of the 9 models, 25 samples of perpendicular distances were generated for 4 sample sizes ($n = 40, 60, 100, 400$). An estimate of $f(0)$ was calculated for each of the 25 samples and an average estimate $\hat{f}(0)$ was then calculated. From those, the average PRB, the standard error of $\hat{f}(0)$, and the MSE were calculat-

ed. The average PRB can be represented by

$$\text{PRB} = \left(\frac{\hat{f}(0) - f(0)}{f(0)} \right) \times 100,$$

where $f(0)$ is the true value for the particular detection function. From the simulation, the sampling variance of $\hat{f}(0)$ is estimated by

$$\hat{\text{var}}(\hat{f}(0)) = \frac{\sum_{i=1}^R (\hat{f}_i(0) - \hat{f}(0))^2}{R - 1},$$

the standard error is

$$\hat{s}\text{e}(\hat{f}(0)) = (\hat{\text{var}}(\hat{f}(0))/R)^{1/2},$$

and the MSE is

$$\text{MSE} = \hat{\text{var}}(\hat{f}(0)) + (f(0) - \hat{f}(0))^2,$$

where R is the number of replications (e.g., 25 in this case). Those statistics were used to compare the various estimators in terms of their robustness and efficiency.

Not all of the estimators presented in this monograph were used in the computer simulation study. Some of the methods have not been developed fully and others are not generally useful or are without sound theoretical development. However, each of the general classes of estimators is represented: simple parametric (negative exponential, NEXP; half-normal, HN; and modified beta, MB), generalized parametric (exponential polynomial, EPOLY; and exponential power series, EPS), nonparametric linear models (Fourier series, FS) and nonparametric methods based on grouped data (COX 1 and COX 2) and grouped isotonic regression (ISREG). The FS estimator has been discussed extensively and the exponential polynomial, exponential power series, half-normal, and negative exponential are presented under *Parametric Models* (in PART 3). The remaining estimators, Cox, modified beta, and grouped isotonic regression are presented in PART 4 (COMMENTS ON OTHER ESTIMATORS).

The data were analyzed ungrouped,

except for the Cox and grouped isotonic regression estimators. Although the other estimators used in the simulation study can be used for grouped data, they were designed for ungrouped data, whereas Cox and grouped isotonic regression are applicable only to grouped data. Estimation of $f(0)$ for the ungrouped data estimators is straightforward. However, because implementation of the grouped data estimators is more subjective, we present a description of the approach we used. The sample of observed perpendicular distances generated from the detection function was grouped arbitrarily into equal intervals of size 5, from 0 to the particular width (w) of the detection function. Within that framework, the grouped isotonic regression estimator was used in the manner detailed in PART 4 (METHODS BASED ON PERPENDICULAR DISTANCES) for equal intervals. The Cox estimator was also used within that framework, but its use requires further decision about the cut points c_1 and c_2 before estimation can proceed. Two rules were developed to decide on the c_i , which correspond to suggestions given by Eberhardt (1978a). In the simulation, they were considered as 2 different estimators, COX 1 and COX 2. For COX 1, the 2 intervals defined by the cut points c_1 and c_2 were chosen with equal width such that approximately 90 percent of the data were included in the 2 intervals. For COX 2, c_1 was set to 5 (the first interval in the data) and the second cut point c_2 was specified by the criterion that approximately 90 percent of the data were included in the 2 intervals. The problem of simulating those methods emphasizes the subjective nature of the estimators.

Simulation Results

The results of the simulation for 2 of the 4 sample sizes used are presented in Table 22. The average PRB, the standard error and the MSE of $\hat{f}(0)$ are presented for comparison (the latter 2 values were multiplied by 10,000 to enable easier comparison between estimators).

TABLE 22.—SUMMARY OF COMPUTER SIMULATION RESULTS FOR $n = 60$ AND 100 . THE PRB, STANDARD ERROR OF $f(0)$, AND MSE ARE PRESENTED FOR COMPARISON. THE MNEMONICS FOR THE ESTIMATORS ARE EXPLAINED IN THE TEXT

Detection function	Sample size	Estimators								
		Grouped data					Ungrouped data			
		COX 1	COX 2	ISREG	EPOLY	EPS	FS	NEXP	HN	MB
Negative exponential $w = 35$	60	-19.6 ¹	-5.6	-12.8	-9.6	12.3	-13.7	-3.4	-29.8	-28.3
		17.8 ²	33.4	25.4	27.7	62.0	28.6	23.6	11.1	15.8
		2.8 ³	3.0	2.5	2.4	10.4	3.0	1.5	5.0	4.8
	100	-15.7	-4.5	-13.2	-2.2	17.6	-14.8	-0.3	-28.8	-27.5
		18.8	25.5	21.0	30.5	60.6	26.0	24.8	11.2	17.3
		2.2	1.7	2.0	2.3	10.8	2.8	1.5	4.7	4.7
Negative exponential $w = 65$	60	-17.8	-3.4	-10.1	-4.4	16.4	-11.2	3.0	-37.4	-24.3
		15.1	23.4	18.4	18.6	43.7	28.2	16.4	7.7	13.5
		1.9	1.4	1.3	1.0	5.9	2.5	0.7	6.2	3.0
	100	-22.4	-4.3	-12.0	-15.7	5.7	-12.5	-1.8	-39.1	-26.6
		11.2	22.4	19.4	14.4	34.6	25.0	14.7	8.0	15.0
		2.5	1.3	1.6	1.6	3.1	2.3	0.6	6.8	3.6
Negative exponential $w = 100$	60	-21.4	-9.7	-16.4	-8.7	-5.6	-20.4	0.9	-40.9	-16.8
		18.1	19.7	16.7	20.1	31.5	17.4	18.2	12.2	20.3
		2.8	1.4	1.8	1.3	2.6	2.5	0.8	7.5	2.2
	100	-22.2	-6.2	-13.5	-5.8	-6.5	-16.5	1.9	-40.2	-15.4
		11.7	18.5	15.4	18.9	22.7	21.8	15.1	10.3	16.9
		2.4	1.0	1.4	1.0	1.5	2.3	0.6	7.1	1.7
Half-normal $w = 35$	60	9.3	9.9	18.7	13.2	12.3	4.7	15.4	3.2	1.3
		17.1	31.4	19.7	12.1	16.3	15.0	16.8	9.8	10.1
		0.8	2.6	1.4	0.6	0.8	0.6	1.0	0.3	0.3
	100	4.4	4.0	10.6	6.7	6.8	0.1	10.4	0.8	-1.1
		13.2	16.5	9.5	8.7	13.3	9.3	10.7	6.7	6.2
		0.5	0.7	0.4	0.2	0.5	0.2	0.4	0.1	0.1
Half-normal $w = 65$	60	11.6	-4.6	8.4	6.4	0.8	-0.6	24.3	-0.4	1.0
		8.6	18.8	10.9	10.8	10.6	7.8	9.1	4.8	6.3
		0.3	0.9	0.3	0.4	0.3	0.2	0.6	0.1	0.1
	100	12.0	-4.3	6.4	7.7	1.8	1.1	26.5	1.0	2.5
		9.6	12.5	7.2	8.4	9.6	8.0	8.7	4.5	5.3
		0.3	0.4	0.2	0.2	0.2	0.2	0.6	0.1	0.1
Half-normal $w = 100$	60	16.3	11.4	21.8	17.9	3.0	2.4	61.4	3.7	21.2
		7.5	20.8	13.5	7.2	7.5	8.8	7.1	4.7	8.1
		0.3	1.2	0.7	0.3	0.2	0.2	2.3	0.1	0.4
	100	13.1	7.8	10.9	9.3	2.2	2.5	59.7	2.2	17.7
		5.9	14.7	10.6	6.8	5.6	5.6	4.7	2.5	5.0
		0.2	0.6	0.4	0.2	0.1	0.1	2.1	0.1	0.2
Modified beta $a = 0.5$	60	7.4	11.4	28.2	7.5	-15.8	5.4	67.4	10.7	-0.5
		5.4	13.7	8.6	3.9	4.1	4.1	3.7	2.4	4.1
		0.1	0.5	0.4	0.1	0.1	0.1	1.1	0.1	0.1
	100	-1.3	-13.0	6.3	1.9	-15.0	-5.3	65.9	11.5	1.0
		4.7	9.9	6.2	3.5	2.9	4.2	3.9	2.3	3.9
		0.1	0.3	0.1	0.1	0.1	0.2	1.0	0.1	0.1
Modified beta $a = 1.0$	60	0.4	8.3	14.0	2.7	-16.7	-8.2	48.3	-3.1	-0.2
		5.9	14.4	11.6	6.0	5.5	5.2	4.2	2.4	4.1
		0.1	0.6	0.4	0.1	0.2	0.1	1.0	0.1	0.1
	100	-2.6	0.5	5.0	-2.0	-17.9	-7.1	47.7	-3.4	-1.1
		6.0	13.3	9.9	5.5	4.4	5.7	4.4	2.4	4.1
		0.1	0.4	0.3	0.1	0.2	0.1	1.0	0.1	0.1

TABLE 22.—CONTINUED

Detection function	Sample size	Estimators								
		Grouped data				Ungrouped data				
		COX 1	COX 2	ISREG	EPOLY	EPS	FS	NEXP	HN	
Modified beta $a = 2.5$	60	-14.1	2.1	1.7	-8.8	-18.0	-14.9	27.2	-19.2	
		10.8	21.3	17.1	10.3	10.5	12.2	9.5	5.5	
		0.5	1.1	0.7	0.4	0.7	0.6	1.1	0.5	
	100	-2.2	0.3	1.7	0.1	-6.9	-5.3	32.8	-17.4	
		8.9	17.3	11.9	11.1	10.2	10.9	7.9	4.2	
		0.2	0.8	0.4	0.3	0.3	0.3	1.5	0.4	
¹ PRB—average percent relative bias for 25 replications.										
² Standard error of $f(0) \times 10^4$.										
³ MSE—Mean square error $\times 10^4$.										

Several conclusions can be drawn from examining the results in regard to *model robustness*. It is evident, as would be expected, that the simple parametric estimators (NEXP, HN, MB) provide only negligible bias when the underlying detection function is the same as the estimator used. However, when the underlying detection function is only slightly different, the estimate is subject to extreme bias. For example, the NEXP has a minimal relative bias for the negative exponential detection function, but it has a relative bias on the order of 30 percent for the modified beta with $a = 2.5$ which has a shape similar to the negative exponential. That implies the need for a more flexible estimator such as the exponential polynomial, exponential power

series, Fourier series or the Cox and isotonic regression estimators. All of those estimators are fairly *model robust*. Some are more robust than others, and there is no single estimator that performs better than all others for all of the detection functions. For example, the FS estimator performs well with concave to slightly convex functions, but its bias increases dramatically with severely spiked functions (like the negative exponential). However, that is not a great concern because in real data it is not likely that the detection function will be so spiked. The exponential polynomial estimator, on the other hand, is robust to convex functions but does not perform as well as the FS estimator on concave functions (i.e., half-normal). Finally, it can be noted from the

TABLE 23.—RATIOS (IN %) OF ESTIMATED SAMPLING VARIANCES AND RATIOS OF MSE'S FROM THE EXPONENTIAL POLYNOMIAL AND EXPONENTIAL POWER SERIES ESTIMATORS (NUMERATOR OF THE RATIOS), TO THE FOURIER SERIES ESTIMATOR (DENOMINATOR OF RATIOS) FOR SAMPLE SIZE 100. THE MSE RATIOS ARE ESTIMATED RELATIVE EFFICIENCIES OF THE RESPECTIVE PARAMETRIC ESTIMATORS TO THE FS ESTIMATOR AS DEFINED THEORETICALLY BY EQ. (3.12). THE RATES ARE COMPUTED FROM DATA IN TABLE 22

Estimator	Detection function								
	Negative exponential			Half-normal			Modified beta		
	w = 35	w = 65	w = 100	w = 35	w = 65	w = 100	a = 0.5	a = 1	
Exponential polynomial	137 ¹	34	76	88	110	146	69	92	102
	82 ²	70	43	100	100	200	50	100	100
Exponential power series	543	190	108	204	144	100	100	59	88
	386	135	65	250	100	100	100	200	100

¹ Ratio of variances.² Ratio of MSE (i.e., estimated relative efficiency).

results that the Cox estimator is affected by cut point choices that are not selected objectively. The need for an objective estimator that is *model robust* leads to consideration of only the generalized parametric or nonparametric linear models for ungrouped data.

A comparison of the statistical efficiency of the estimators enables further discrimination in choosing among them. By limiting the range of estimators to the exponential polynomial, exponential power series, and Fourier series, one can compare the standard error and MSE. The percentage ratios of the standard error and MSE of the exponential polynomial and exponential power series to the Fourier series are given in Table 23. If the percentage is greater than 100, the Fourier series is more efficient (hence, to be preferred) and vice versa if the opposite is true. The ratios are computed directly from Table 22 so some rounding is involved. From Table 23, it appears that the Fourier series is more efficient than the exponential power series. However, that is not the case with the exponential polynomial estimator, which disagrees with the results of the previous section on theoretical efficiencies. That effect could be a result of the small sample size

(25 replications). However, there is one important difference in the empirical standard errors compared with the theoretical standard errors for the exponential polynomial. In the simulation, the parameters were both constrained to be greater than zero, which is reflected in a smaller variation in the estimate of $f(0)$. The theoretical variances for the exponential polynomial do not take that constraint into account. However, that does not increase the usefulness of the method because most line transect studies will be such that the variance of density will be constructed from the theoretical variance of $f(0)$ and not from replications.

The results of the simulation study are useful in showing that some estimators are definitely not *model robust* (e.g., negative exponential and half-normal). They also demonstrate the subjective nature of estimators like Cox and that the choices to be made are crucial in terms of bias. However, the results do not show any single estimator as better than all others in terms of bias and efficiency; this suggests that the *pooling robustness* criterion and the *shape criterion* must be used to limit the possible range of estimators to be considered.

PART 4

COMMENTS ON OTHER ESTIMATORS

In this section, most of the other estimation schemes we have examined are reviewed briefly. We have drawn on material by Seber (1973), Eberhardt (1978a), Gates (1979a), Quinn (unpublished thesis), and Hayes (unpublished thesis). The interested reader is encouraged to review those papers for more detail on previous methods.

We have tried to give a general assessment concerning the present usefulness of each method (often a fairly subjective judgment). The assessments were based on what we believe are 4 desirable properties of an estimator that is to be broadly applicable for general line transect samples. In order of importance, they are:

- (1) The estimator should be robust and we believe this implies a nonparametric approach (this property is called *model robustness*),
- (2) The estimator should satisfy the condition

$$n\hat{f}(0) = \sum_{i=1}^k n_i \hat{f}_i(0),$$

enabling pooling data from different detection functions without bias (this property is called *pooling robustness*),

- (3) The derivative of $f(x)$ should be zero at the origin; i.e., $f'(0) = 0$ (this property is called the *shape criterion*), and
- (4) The estimator should be statistically efficient; i.e., have a small sampling variance (this property is called *estimator efficiency*).

We believe those criteria are important and provide a qualitative basis for judging the overall usefulness of various approaches. In addition, the fit of the model to the observed data must be considered (see PART 1). This is particularly important with the simple parametric methods (e.g., negative exponential and half-normal).

A convenient summary of some char-

acteristics of various methods is given in Table 24. No malice is intended when it is stated that several of the previous methods are no longer useful. Instead, it is merely a reflection of the many important advances that have been made in the field. Indeed, we look forward to the day when better methods will supersede the ones now recommended in PARTS 2 and 3. The methods presented in detail in PARTS 2 and 3 are not discussed below. The entries in Table 24 generally reflect the method as it was originally published. In many cases, we have extended methods previously published, but those extensions are not reflected in the table. The final column is our subjective appraisal of the overall usefulness of each specific method. That assessment is based on 5 conditions: (1) sound theoretical development, (2) *model robustness*, (3) *pooling robustness*, (4) the *shape criterion*, $f'(0) = 0$, and (5) high *estimator efficiency*. We judge *estimator efficiency* only for models that have the *model robust* property (e.g., the estimator of Gates et al. 1968 is very efficient for the negative exponential model for ungrouped, untruncated data, but it is not *model robust*). Gates (1979b) has produced a computer program to compute point estimates of several line transect estimators in Table 24 as well as several new methods he has developed. Estimates of sampling variances and computation of many test statistics are not yet derived but may eventually be incorporated into his program LINETRAN.

Several recent papers attempted to compare parametric and nonparametric approaches, and several implied that nonparametric methods are robust only at the expense of estimation efficiency (e.g., Gates 1979a, Quinn unpublished thesis). As shown in PART 3, however, several nonparametric estimators are more efficient than the appropriate para-

TABLE 24.—SUMMARY OF VARIOUS ESTIMATORS OF DENSITY ON ASSOCIATED PROPERTIES AND ATTRIBUTES (na = NOT APPLICABLE). SEE TEXT FOR EXPLANATIONS

	Estimation method	Generally useful method
<i>Based on Perpendicular Distances</i>		
Emlen	X	X
Frye	X	X
Kelker	X	X
Leopold et al.	X	X
Gates et al. negative exponential	X	X
Hemmingsway, Quinn half-normal	X	X
Eberhardt power series	X	X
Modified beta	X	X
Sen et al.	X	X
Eberhardt reversed logistic	X	X
Anderson et al. log quadratic	X	X
Anderson and Pospahala	X	X
Cox	X	X
Truncated histogram estimators	X	X
Isotonic regression	X	X
Crain et al. Fourier series	X	X
Polynomial	X	X
Exp (quadratic)	X	X
Pollock exponential power series	X	X
<i>Based on Sighting Distances and Angles</i>		
King	X	X
Gates	X	X
Webb	X	X
Yapp	X	X
Hayne	X	X
Modified Hayne	X	X
Generalized Hayne	X	X

¹ Of course, ungrouped data can always be grouped. An X in those columns indicates the data used in the estimator. If an estimator is applicable for grouped data it can always be used for data taken ungrouped.

² If an estimator exists for $w^* < \infty$ it is essentially applicable for $w = \infty$. Alternatively, if an estimator is developed on the assumption that $w = \infty$, it will perform poorly on truncated data.

³ In original publication.

⁴ Efficient for arbitrary data relative to other estimators, this is only applicable to *model robust* estimators.

⁵ Within a given species where the class intervals c_i are the same.

metric estimators for the sample sizes likely to be encountered in practice.

The methods discussed in PART 4 tend to fall into several broad classes (also see Hayes unpublished thesis:A2). First, is a series of simple parametric approaches toward modeling $f(x)$ (e.g., negative exponential, half-normal, Eberhardt's power series, and the modified beta distribution). Those methods have generally been published for only ungrouped and untruncated data, although they can be extended to handle properly grouped and truncated data. Generally, those approaches lack both *model robustness* and *pooling robustness* and can often be expected to fit the data poorly.

Second, a number of generalized parametric forms (2 or more parameters) serve as models for $f(x)$ (e.g., incomplete gamma, reversed logistic, mixtures of simple pdfs, exp(quadratic), and the exponential power series). The development of that class of approaches is less complete and their use is often considerably more difficult than the simple parametric methods. Usually, only the ungrouped and untruncated case has been explored for any given model. Such approaches often seem attractive because they typically are *model robust*, can meet the *shape criterion* ($f'(0) = 0$, or at least include that condition), and usually fit the data. Our experience, however, has been that such generalizations have low estimation efficiencies and tend to lack the property of *pooling robustness*. That general approach could be explored further by using other distributions and it could be developed by means of standard theory for any hypothesized pdf $f(x;\theta)$.

Third, a wide variety of nonparametric models exist (see Wegman 1972a, 1972b) where, conceptually, $f(x)$ is modeled as a function of infinitely many parameters (e.g., an orthogonal series). In general, one can express $f(x)$ as a series

$$f(x) = \frac{1}{w^*} + \sum_{i=1}^{\infty} a_i \phi_i(x)$$

where the functions $\phi_i(x)$ are known.

Here, the expansion of $f(x)$ in terms of the $\phi_i(x)$ might be a Fourier series, Taylor series, exp(Fourier series), exp(Taylor series), or some other series. The estimation problem then also includes a rule to select the number of terms in the model for a given data set. That class of estimation schemes can be developed for either grouped or ungrouped data, as well as for truncated or untruncated data. Such a general approach can be used to develop methods that are *model robust* and *pooling robust*, that allow the *shape criterion* $f'(0) = 0$, and are highly efficient for realistic sample sizes (i.e., $n = 30-200$). Several of the nonparametric estimators (Kelker, Cox, jackknife, truncated histogram estimators, and polynomial expressions) are closely related special cases because they are all based on a Taylor series expansion or a polynomial approximation of $f(x)$. Finally, it is curious to note that most simple nonparametric approaches presented in the literature rely on grouped data, in sharp contrast to the development of parametric models that have traditionally used ungrouped data. That dichotomy certainly is not necessary because nonparametric estimators exist for the ungrouped case and parametric estimators exist for the grouped case.

Fourth, several less orthodox approaches have been suggested (e.g., splines, isotonic regression, use of order statistics, and information theoretic and kernel estimators) but not fully developed or explored. Those are difficult approaches because they are often not based on classical statistical theory, but the research potential is high.

The dated concept of "effective strip width" is not covered in the material below, although many of the methods developed previously were originally viewed in that light. Although the concept is not incorrect, it certainly has clouded the issues for several decades in terms of both field procedures and theoretical development.

In the material below, we do not make any distinction between "fixed flushing

radius" models and "variable flushing radius" models. As Hayes (unpublished thesis) pointed out, they are merely special cases of the general probabilistic model. The general analysis is the same for those dichotomies (see original discussion of Eberhardt 1968, then 1978a). We also point out that many of the methods in the literature have been misused. For example, we have seen many instances where the method of Gates et al. (1968) has been used with grouped or truncated data, for which the equations of Gates et al. are not applicable. In addition, Hayne's (1949) method has been used with perpendicular distance data, giving ridiculous results.

Most published methods have dealt with ungrouped, untruncated data. In many cases, we have extended the theory to cover grouped or truncated cases (or both) and these options are available in program TRANSECT. Extension for the truncated case is very important in practice, because of the desirability of deleting extreme observations that commonly occur when data are taken with $w = \infty$. It is often advantageous to delete 1–3 percent of the extreme ungrouped distance data (outliers) and employ a truncated analysis scheme.

In the following material, we often make use of the histogram frequencies n_1, n_2, \dots, n_k from the perpendicular distances x_1, x_2, \dots, x_n within the (perhaps unequal) class intervals $0 < c_1 < c_2 < \dots < c_k$. We used Δ for the interval length if the class intervals are equal; i.e., $c_i = ic_1$.

METHODS BASED ON PERPENDICULAR DISTANCES

Ad Hoc Methods

EMLEN (1971, 1977) developed a non-mathematical approach to estimate the density of song birds on the basis of the assumption that a characteristic proportion of birds of any species will be detected in area $2Lw^*$. Given a fixed transect width w^* and $g(0) = 1$, he defined a coefficient of detectability CD as the pro-

portion detected. Thus, the reciprocal of CD is the average detection probability. \bar{CD} can be expressed as

$$\widehat{CD} = \frac{n}{w^*h(0)} = \frac{1}{w^*\hat{f}(0)},$$

where $h(0)$ is an estimator of $nf(0)$ determined from a smoothed frequency histogram of perpendicular distances. The estimator of density is a particular case of the general theory; i.e.,

$$\hat{D} = \frac{n}{2Lw^*\widehat{CD}} = \frac{nf(0)}{2L}.$$

The method typically uses data from only 2 to 4 distance groupings, often of unequal width (e.g., 0–20, 20–100, 100–400 m) and the estimation efficiency can be expected to be fairly low. The method requires unlikely assumptions regarding the constancy of the CD and probably is best considered as a crude index to abundance (also see Hayes unpublished thesis and Gates 1979a). Eberhardt (1978a) discussed the method and suggested a modification that may be useful. Although Emlen probably did not intend such an interpretation, his paper almost leads one to believe that small sample sizes of 10–15 might be adequate for density estimation (see Balph et al. 1977 in that same respect). Little can be learned from small samples *regardless* of the specific method. It seems that neither the suggested field procedure of using only 2 to 4 distance groups nor the analysis method will be very useful.

FRYE'S METHOD is best described by Overton and Davis (1969). Only an index to density is computed because only those objects observed within the distance interval 0 to \bar{x} are employed (i.e., $w^* \equiv \bar{x}$) in the otherwise standard estimation scheme; i.e.,

$$\text{index } \hat{D} = \frac{n'\hat{f}(0)}{2L}$$

and

$$f(0) = \frac{1}{\bar{x}}$$

where $n' < n$ and is the number of objects observed from the line to the average perpendicular distance. Its practical use often is similar to the use of Kelker's (unpublished dissertation) method. Frye's method is no longer useful as a practical technique unless, perhaps, only a crude index to density is required.

KELKER (unpublished dissertation) proposed a simple method to estimate density on the basis of a frequency histogram. By inspection, one determines the distance within which all animals or objects probably were observed but beyond which some were missed (Robinette et al. 1974). In that method,

$$\hat{D} = \frac{n\hat{f}(0)}{2L} = \frac{n_1}{2L\Delta} \text{ and } \hat{f}(0) = n_1/(n\Delta).$$

The method is well known and discussed by Eberhardt (1978a), Hayes (unpublished thesis) and Gates (1979a). It is a special case of the Truncated Histogram Estimators discussed later in PART 4. The method is very inefficient and Hayes (unpublished thesis) gave 2 reasons for potential bias. Furthermore, the procedure requires arbitrary grouping of the distance data x_i into class intervals. Balph et al. (1977) rediscovered the basic method. Studies by Robinette et al. (1974) indicate that this estimator performed quite well with respect to bias, at least for large sample sizes of immotile objects. In general, we believe the method is no longer very useful except as a rough guide to D .

LEOPOLD ET AL. METHOD (1951) is essentially like the estimator of Gates et al. (1968), which is based on the negative exponential distribution. The form of the density estimator is identical to the one proposed by Gates et al. (1968) except for the correction for bias

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where

$$\hat{f}(0) = \frac{1}{\bar{x}}.$$

That method has been superseded en-

tirely by the method of Gates et al. (1968) if the assumption of negative exponential sighting is valid.

Parametric Methods

THE NEGATIVE EXPONENTIAL DISTRIBUTION was used by Gates et al. (1968) as the basis for the first probabilistic treatment of line transect sampling theory. Based on a set of underlying assumptions, they found an optimal estimator of density as a function of the measured perpendicular distances. The detection function was assumed to be the negative exponential

$$g(x) = \exp(-ax), \quad 0 < x.$$

Their widely used method is applicable to untruncated, ungrouped data and has the form

$$\hat{D} = \frac{n\hat{f}(0)}{2L},$$

where

$$\hat{f}(0) = \frac{1}{\bar{x}} \left(\frac{n-1}{n} \right) \quad (\text{corrected for bias})$$

or

$$\hat{D} = \frac{n-1}{2L\bar{x}}.$$

The work of Gates et al. (1968) is a landmark paper in terms of its rigorous formalization of the statistical problem.

The method is excellent if the assumptions (particularly the negative exponential distribution) are met. Epstein (1960) provided some powerful tests for the negative exponential distribution without having to group the distance data and, therefore, causing a substantial loss of power of the test. We recommend the use of such tests before naively accepting the assumption that line transect data have an underlying negative exponential distribution.

As a useful method for the analysis of field data, the method is of limited use because it lacks *pooling robustness* and

model robustness and does not meet the *shape criterion*. The assumption of negative exponential sighting is too restrictive and the estimator is not robust to failure of this assumption (Gates 1969, 1979a; Robinette et al. 1974; Quinn unpublished thesis; Eberhardt 1978a). We believe the method has applicability for the analysis of data only if a good test of that assumption is made (cf. Gates 1979a). Such tests, as well as this estimation method, are available as options in program TRANSECT. In addition, we have extended the method to cover grouped and truncated data.

THE HALF-NORMAL DISTRIBUTION has been suggested by Hemingway (1971) as a realistic form for $g(x)$. The functional form for the half-normal detection function can be expressed as

$$g(x) = \exp(-ax^2) \quad 0 \leq x \leq \infty.$$

Although this form shows the close relation between the half-normal and negative exponential distributions, one usually employs the (equivalent) expression for the half-normal distribution,

$$g(x) = \exp(-x^2/2\sigma^2) \quad 0 \leq x \leq \infty.$$

Hayes (unpublished thesis) reviewed the development of that model, and Quinn (unpublished thesis) examined an estimator of density based on it and developed point estimators and sampling variances. In our general framework, the estimate of density for the half-normal distribution, if the data are ungrouped and untruncated, has

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

and

$$\hat{f}(0) = \left(\frac{\pi/2}{\sum x_i^2/n} \right)^{1/2} \cdot \frac{(n - 0.8)}{n}.$$

The estimate of $f(0)$ is expressed as an approximation because the term $(n - 0.8)/n$ is a close approximation to the optimal bias correction term given by Quinn. Hayes provided an alternative estimator of $f(0)$ based on the method of moments.

The half-normal is often more tenable than the negative exponential model. The derivative of the function is zero at the origin ($g'(0) = 0$) which corresponds to the assumption that an observer can see objects at least out to some small distance with probability 1, whereas the negative exponential model suggests that the probability of sighting drops off immediately away from the line. However, the half-normal model is extremely restrictive (as is the negative exponential) and we cannot recommend its general use. The procedure is, however, available as an option in program TRANSECT, including extensions to allow for the analysis of grouped and ungrouped and truncated and untruncated data.

THE POWER SERIES DISTRIBUTION was proposed by Eberhardt (1968) as a possible family of curves to model the sighting probability functions $g(x)$. His intent was to suggest a broad family of curves that vary from concave to convex. The form he proposed was

$$g(x) = 1 - (x/w)^a \quad 0 \leq x \leq w, \quad a \leq 0.$$

That detection function has $g(w) = 0$, which may be unrealistic unless the data are virtually untruncated. Estimation and inference are simple if w is considered known, however, it is often necessary to treat w as an unknown parameter. One approach is to estimate w by the largest order statistic; then, a ML solution for a can then be obtained by iterative numerical methods as outlined by Seber (1973:35). The estimate of density can be expressed as

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

and

$$\hat{f}(0) = \frac{\hat{a} + 1}{\hat{a}\hat{w}}.$$

Hayes (unpublished thesis) discussed that approach at some length. Eberhardt (1968) gave the form for ML estimation of both w and a , but the estimators do not exist in closed form.

The original work is important because it promoted a *family* of curves to cover a variety of shapes from concave to convex. The form of $g(x)$ is parametric and has undesirable characteristics near $g(w)$ that affect estimation of $f(0)$, including the fact that w is not known. For that reason, we believe better approaches are now available.

THE MODIFIED BETA DISTRIBUTION is a modification of the function suggested by Eberhardt (1968) such that

$$g(x) = (1 - x/w)^a \quad 0 \leq x \leq w, \quad a \geq 0.$$

That function provides a similar array of shapes, but the ML estimator of a exists in closed form. Also, the derived pdf $f(x)$ can be expressed as a modification of a beta distribution with parameters $(1, a+1)$ where the beta is described by

$$\beta(a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \left(\frac{x}{w}\right)^{a_1-1} \left(1 - \frac{x}{w}\right)^{a_2-1}$$

$$0 \leq x \leq w$$

and

$$a_1, a_2 > -1.$$

The modification is a simple transformation of variables which allows x to range in the interval $[0, w]$. An estimator for $f(0)$ is

$$\hat{f}(0) = \frac{\hat{a} + 1}{\hat{w}},$$

where \hat{w} is estimated, for example, by the maximum order statistic. That estimator is valid for ungrouped, untruncated data. We have studied the properties of that estimator through computer simulation and have found it to be fairly inflexible because of the constraint that $g(w) = 0$. Also, the estimator for w seems somewhat arbitrary. We do not recommend using this estimator.

THE GAMMA DISTRIBUTION was studied by Sen et al. (1974) as a generalized parametric extension of the method of Gates et al. (1968). Sen et al. (1974) assumed that

$$g(x) = \int_z^{\infty} \frac{a_2^{a_1}}{\Gamma(a_1)} x^{a_1-1} e^{-a_2 x} dx$$

for untruncated and ungrouped data. A mistake was made in their paper because the incomplete gamma is the correct model (see Burnham and Anderson 1976:330–331), so the results they presented are not useful. An estimate of $f(0)$ could be made from the incomplete gamma distribution and the general form would still apply, i.e.,

$$\hat{D} = \frac{n\hat{f}(0)}{2L}.$$

Use of those distributions may fit data well and their estimators can be expected to have small bias. Estimation efficiency is low, which represents a substantial drawback, and the numerical difficulties to be overcome are very formidable, especially for the grouped and truncated cases. We cannot recommend the use of that approach.

THE REVERSED LOGISTIC DISTRIBUTION has been suggested by Eberhardt (1968, 1978a) as a flexible model to represent the underlying sighting function $g(x)$. He used the reversed logistic to determine the sensitivity of the negative exponential estimator (Gates et al. 1968) to the assumption that $g(x)$ follows a negative exponential distribution. If the cdf of the logistic distribution is represented by

$$H(x) = \frac{1 - e^{-a_1 x}}{1 + a_2 e^{-a_1 x}}$$

$$0 \leq x < \infty, \quad a_1, a_2 > 0$$

for the positive values of x , an appropriate sighting function can be represented by

$$g(x) = 1 - H(x) = \frac{(1 + a_2)e^{-a_1 x}}{1 + a_2 e^{-a_1 x}}.$$

That function will assume the form of the negative exponential if $a_2 = 0$, the uniform distribution if $a_1 = 0$, and a variety of forms in between.

However, Eberhardt (1978a) did not develop an estimator based on the reversed logistic. Instead, he used it as a flexible model to generate data for simulation studies. If it is assumed that $g(x)$

follows a logistic distribution and the data are untruncated, the resulting estimator is

$$\hat{f}(0) = \frac{\hat{a}_1 \hat{a}_2}{\ell n(1 + \hat{a}_2)(1 + \hat{a}_2)}.$$

A ML approach to estimation of a_1 and a_2 requires an iterative numerical solution. The logistic appears to be a fairly flexible function, but it has been our experience that generalized parametric forms have relatively low estimation efficiencies (also see PART 3).

THE LOG-QUADRATIC ESTIMATOR developed by Anderson et al. (1978) was an attempt to provide a grouped data analog to the exponential quadratic estimator (see PART 2). The estimator assumes the underlying detection function is

$$g(x) = \exp(a_1 x + a_2 x^2)$$

$$0 \leq x \leq \infty, \quad a_1, a_2 \leq 0.$$

Assuming that $g(0) = 1$, $w = \infty$, and equal intervals, each cell count n_i is divided by the number seen in the first interval n_1 . That gives a crude discrete approximation to the sighting function $g(x)$. The probability in the i^{th} interval is approximately n_i/n_1 and it is weighted on the midpoint x_i . A log transformation of the n_i/n_1 puts the problem into the realm of the general linear model (see Graybill 1976).

Once the parameters have been estimated, the estimate of density can be obtained directly. In the general framework, it can be expressed as

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

and

$$\hat{f}(0) = \left(\int_0^{w*} \exp(\hat{a}_1 x + \hat{a}_2 x^2) dx \right)^{-1}.$$

The ML approach with multinomials is examined in PART 3 and is available as part of the program TRANSECT. The interested reader may wish to see the paper by Anderson, Burnham, and Crain (1979) on a related approach. We do not rec-

ommend the estimation procedure outlined by Anderson et al. (1978).

Nonparametric Methods

ANDERSON AND POSPAHALA (1970) proposed a general method for truncated sampling ($w < \infty$) to estimate a correction factor for objects not observed. A smooth curve $h(y)$, such as a polynomial, was fitted to the frequencies n_i , using the interval midpoints as the independent variable y . The estimated proportion of objects seen is

$$\hat{p} = \left[\int_0^w h(y) dy \right] / \hat{f}(0)w.$$

The estimated number of objects in the area $2Lw$ searched by the observer is then $\hat{N} = n/\hat{p}$ and can be written as

$$\hat{N} = \frac{n\hat{f}(0)w}{\int_0^w h(y) dy},$$

or

$$\hat{D} = \frac{n\hat{f}(0)}{2L \int_0^w h(y) dy}.$$

Because $\int_0^w h(y) dy \doteq n$ and $\hat{f}(0) = \hat{h}(0)/n$, the Anderson-Pospahala method is formally expressible as

$$\hat{D} = \frac{n\hat{f}(0)}{2L}.$$

Gates (1979a) discussed the method and noted that it could be improved if $h(y)$ were made to integrate to n .

Anderson and Pospahala (1970) attempted to illustrate the method by choosing a simple polynomial for the analysis of data on duck nests. After some analysis of polynomial regressions, they chose

$$h(y) = a_1 + a_2 y^2.$$

Using the same approach (i.e., selecting an appropriate simple polynomial regression), Robinette et al. (1974) found that the method performed well with respect

to bias in replicated field experiments of inanimate objects. Additional information on this method can be found in Anderson et al. (In press), Hayes (unpublished thesis), and extensions are derived by Gates (1979a).

Pollock (1978) found that an estimator based on the exponential power series for grouped data was similar to the Anderson-Pospahala (1970) method. A polynomial type estimator, similar to the Anderson-Pospahala procedure, can also be derived from a jackknife procedure. We do not recommend the algorithm used in the Anderson-Pospahala paper. However, the use of ordinary (or orthogonal) polynomials to model $g(x)$ looks very promising. A general procedure for obtaining ML estimates for the polynomial procedure (without the linear term) is outlined in PART 3. Such a polynomial expression of $g(x)$ satisfies *model robustness*, *pooling robustness*, and the *shape criterion* and the *estimator efficiency* is excellent.

Cox's METHOD was derived for line transect studies by Eberhardt (1978a), based on the work of Cox (1969). The estimator is based on grouped data, but uses data for only the 2 groups nearest the line. For group intervals of arbitrary size, the general form of the estimation procedure is

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where

$$\hat{f}(0) = \frac{(c_1 + c_2)}{c_1 c_2} \left(\frac{n_1}{n} \right) - \frac{c_1}{c_2(c_2 - c_1)} \left(\frac{n_2}{n} \right).$$

If the 2 class intervals are of equal length, say Δ , then the estimator of density (see Eberhardt 1978a) is simply

$$\hat{D} = \frac{3n_1 - n_2}{4L\Delta}.$$

That method is a special case of a general class of estimators we call TRUNCATED HISTOGRAM ESTIMATORS (discussed below). Note that the intervals $c_{i+1} - c_i$ need not be equal in the general theory. The

method that Eberhardt calls the Cox method has the properties of *pooling* and *model robustness* but has a large variance and requires substantial subjectivity in the selection of Δ , particularly if the data were not originally grouped. In addition, the *shape criterion* $f'(0) = 0$ is not met with that procedure. The method has the advantage that it is very easy to compute. Eberhardt (1978a) believed the method to be useful mainly as a cross-check and when other methods fail.

TRUNCATED HISTOGRAM ESTIMATORS represent a broad class of estimators based on the n_1, n_2, \dots, n_{k+1} in the intervals $0 < c_1 < c_2 < \dots < c_{k+1}$, respectively. In general,

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where

$$\hat{f}(0) = \sum_{i=1}^{k+1} a_i \left(\frac{m_i}{nc_i} \right)$$

where $m_i = n_1 + n_2 + \dots + n_i$, and

$$a_i = \frac{(-1)^{i-1} \left(\frac{1}{c_i} \right)^k}{\prod_{j=1}^{i-1} \left(\frac{1}{c_j} - \frac{1}{c_i} \right) \prod_{j=i+1}^{k+1} \left(\frac{1}{c_i} - \frac{1}{c_j} \right)},$$

where terms such as $\prod_{j=1}^0$ and $\prod_{j=k+2}^{k+1}$ are defined to be 1. Estimates of the sampling variance of $\hat{f}(0)$ can be derived by using the fact that the n_i are multinomial random variables.

It is interesting to note that this expression for $\hat{f}(0)$ can be derived in several ways:

- (1) by a Taylor series approximation to $g(x)$,
- (2) by assuming a k^{th} order polynomial as an approximation to $g(x)$,
- (3) by fitting a k^{th} order polynomial to the sample frequency data (n_i), or,
- (4) by using the theory of the generalized jackknife on the frequency data.

Although mathematically equivalent, those approaches are conceptually different. No goodness of fit test is possible for that class of methods. The Cox and Kelk-

er methods are special cases of that general class. Those methods all require important and subjective decisions to specify k and c_1, c_2, \dots, c_{k+1} , and the estimator can depend critically upon those choices. Although the methods may be robust, their efficiency appears to be poor and their subjectivity is high, particularly if only 2 or 3 groups are chosen. We do not recommend the use of estimators based on the truncated histogram approach.

JACKKNIFE ESTIMATION has been investigated, based on a series of naive density estimators, as

$$\hat{D}_i = \frac{m_i}{2Lc_i}, \quad i = 1, \dots, k + 1$$

where c_i are the frequency intervals; $0 = c_0 < c_1 < c_2 \dots < c_{k+1} = w$. Because the n_i tend to decrease, one sees that bias in the naive density estimates will increase. Hence, a basis for "bias reduction" is possible by using jackknife techniques (see Gray and Schucany 1972) on the series of estimates \hat{D}_i .

The expected value of \hat{D}_i is

$$\begin{aligned} E(\hat{D}_i) &= \frac{E(m_i)}{2Lc_i} = \frac{D2Lc_i}{2Lc_i} \int_0^{c_i} g(x)dx \\ &= D \int_0^{c_i} g(x)dx. \end{aligned}$$

The integral given above can be approximated by the first k terms of a Taylor series expansion of $g(x)$. After integration,

$$E(\hat{D}_i) = D[c_i + a_1(c_i)^2 + \dots + a_k(c_i)^{k+1}], \quad i = 1, \dots, k + 1$$

is obtained, where the a_j are unknown parameters, although structurally one has

$$a_j = \frac{g^{j-1}(0)}{j!}.$$

The $k+1$ equations for $E(\hat{D}_i)$ have $k+1$ unknown parameters (D, a_1, \dots, a_k). If the function $g(x)$ is exactly expressible as a $k+1^{\text{th}}$ order polynomial, the system of equations can be solved for \hat{D} that is unbiased. Even when the polynomial is

only an approximation, the density estimator will have a smaller order of bias relative to $\hat{D}_1, \dots, \hat{D}_{k+1}$.

Estimators derived from the generalized jackknife are identical to those we have called the truncated histogram estimators. Conditional on n , the n_i are multinomial random variables, and estimators of sampling variances can be derived. No goodness of fit test can be developed and substantial subjectivity is required, especially if the data were taken as ungrouped. We do not recommend estimators based on the jackknife method.

ISOTONIC REGRESSION was suggested by Burnham and Anderson (1976) as a robust estimation procedure for line transect data (see Barlow et al. 1972). The procedure is based on the order statistics and can be used on grouped or ungrouped data regardless of the value of w . In all cases, density can be estimated from

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where, for ungrouped data,

$$\hat{f}(0) = \max_i \left\{ \frac{i}{nX_{(i)}} \right\}$$

and $X_{(i)}$ is the i^{th} order statistic of the perpendicular distances x_i . For grouped data, the form is

$$\hat{f}(0) = \max_j \left(\sum_{i=1}^j \frac{n_i}{n\Delta_j} \right).$$

That form assumes the groups are of equal size; i.e., $c_i - c_{i-1} = \Delta$ for all i , however, generalization to unequal group sizes is straightforward.

The potential advantage of an isotonic regression estimator is that it forces $g(x)$ to be monotonically nonincreasing. That is useful if some animals were moving slightly from the center line before being observed. Estimation efficiencies are extremely low, particularly for ungrouped data (Hayes unpublished thesis). The scheme has very poor properties, especially for ungrouped data, and we do not recommend its use.

METHODS BASED ON SIGHTING ANGLES AND DISTANCES

The methods that have been proposed for this class of estimation methods have assumed untruncated data ($w = \infty$) and ungrouped sighting distance and angle data. Methods by King (Overton and Davis 1969) and Hayne (1949) do not allow for grouping or truncation. Parametric models (e.g., Gates 1969) could be extended to cover both grouped and truncated data, but apparently that has never been done. Several of the early methods were based on only the sighting distances. However, we believe strongly that both sighting distances and sighting angles θ_i are needed for a satisfactory estimation method. Hayne's (1949) method is covered briefly, but the reader is referred to PARTS 2 and 3 where more detail is given and 2 generalizations are derived.

In general, we believe the best analysis and inference procedures for line transect data employ perpendicular distance data. Attempts to model surveys that involve sighting distances and angle data are really not satisfactory. We believe Hayne's (1949) method and recent extensions are somewhat useful, but not as good as methods available for treating perpendicular distance data.

KING'S METHOD (Leopold 1933, Overton and Davis 1969) might be called the classic of early line transect methodology. His ad hoc method is based only on the ungrouped and untruncated measurements of the sighting distances r_i :

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where

$$\hat{f}(0) = \frac{1}{\bar{r}}$$

Hayne (1949) showed that the observed value of \bar{r} is a biased estimator of the average flushing distance of the population under the assumption of a *circular* flushing radius.

Gates (1969) reviewed the method and Kovner and Patil (1974) examined its statistical properties under the assumption of a negative exponential distribution, the methods described by Gates (1969) are superior in all respects. Under simulated field conditions, Robinette et al. (1974) found that the King estimator performed fairly well, as did the methods of Gates (1969).

We believe it is essential to measure both r_i and θ_i (or r_i and x_i) and to base an estimator on both measurements. Otherwise, very restrictive and unlikely assumptions are required. For that reason, we do not recommend the use of the King method.

HAYNE'S METHOD (1949) is an estimator based on ungrouped, untruncated data on the sighting distances r_i . The method has the usual form

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where

$$\hat{f}(0) = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i}$$

This useful and popular method has been discussed by Gates (1969), Overton and Davis (1969), Seber (1973), Kovner and Patil (1974), Burnham and Anderson (1976), and Eberhardt (1978a). The method is often derived by assuming a fixed, circular flushing distance. The critical assumption of the model is that $\sin(\theta)$ is a uniform random variable on the interval $[0,1]$. Under those assumptions $E(\theta) = 32.7^\circ$. Eberhardt (1978a) and others have indicated the Hayne estimator is fairly unbiased over a range of underlying distributions. We believe it is not a good procedure if θ is not approximately 32.7° . Often in field data, we have found $\theta > 32.7^\circ$; often being nearer to 40° . Use of the Hayne method on studies of inanimate objects failed to produce satisfactory results (Robinette et al. 1974). In general, we feel the method is useful in some cases but that its value has been overrated in the published literature.

Several workers (e.g., Seber 1973, Kovner and Patil 1974, Quinn unpublished thesis) noted that the $\text{var}(\hat{D})$ does not exist under Hayne's model because $E\left(\left(\frac{1}{r_i}\right)^2\right) = \infty$. This is true only for isolated underlying probability models for r (e.g., negative exponential) and is not true in general. Hayne's model is based on the concept of a flushing radius for each object. Therefore, in theory, if not by convention, $r_i > 0$ and we see that an empirical estimator of the sampling variance of $f(0)$ can be derived by using

$$\hat{f}_i(0) = \frac{1}{r_i}$$

and the average is $\hat{f}(0)$. Then, from Burnham and Anderson (1976:332), we see that

$$\hat{\text{var}}(\hat{f}(0) | n) = \frac{\sum_{i=1}^n [\hat{f}_i(0) - \hat{f}(0)]^2}{n(n-1)}.$$

Cox (1969:515) discussed a similar problem in a similar context (also see Hayes unpublished thesis). We believe the estimators of D or $f(0)$ under Hayne's model have a sampling variance that can be estimated.

The *pooling robustness* property is satisfied by the Hayne method, and this is the primary reason for its robustness, *if sin(θ) is uniform*. However, the estimator is sensitive to very small values of the sighting distance.

Even after nearly 30 years since it was developed, we believe Hayne's method still has some merits. However, we see the 2 extensions of Hayne's model (see PARTS 2 and 3) as substantial improvements. Still, those methods for sighting angles and distances are not on a par with those for perpendicular distance data.

In Hayne's model, estimation is possible without the sighting angles θ_i , but testing requires the angle measurements. The 2 extensions of the method require θ_i for *both* estimation and testing. Those 3 estimation methods are available in

program TRANSECT for ungrouped and untruncated data.

THE NEGATIVE EXPONENTIAL METHODS derived by Gates (1969) require only ungrouped, untruncated ($w = \infty$) measurements of the sighting distances r_i . Both estimators are of the usual form

$$\hat{D} = \frac{n\hat{f}(0)}{2L}.$$

Gates (1969) found a relation between estimators of density; i.e.,

$$\begin{aligned} \text{King} &< \text{Gates}_{(\text{Geom.})} \\ &< \text{Hayne} < \text{Gates}_{(\text{Arith.})}. \end{aligned}$$

Seber (1973), Kovner and Patil (1974), and Gates (1979a) discussed those methods. However, the most comprehensive description is to be found in the original paper (Gates 1969). The methods have a good theoretical basis and are essentially optimal estimation schemes under the assumption of an underlying negative exponential distribution.

We believe those methods to be useful in only unusual situations and after careful testing of the critical assumption that the sighting function is negative exponential. We maintain it is essential to measure either r_i and θ_i (or r_i and x_i) if those types of models are to be useful. In summary, we regard those 2 methods as clearly conceived and fully developed; however, the restrictive parametric assumptions seem unlikely to hold in practice. Further work in this direction is needed, with special attention to the auxiliary use of the sighting angles θ_i . The method could be extended to cover sighting distance data that are grouped.

WEBB'S METHOD (1942) is similar to the methods proposed by Leopold et al. (1951) and Gates et al. (1968). The estimate of density is

$$\hat{D} = \frac{n\hat{f}(0)}{2L}$$

where

$$\hat{f}(0) = \frac{1}{\bar{r} \sin(\bar{\theta})}.$$

The expression $\bar{r} \sin(\bar{\theta})$ is approximately equal to \bar{x} . Hayne (1949) discussed some aspects of Webb's method and the method was reviewed by Hayes (unpublished thesis). We believe the method has been superseded by the method of Gates et al. (1968).

YAPP'S METHOD (1956) relates to a dynamic system whereby points (objects) and the observer are moving in random directions at various velocities. The theory was developed along the lines of Brownian motion. No practical results were found, and we do not expect any under such a complex situation. Skellam's (1958) method was primarily a discussion and elaboration of Yapp's (1956) work.

In general, the Brownian motion approach (random movement) does not address the real situation. Generally, animals move toward or away from the line and the observer rather than exhibiting random movement. For example, some species of porpoise seem to be attracted to ships, and Hayes (unpublished thesis) gave a similar example involving giraffes. Rabbits and hares often move away from the line and the observer for some distance before being detected. Random movement is perhaps far more the exception than the rule.

LITERATURE CITED

- ANDERSON, D. R., K. P. BURNHAM, AND B. R. CRAIN. 1978. A log-linear model approach to estimation of population size using the line transect sampling method. *Ecology* 59(1):190-193.
- _____, ____, AND _____. 1979. Line transect estimation of population size: The exponential case with grouped data. *Commun. Statist. - Theor. Meth.* A8(5):487-507.
- _____, ____, AND _____. 1980. Some comments on Anderson and Pospahala's correction of bias on line transect sampling. *Biomet. J.* In press.
- _____, J. L. LAAKE, B. R. CRAIN, AND K. P. BURNHAM. 1976. Guidelines for line transect sampling of biological populations. Utah St. Univ. Press, Logan, Utah. 27 pp.
- _____, ____, ____, AND _____. 1979. Guidelines for line transect sampling of biological populations. *J. Wildl. Manage.* 43(1):70-78.
- _____, AND R. S. POSPAHALA. 1970. Correction of bias in belt transects of immobile objects. *J. Wildl. Manage.* 34(1):141-146.
- ANDREWS, D. F., P. J. BICKEL, F. R. HAMPEL, P. J. HUBER, W. H. ROGERS, AND J. W. TUKEY. 1972. Robust estimates of location. Princeton Univ. Press, Princeton, N. J. 373 pp.
- AOKI, M. 1971. Introduction to optimization techniques. The Macmillan Co. New York, N.Y. 335 pp.
- BALPH, M. H., L. C. STODDART, AND D. F. BALPH. 1977. A simple technique for analyzing bird transect data. *Auk* 94(3):606-607.
- BARLOW, R. E., D. J. BARTHOLOMEW, J. M. BRENNER, AND H. D. BRUNK. 1972. Statistical inference under order restrictions. John Wiley & Sons, New York, N.Y. 388 pp.
- BONEVA, L. J., D. G. KENDALL, AND I. STEFANOV. 1971. Spline transformations: Three new diagnostic aids for the statistical data-analyst. *J. Roy. Statist. Soc. b* 33(1):1-71.
- BRADLEY, J. V. 1968. Distribution-free statistical tests. Prentice-Hall, Inc., Englewood Cliffs, N.J. 388 pp.
- BRUNK, H. D. 1976. Univariate density estimation by orthogonal series. Dept. Statist., Oregon St. Univ., Tech. Rept. 51.
- _____. 1978. Univariate density estimation by orthogonal series, with application to estimation of wildlife populations by line transect surveys. *Biometrika* 66(3):521-528.
- BURNHAM, K. P. 1979. A parametric generalization of the Hayne estimator for line transect sampling. *Biometrics* 35(3):587-595.
- _____, AND D. R. ANDERSON. 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics* 32(2):325-336.
- _____, _____, AND J. L. LAAKE. 1979. On robust estimation from line transect data. *J. Wildl. Manage.* 43(4):992-996.
- CHURCHILL, R. V. 1963. Fourier series and boundary value problems. McGraw-Hill Book Co., New York, N.Y. 206 pp.
- COCHRAN, W. G. 1963. Sampling techniques. John Wiley & Sons, Inc., New York, N.Y. 413 pp.
- _____, AND G. M. COX. 1957. Experimental design. John Wiley & Sons, Inc., New York, N.Y. 611 pp.
- COX, D. R. 1958. Planning of experiments. John Wiley & Sons, Inc., New York, N.Y. 308 pp.
- _____. 1969. Some sampling problems in technology. Pp. 506-527. In N. L. Johnson (Ed.). New developments in survey sampling. Wiley-Interscience, New York, N.Y. 732 pp.
- CRAIN, B. R. 1973. A note on density estimation using orthogonal expansions. *J. Amer. Statist. Ass.* 68(4):964-965.
- _____. 1974. Estimation of distributions using orthogonal expansions. *Ann. Statist.* 2(3):454-463.
- _____, K. P. BURNHAM, D. R. ANDERSON, AND J. L. LAAKE. 1978. A Fourier series estimator of population-density for line transect sampling. Utah St. Univ. Press, Logan, Utah. 25 pp.

- DOI, T. 1970. Re-evaluation of population studies by sighting observation of whale. *Bull. Tokai Reg. Fish. Lab.* 63:1-9.
- . 1974. Further development of whale sighting theory. Pp. 359-368. In W. E. Schevill (Ed.). *The whale problem*. Harvard Univ. Press, Cambridge, Mass. 419 pp.
- . 1975. Further development of sighting theory on whale. *Bull. Tokai Reg. Fish. Lab.* 68:1-22.
- EBERHARDT, L. L. 1968. A preliminary appraisal of line transects. *J. Wildl. Manage.* 32(1):82-88.
- . 1978a. Transect methods for population studies. *J. Wildl. Manage.* 42(1):1-31.
- . 1978b. Appraising variability in population studies. *J. Wildl. Manage.* 42(2):207-238.
- EKLUND, C. R., AND E. L. ATWOOD. 1962. A population study of antarctic seals. *J. Mammal.* 43(2):229-238.
- EMLEN, J. T. 1971. Population densities of birds derived from transect counts. *Auk* 88(2):323-342.
- . 1977. Estimating breeding season bird densities from transect counts. *Auk* 94(3):455-468.
- EPSTEIN, B. 1960. Tests for the validity of the assumption that the underlying distribution of life is exponential: Part I. *Technometrics* 2(1):83-101.
- FORBES, S. A., AND A. O. GROSS. 1921. The orchard birds of an Illinois summer. *Ill. Nat. Hist. Surv. Bull.* 14:1-8.
- GATES, C. E. 1969. Simulation study of estimators for the line transect sampling method. *Biometrika* 56(2):317-328.
- . 1979a. Line transects and related issues. Pp. 71-154. In R. M. Cormack, G. P. Patil, and D. S. Robson (Eds.). *Sampling biological populations*. Internat'l. Co-op. Publ. House, Fairland, Md.
- . 1979b. LINETRAN user's guide. Statist. Lab. No. 133, Univ. Wyoming. 33 pp.
- . W. H. MARSHALL, AND D. P. OLSON. 1968. Line transect method of estimating grouse population densities. *Biometrics* 24(1):135-145.
- GRAY, H. L., AND W. R. SCHUCANY. 1972. The generalized jackknife statistic. Marcel Dekker, Inc., New York, N.Y. 308 pp.
- GRAYBILL, F. A. 1976. Theory and application of the linear model. Duxbury Press, North Scituate, Mass. 704 pp.
- GROSS, J. E., L. C. STODDART, AND F. H. WAGNER. 1974. Demographic analysis of a northern Utah Jackrabbit population. *Wildl. Monogr.* No. 40:1-68.
- HAYNE, D. W. 1949. An examination of the strip census method for estimating animal populations. *J. Wildl. Manage.* 13(2):145-157.
- HEMINGWAY, P. 1971. Field trials of the line transect method of sampling large populations of herbivores. Pp. 405-411. In E. Duffey and A. S. Watts (Eds.). *The scientific management of animal and plant communities for conservation*. Blackwell Sci. Publ., Oxford, Eng.
- HOBSON, R. 1976. Properties preserved by some smooth functions. *J. Amer. Statist. Ass.* 71:763-766.
- HOLLANDER, M., AND D. A. WOLFE. 1973. Nonparametric statistical methods. John Wiley & Sons, Inc., New York. N.Y. 503 pp.
- HOLT, R. S., AND J. E. POWERS. 1980. Abundance estimation of dolphin stocks involved in the Eastern Tropical Pacific yellowfin tuna fishery determined from aerial and ship surveys. *Natl. Mar. Fish. Serv.* In Press.
- JARVINEN, O., AND R. A. VAISANEN. 1975. Estimating relative densities of breeding birds by the line transect methods. *Oikos* 26:316-322.
- JOHNSON, N. L., AND S. KOTZ. 1969. Distributions in statistics: Discrete distributions. Houghton Mifflin Co., Boston, Mass. 328 pp.
- KOVNER, J. L., AND S. A. PATIL. 1974. Properties of estimators of wildlife population density for the line transect method. *Biometrics* 30(2):225-230.
- KRONMAL, R., AND M. TARTER. 1968. The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Ass.* 63(3):925-952.
- LAAKE, J. L., K. P. BURNHAM, AND D. R. ANDERSON. 1979. User's manual for program TRANSECT. Utah St. Univ. Press, Logan, Utah. 26 pp.
- LEATHERWOOD, S., J. R. GILBERT, AND D. G. CHAPMAN. 1978. An evaluation of some techniques for aerial censuses of bottlenosed dolphins. *J. Wildl. Manage.* 42(2):239-250.
- LEOPOLD, A. 1933. Game management. Charles Scribner's Sons, New York, N.Y. 481 pp.
- LEOPOLD, A. S., T. RINEY, R. MCCAIN, AND L. TEVIS, JR. 1951. The Jawbone deer herd. Cal. Div. Fish Game, Game Bull. No. 4:1-139.
- MAISEL, H., AND G. GNUGNOLI. 1972. Simulation of discrete stochastic systems. *Sci. Res. Ass.*, Inc. 465 pp.
- MARQUARDT, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* 11(2):431-441.
- MILLER, R. G. 1974. The jackknife—a review. *Biometrika* 61(1):1-15.
- MYRBERGET, S. 1976. Field tests of line transect methods for grouse. *Norw. J. Zool.* 24:307-317.
- OVERTON, W. S., AND D. E. DAVIS. 1969. Estimating the number of animals in wildlife populations. Pp. 405-455. In R. H. Giles, Jr. (Ed.). *Wildlife management techniques*. The Wildlife Society, Washington, D.C.
- PATIL, G. P., C. TAILLIE, AND R. L. WIGLEY. 1979. Transect sampling methods and their application to the deep-sea red crab. Pp. 51-75. In R. M. Cormack, G. P. Patil, and D. S. Robson (Eds.). *Sampling biological populations*. Internat'l. Co-op. Publ. House, Fairland, Md.
- PATIL, S. A., K. P. BURNHAM, AND J. L. KOVNER. 1979. Nonparametric estimation of plant den-

- sity by the distance method. *Biometrics* 35(3):597–604.
- POLLOCK, K. H. 1978. A family of density estimators for line-transect sampling. *Biometrics* 34(3):475–478.
- RAMSEY, F. L. 1979. Parametric models for line-transect surveys. Dept. Statist., Oregon St. Univ., Tech. Rept. 65. 28 pp.
- , AND J. M. SCOTT. 1979. Estimating population density for variable circular plot surveys. Pp. 155–181. In R. M. Cormack, G. P. Patil, and D. S. Robson (Eds.). *Sampling biological populations*. Internat'l. Co-op. Publ. House, Fairland, Md.
- RAO, C. R. 1957. Maximum likelihood estimation for the multinomial distribution. *Sankhyā* 18:139–148.
- . 1973. Linear statistical inference and its applications. John Wiley & Sons, Inc., New York, N.Y. 625 pp.
- ROBINETTE, W. L., D. A. JONES, J. S. GASHWILER, AND C. M. ALDOUS. 1954. Methods for censusing winter-lost deer. *Trans. N. Amer. Wildl. Conf.* 19:511–525.
- , —, —, AND —. 1956. Further analysis of methods for censusing winter-lost deer. *J. Wildl. Manage.* 20(1):75–78.
- , —, AND C. M. LOVELESS. 1974. Field tests of strip census methods. *J. Wildl. Manage.* 38(1):81–96.
- ROSENBLATT, M. 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27(1):832–837.
- ROYAMA, T. 1960. The theory and practice of line transects in animal ecology by means of visual and auditory recognition. *Rep. Yamashima Inst. Ornith. Zool.* 2(14):1–17.
- RUSCH, D. H., AND L. B. KEITH. 1971. Seasonal and annual trends in numbers of Alberta ruffed grouse. *J. Wildl. Manage.* 35(4):803–821.
- SCHWEDER, T. 1977. Point process models for line transect experiments. Pp. 221–242. In J. R. Barba, F. Brodeau, G. Romier, and B. Van Cutsem (Eds.). *Recent developments in statistics*. North-Holland Publ. Co., New York, N.Y.
- SEBER, G. A. F. 1973. The estimation of animal abundance. Hafner Publ. Co., Inc., New York, N.Y. 506 pp.
- . 1979. Transects of random length. Pp. 183–192. In R. M. Cormack, G. P. Patil, and D. S. Robson (Eds.). *Sampling biological populations*. Internat'l. Co-op. Publ. House, Fairland, Md.
- SEN, A. R., J. TOURIGNY, AND G. E. J. SMITH. 1974. On the line transect sampling method. *Biometrics* 30(3):329–340.
- SKELLAM, J. G. 1958. The mathematical foundations underlying the use of line transects in animal ecology. *Biometrics* 14(3):385–400.
- SMITH, G. E. J. 1979. Some aspects of line transect sampling when the target population moves. *Biometrics* 35(1):323–329.
- SMITH, T. D. 1975. Estimates of sizes of two populations of porpoise (*Stenella*) in the Eastern Tropical Pacific Ocean. *Southwest Fish. Cent. Admin. Rept.* LJ-75-67.
- STEELE, R. G. D., AND J. H. TORRIE. 1960. *Principles and procedures of statistics*. McGraw-Hill Book Co., New York, N.Y. 481 pp.
- TAPIA, R. A., AND J. R. THOMPSON. 1978. Nonparametric probability density estimation. The Johns Hopkins Press, Baltimore, Md. 176 pp.
- TARTER, M. E., AND R. A. KRONMAL. 1976. An introduction to the implementation and theory of nonparametric density estimation. *Amer. Statist.* 30(8):105–112.
- TOLSTOV, G. P. 1962. *Fourier series*. Prentice-Hall, Englewood Cliffs, N.J. 336 pp.
- VANCE, E. P. 1954. *Trigonometry*. Addison-Wesley, New York, N.Y. 158 pp.
- WEBB, W. L. 1942. Notes on a method of censusing snowshoe hare populations. *J. Wildl. Manage.* 6(1):67–69.
- WEGMAN, E. J. 1972a. Non-parametric probability density estimation: I. A summary of available methods. *Technometrics* 14(3):533–546.
- . 1972b. Nonparametric probability density estimation. II. A comparison of density estimation methods. *J. Statist. Comp. Simul.* 1:225–245.
- YAPP, W. B. 1956. The theory of line transects. *Bird Study* 3(2):93–104.

APPENDIX A

SYMBOLS, DEFINITIONS, AND ABBREVIATIONS

List of Symbols

a_i	Unknown parameters to be estimated, $i = 1, \dots, m$
x_i	Perpendicular distance, $i = 1, \dots, n$ (also called right angle or lateral distances)
r_i	Sighting (or flushing) distance, $i = 1, \dots, n$
\bar{r}_h	The average of the reciprocals of the sighting distances, $= \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i}$
θ_i	Sighting (or flushing) angle, $i = 1, \dots, n$
y_i	$\sin(\theta_i) = x_i/r_i$
ℓ_j	Length of an individual line
L	Total length of the transect $L = \sum \ell_j$
R	Number of replicate lines

n_i	Number of objects observed in the i^{th} interval (grouped data), $i = 1, \dots, k$ (occasionally used to denote the number of objects n_j observed on the j^{th} line ℓ_j)	$f(x)$	The probability density function (pdf) of the perpendicular distances
n	Total number of objects seen $= \sum_{j=1}^k n_j$	$f(0)$	The pdf at zero distance from the line
m_i	Cumulative count in the intervals $1, \dots, i$; i.e., $m_i = \sum_{j=1}^i n_j$	$f'(0)$	The first derivative of the pdf $f(x)$ at zero distance from the line
c_i	Boundaries for the class intervals for grouped data. The x_i data would be grouped into (perhaps unequal) intervals $0 < c_1 < c_2 < \dots < c_k$.	$\Gamma(\cdot)$	The gamma function
$\hat{\theta}$	Denotes an estimator	$\text{var}(\hat{\theta})$	The sampling variance for some parameter estimator, say $\hat{\theta}$
D	Average density of the population under study	$\text{cov}(\hat{\theta}_1, \hat{\theta}_2)$	The sampling covariance of 2 parameter estimators, say $\hat{\theta}_1$ and $\hat{\theta}_2$
\hat{D}_H	Density estimator for Hayne's (1949) model	$\text{var}(n)$	The sampling variance of the (random) sample size
\hat{D}_{MH}	Density estimator for the modified Hayne model	$\hat{\theta}$	The "hat" ($\hat{\cdot}$) denotes an estimator of θ
\hat{D}_{GH}	Density estimator for the generalized Hayne model	$\text{se}(\hat{\theta})$	The standard error of the estimator $\hat{\theta}$ ($= \sqrt{\text{var}(\hat{\theta})}$)
\hat{D}_{FS}	Density estimator for the Fourier series model	s	Group size, clustered populations (school, herds, ponds, prides, coveys, packs, etc.)
N	The size of the population under study, $N = DA$ assuming a well-defined study area	π	3.14159 ...
w	The width of the transect, determined before the sample is taken. The width can be finite (e.g., 25 m) or unbounded.	p_i	Cell probability in the multinomial distribution for grouped data
w^*	A perpendicular distance such that all x_i greater than w^* are discarded in the analysis, $w^* \leq w$.	$\binom{n}{y}$	Binomial coefficient =
μ	Effective strip width (ESW) or $1/f(0)$		$\frac{n!}{y!(n-y)!}$
Δ	The class interval length for grouped data if all intervals are of equal length	$Z_{(i)}$	The i^{th} order statistic, $i = 1, \dots, n$ of a sample $Z_{(1)}, \dots, Z_{(n)}$
$g(x)$	The detection function that represents the conditional probability of observing a point, given it is at perpendicular distance x from the line. Note, $g(0) = 1$ by assumption.	\equiv \doteq \neq $E(\cdot)$ $\sum_{i=1}^k$ $\prod_{i=1}^k$ $\exp(x)$ \approx $ a_k $ z $\mathcal{L}(\cdot)$	Defined to be Approximately equal to Not equal to Expectation operator Summation operator Product operator e^x Infinity The absolute value of a_k A test statistic distributed as a normal distribution with mean 0 and variance 1 Likelihood function

List of Abbreviations

cdf	Cumulative distribution function, $H(x)$ or $F(x)$
cv	Coefficient of variation
CI	Confidence interval
EPOLY	Exponential polynomial function
EPS	Exponential power series
ESW	Effective strip width
FS	Fourier series
HN	Half-normal function
ISREG	Isotonic regression
MB	Modified beta function
MISE	Mean integrated squared error
ML	Maximum likelihood
MSE	Mean square error
NE	Negative exponential function (also NEXP)
pdf	Probability density function
Pr(e)	Probability of event e
PRB	Percentage relative bias
RE	Relative efficiency
SA	Sighting angle
SD	Sighting distance

detection functions used in the simulation studies. The data differ from sample to sample because of sampling variation. Two sample sizes (40 and 100) and 2 different groupings are illustrated. The 4 replicates shown in each figure were chosen randomly from 25 generated in the computer simulation studies. Apparent “outliers” are especially noticeable in Fig. B4.

Even with sample sizes of 100, one has difficulty in inferring the true underlying detection function (e.g., the underlying function in Fig. B5 is linear, but that is not at all obvious from the data). The numbers and sizes of the groups can make the data appear markedly different (e.g., Fig. B1). Data from a negative exponential detection function often appear nearly half-normal if samples are small (e.g., Fig. B3, second and third from the top left). Data from an underlying half-normal detection function can take on an array of possible shapes, again depending on the grouping selected. Even movement away from the line may be suspected (e.g., Fig. B1, second from bottom left; Fig. B2, second from top left) when, in fact, no movement occurred. Those results support the need for robust estimation methods that are not dependent on subjective examination of the data. This is further necessary because the power of statistical goodness of fit tests is not great for determining the “correct” model from an assessment of grouped data.

Fig. B7 shows the illusion that data come from an underlying negative exponential distribution as larger intervals are employed for grouping. This also supports the need for the *shape criterion* to be used in modeling line transect data.

APPENDIX B

SAMPLE HISTOGRAMS OF VARIOUS SIMULATED DATA SETS

This appendix is intended to provide insight into the stochastic (variable) nature of sample data from line transect studies and it presents some of the data (shown as frequency histograms) used in our computer simulation studies.

The first 6 figures (Figs. B1–B6) illustrate the stochastic nature of samples of observed perpendicular distances for 3

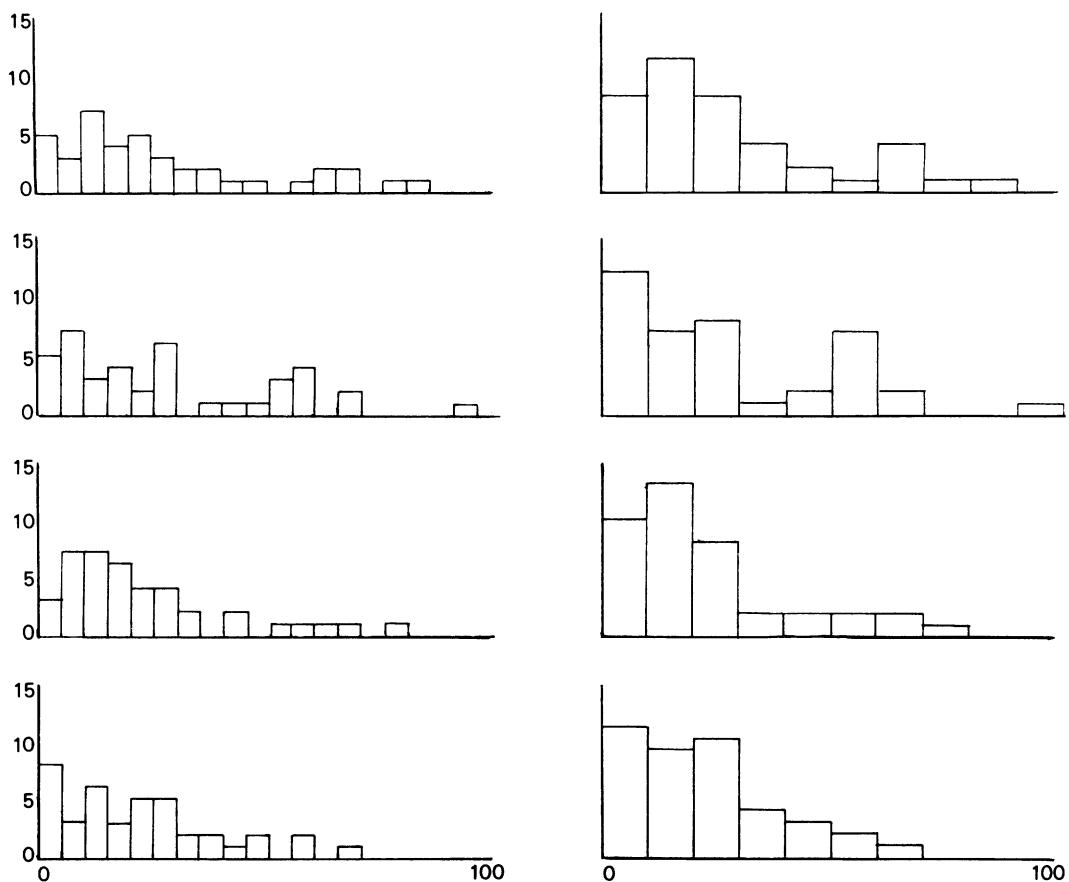


FIG. B1. Four sample histograms for samples of size 40 from an underlying half-normal detection function with $\sigma=33.3$ and $w=100$ m (essentially no truncation of the data). The left column shows the perpendicular distance data grouped into intervals of 5 m while the right column shows the same data grouped into intervals of 10 m. The data represent 4 data sets of the 25 used in the computer simulation studies with this detection function. The results are independent of the measurement unit; here, the units are arbitrarily denoted as meters.

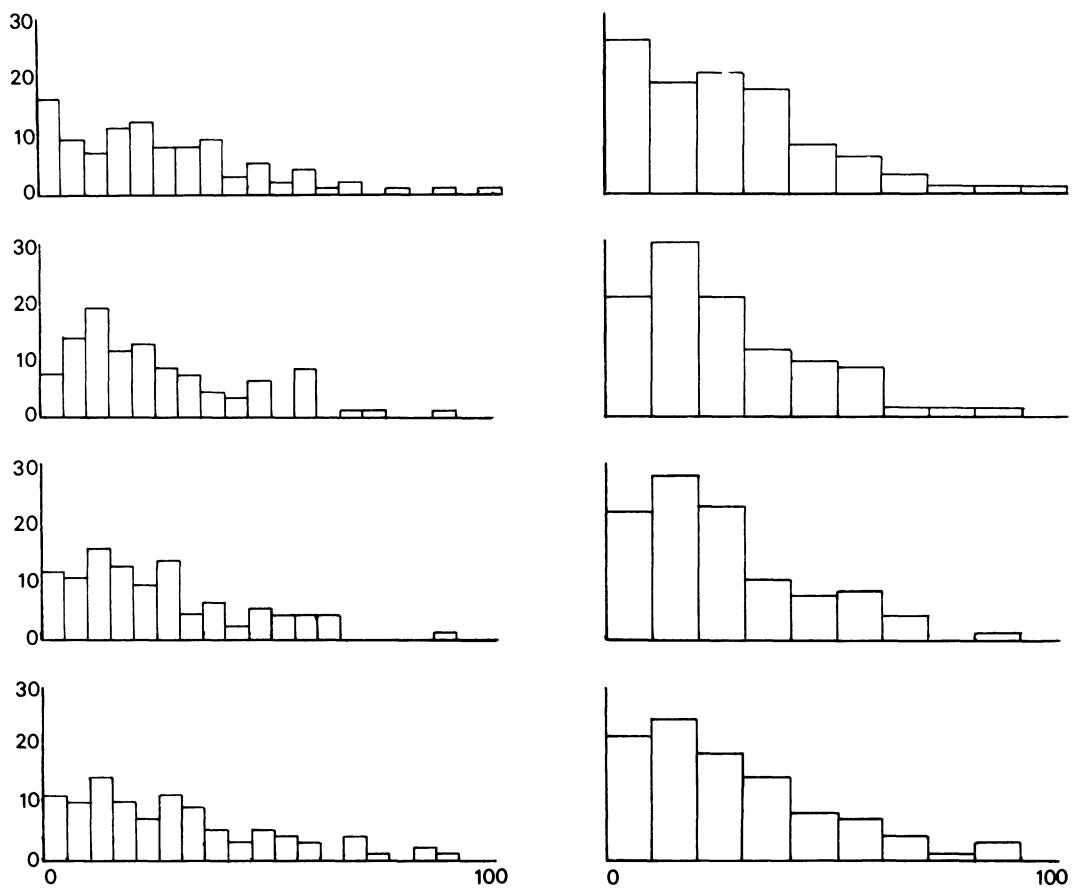


FIG. B2. Four sample histograms for samples of size 100 from an underlying half-normal detection function with $\sigma=33.3$ and $w=100$ m (essentially no truncation of the data). The left column shows the perpendicular distance data grouped into intervals of 5 m while the right column shows the same data grouped into intervals of 10 m. The data represent 4 data sets of the 25 used in the computer simulation studies with this detection function. The results are independent of the measurement unit; here, the units are arbitrarily denoted as meters.

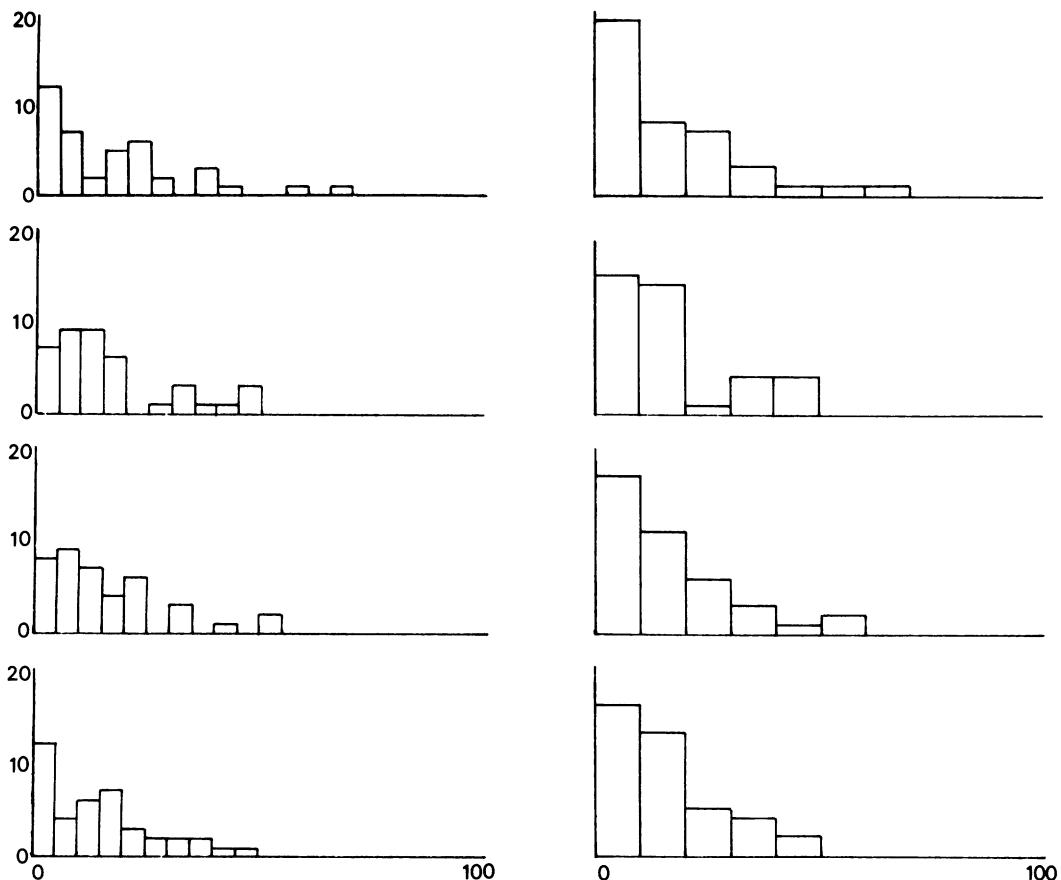


FIG. B3. Four sample histograms for samples of size 40 from an underlying negative exponential detection function with $a=0.065$ and $w=100$ m (essentially no truncation of data). The left column shows the perpendicular distance data grouped into intervals of 5 m while the right column shows the same data grouped into intervals of 10 m. The data represent 4 data sets of the 25 used in the computer simulation studies with this detection function. The results are independent of the measurement unit; here, the units are arbitrarily denoted as meters.

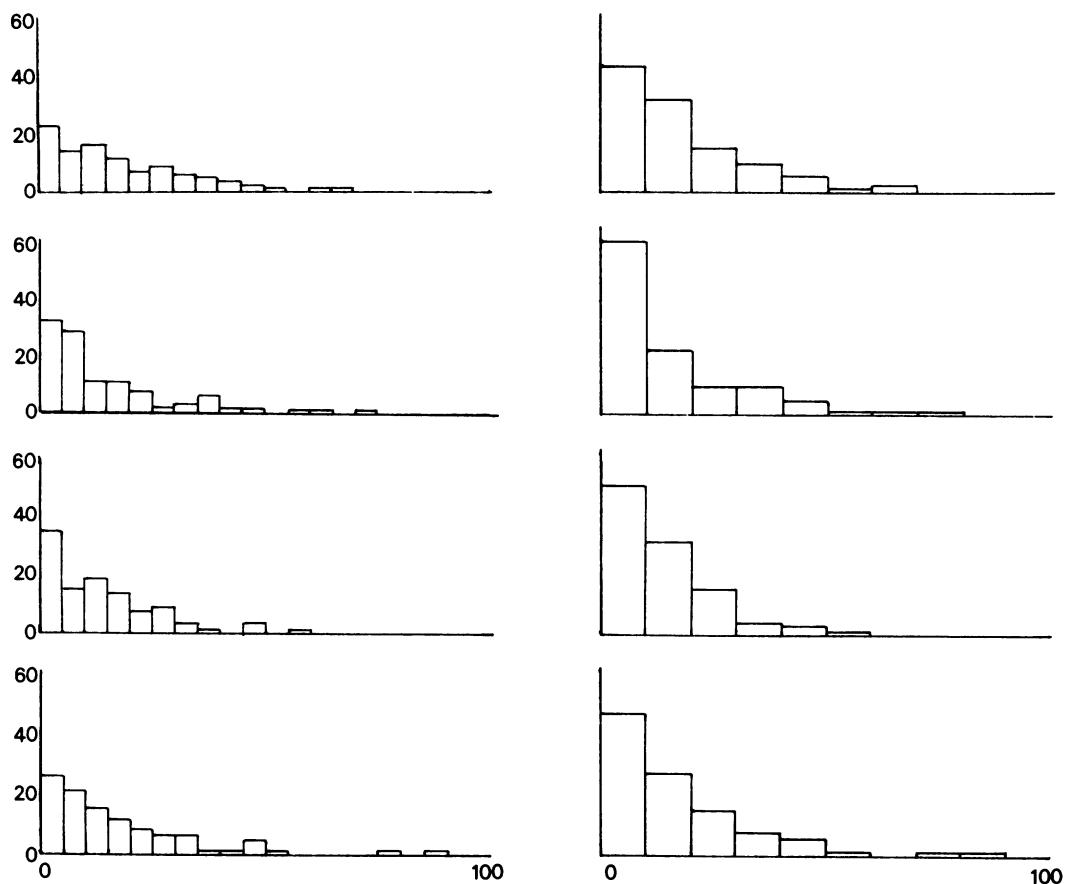


FIG. B4. Four sample histograms for samples of size 100 from an underlying negative exponential detection function with $a=0.065$ and $w=100$ m (essentially no truncation of the data). The left column shows the perpendicular distance data grouped into intervals of 5 m while the right column shows the same data grouped into intervals of 10 m. The data represent 4 data sets of the 25 used in the computer simulation studies with this detection function. The results are independent of the measurement unit; here, the units are arbitrarily denoted as meters. Unless the true underlying function is known, most estimators of $f(0)$ will perform better if some of the extreme outliers are deleted (truncated) during the analysis (e.g., lower graphs).

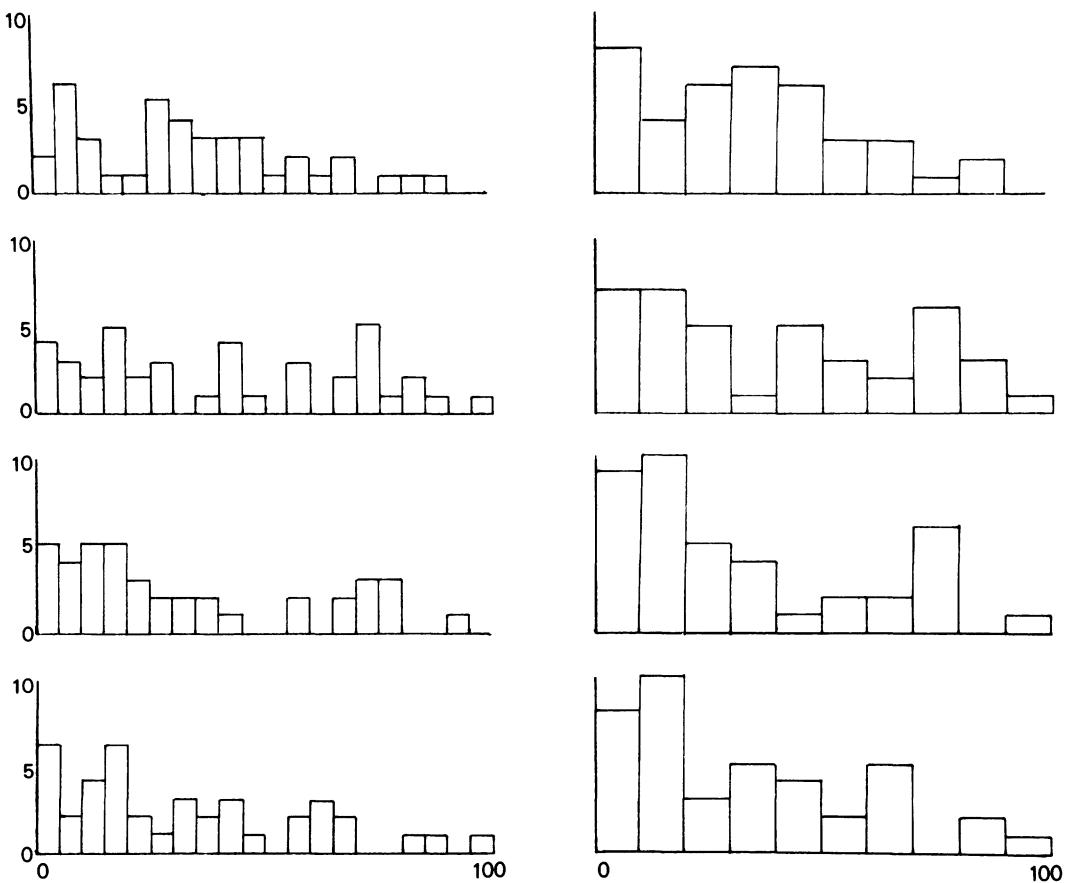


FIG. B5. Four sample histograms for samples of size 40 from an underlying modified beta detection function with $a=1.0$ (linear) and $w=100$ m. The left column shows the perpendicular distance data grouped into intervals of 5 m while the right column shows the same data grouped into intervals of 10 m. The data represent 4 data sets of the 25 used in the computer simulation studies with this detection function. The results are independent of the measurement unit; here, the units are arbitrarily denoted as meters.

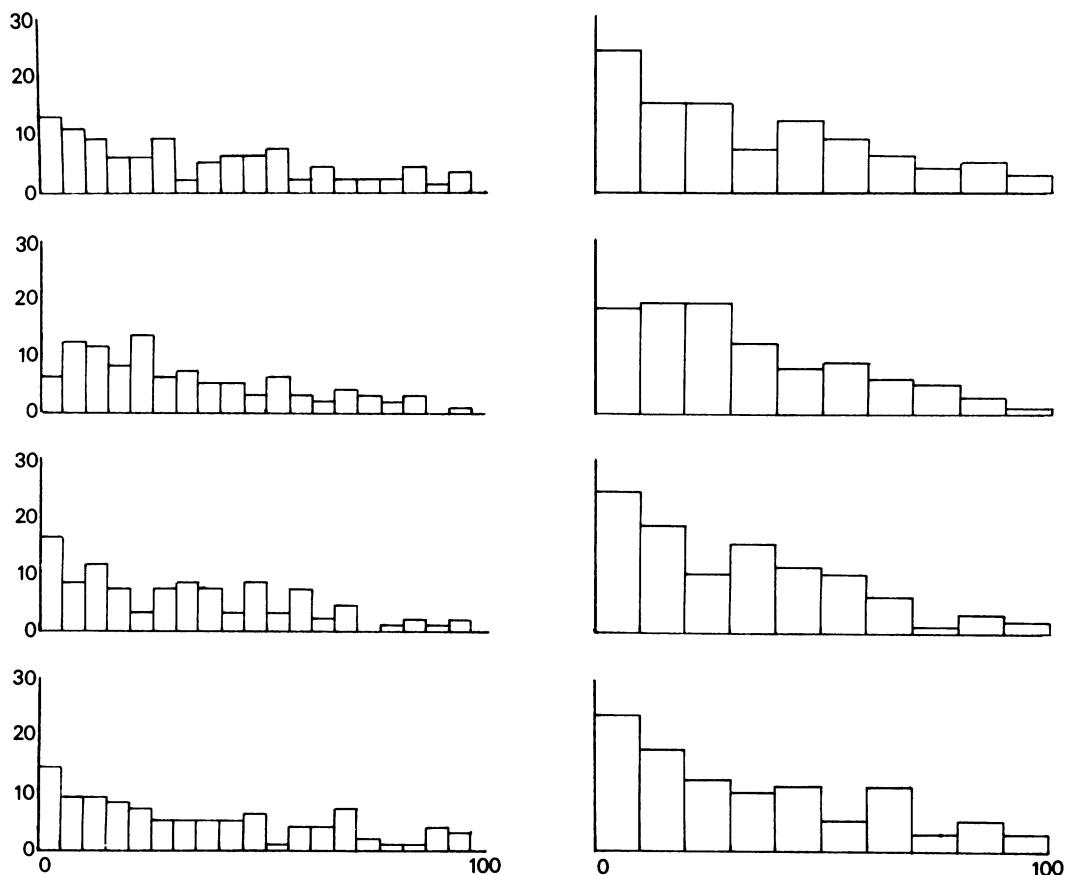


FIG. B6. Four sample histograms for samples of size 100 from an underlying modified beta detection function with $a=1.0$ (linear) and $w=100$ m. The left column shows the perpendicular distance data grouped into intervals of 5 m while the right column shows the same data grouped into intervals of 10 m. The data represent 4 data sets of the 25 used in the computer simulation studies with this detection function. The results are independent of the measurement unit; here, the units are arbitrarily denoted as meters.

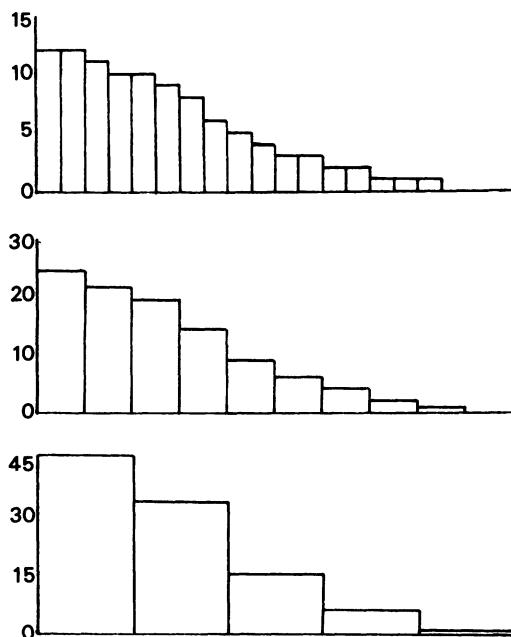


FIG. B7. Three histograms that show expected frequencies for a sample of size 100 from an underlying half-normal detection function with $\sigma=33.3$ and $w=100$. The data are the same in each of the 3 graphs, only the number of groups and the interval size vary. This demonstrates the illusion that as group size increases, the data look as though the underlying detection function is negative exponential.

oretical FS models fitted to several shapes of the probability density function, $f(x)$, for perpendicular distance models.

Construction of Figures, General Considerations

In examining and comparing different detection function models, it is their (relative) shapes that concern us. Thus, the scale at which those curves are drawn is not of importance, yet some scale must be chosen. Moreover, we cannot plot their full range, x in 0 to ∞ , so some finite w must be chosen for display purposes.

Similarly, the scale of the y axis (i.e., $y = g(x)$) is arbitrary, except that it ranges from 0 to 1. We had to choose some scale at which to plot the curves, and our choice was basically governed by the need to illustrate 2 plots per page. Having chosen a scale, we used exactly the same scale for all 10 figures. This is important because it allows valid comparison of the shapes of the functions graphed (either for $g(x)$ or $f(x)$).

Another consideration was which value to choose for $g(w)$. For some functions (e.g., in Fig. C1) we automatically have $g(w) = 0$, but for most detection functions $g(x)$ goes only to zero asymptotically as x goes to infinity. To achieve comparability of curve shapes, we decided to fix $g(w) = 0.02$ (where there was a choice). Thus, in Figs. C2–C7 we forced $g(w)$ through 0.02 for all curves plotted. In Figs. C8–C9, $f(w)$ is 2 percent of $f(0)$.

No loss of generality results from adopting those conventions; in fact, using $g(w) = 0.02$ renders the value of w arbitrary and guarantees that the shapes of the curves may be meaningfully compared within and between figures. Nonetheless, to actually plot curves, one must select a w value, and make other choices about curve parameter values. (Because those technical details are not relevant to the points we wish to make, they are not presented here; they may be requested from the authors.)

APPENDIX C

EXAMPLES OF DETECTION FUNCTION SHAPES AND THE FOURIER SERIES MODEL

To demonstrate detection function shapes, we have plotted some examples of that function. Figs. C1 and C2 show the detection functions used as a basis for the simulation results given in PART 3 (COMPUTER SIMULATION RESULTS). Figs. C3–C6 show the basic shapes of 4 parametric families of detection curves. Fig. C7 presents our opinion of the range in which the detection curve should fall in real line transect studies. Finally, Figs. C8–C10 show some 1-, 2-, and 3-term the-

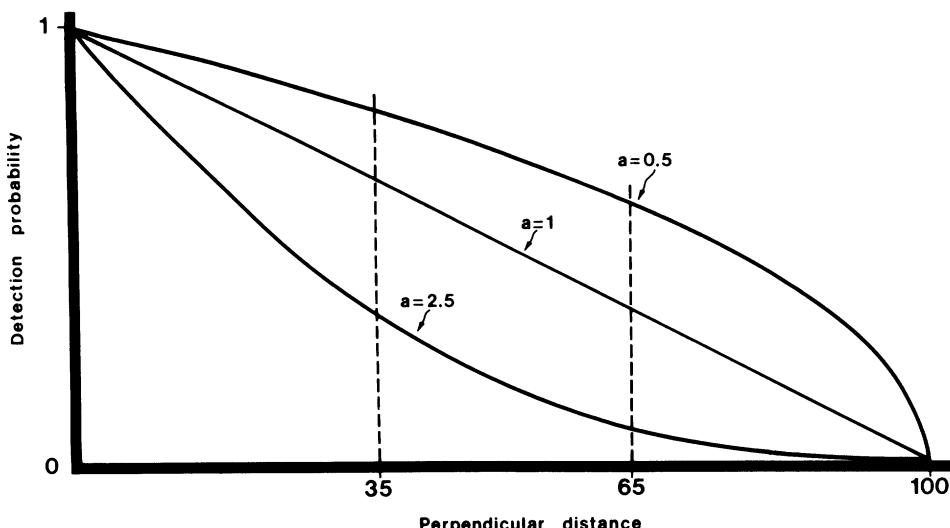


FIG. C1. The 3 cases of the modified beta detection function used in the simulation study.

The Simulation Study

Early in our research we did a simulation study to evaluate the performance of various estimators under varying shapes of the detection function. Models used were the modified beta,

$$g(x) = \left(1 - \frac{x}{w}\right)^{a+1} \quad 0 < x < w, \quad 0 < a,$$

the negative exponential (Gates et al. 1968)

$$g(x) = \exp(-ax) \quad 0 < x \quad 0 < a,$$

and the half-normal

$$g(x) = \exp(-ax^2) \quad 0 < x \quad 0 < a.$$

Fig. C1 shows the 3 basic shapes of the modified beta detection function. In that model, a is a shape parameter. For values of a greater than 2.5, $g(x)$ will lie below the lower curve in Fig. C1. We regard such shapes as undesirable for, and unrealistic with respect to, line transect detection functions. For a less than 0.5, the curves lie above the upper curve in Fig. C1. In all cases, $g(w) = 0$.

Both the negative exponential and half-normal detection functions have only one shape (the single parameter in those functions is a scale parameter, and does not influence the fundamental curve shape). In Fig. C2, which contrasts the shapes of the negative exponential (NE) and half-normal models (HN), we see that the NE model is “spiked” in shaped at the origin ($x = 0$), whereas the HN has a “shoulder” and satisfies the *shape criterion* $g'(0) = 0$.

The Exponential Power Series Model

Fig. C3 shows the range of curve shapes for the exponential power series detection function (Pollock 1978):

$$g(x) = \exp\left(-\left(\frac{x}{a}\right)^b\right) \quad 0 < x < \infty, \quad 0 < a, \quad 0 < b.$$

In that model, a is a scale parameter and the curve shapes are controlled by b . For $b = 1$, we have the NE model, plotted as curve 2 in Fig. C3 (that curve exactly matches the NE curve in Fig. C2). For $b = 2$, we have the HN model, plotted as curve 5 in Fig. C3 (that curve exactly

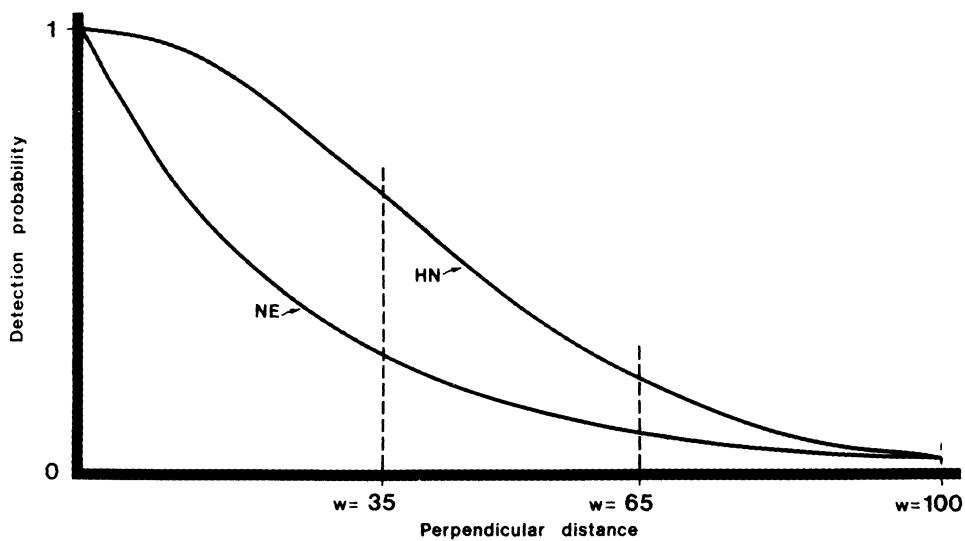


FIG. C2. The half-normal (HN) and negative exponential (NE) detection functions used in the simulation study. Three truncation points ($w=35$, 65 , and 100) are shown.

matches the HN curve in Fig. C2). In general, as the parameter b becomes smaller (less than 1), $g(x)$ becomes more and more spiked at the origin. As b becomes large, $g(x)$ tends to be flatter near $x = 0$. The values of b for curves 1 through 7 in Fig. C3 are, respectively, 0.5, 1.0, 1.333, 1.667, 2.0, 3.5, and 5.0.

That family of detection curves is one of the most flexible parametric models.

The Exponential Quadratic Model

Fig. C4 shows the range of detection function shapes for the exponential qua-

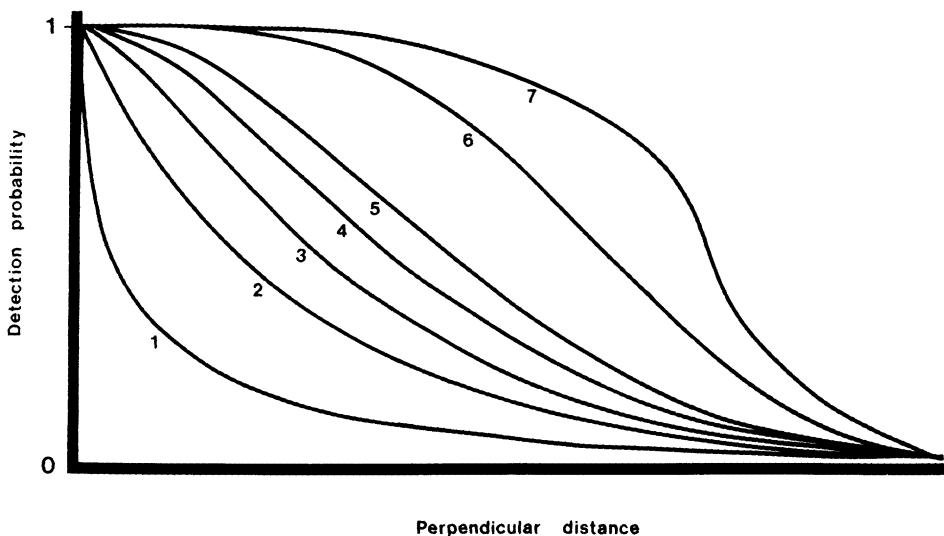


FIG. C3. Several members of the exponential power series detection function are shown to illustrate the various shapes the curves can have.

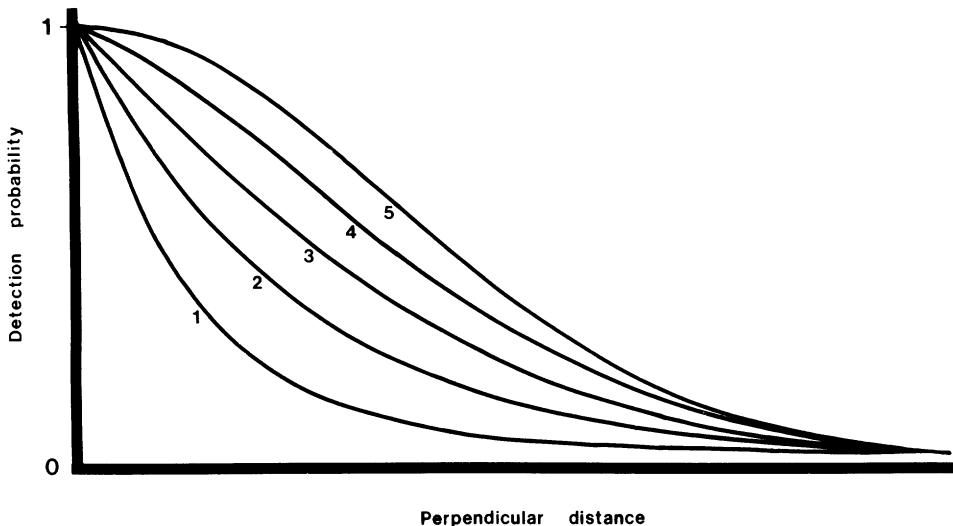


FIG. C4. Several members of the exponential quadratic detection function are shown to illustrate the various shapes the curves can have.

drastic model introduced in this monograph:

$$g(x) = \exp(-(ax + bx^2)) \quad 0 < x < w.$$

It is a valid detection function over the entire range 0 to ∞ if and only if both a and b are nonnegative. It can be used for

values of $a > 0$ and $b < 0$ for the range 0 to $a/|2b|$; however, those special cases are very spiked at the origin. Curve 1 in Fig. C4 is such a curve (b is negative). In all the other curves of Fig. C4, both a and b are nonnegative.

That family of curves includes as special cases the NE (Fig. C4, curve 2) and

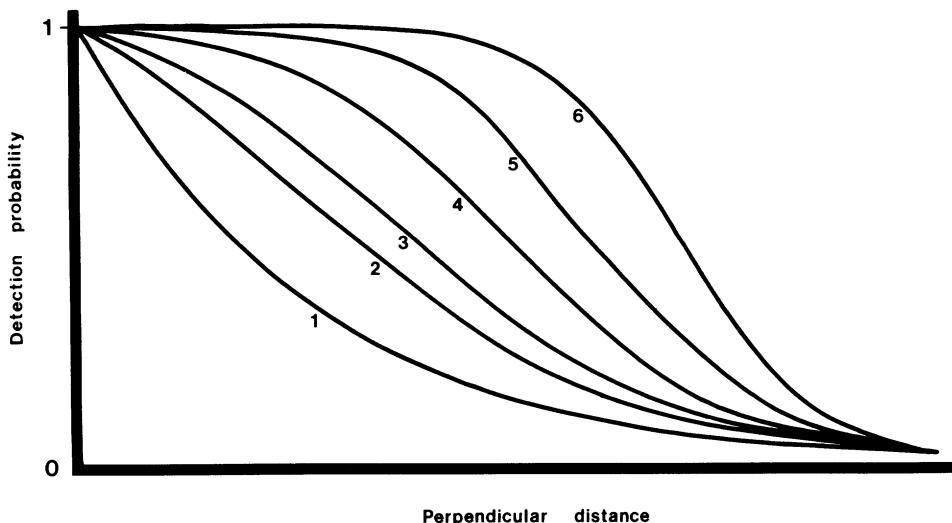


FIG. C5. Several members of the reverse logistic detection function are shown to illustrate the various shapes the curves can have.

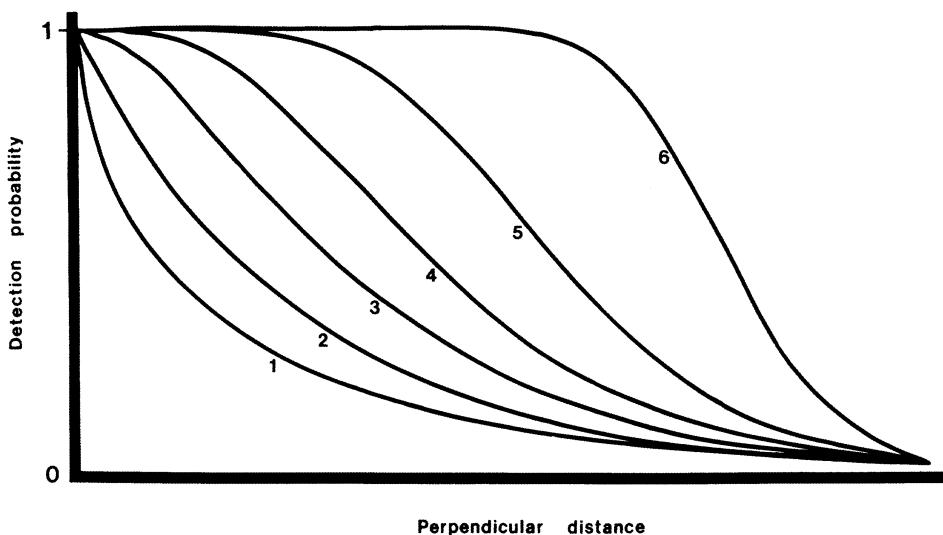


FIG. C6. Several members of the incomplete gamma detection function are shown to illustrate the various shapes the curves can have.

the HN (Fig. C4, curve 5). Those 2 forms are extreme cases of the valid detection function forms for the model when both a and b are nonnegative. Moreover, no valid detection functions of that form lie above curve shape 5 in Fig. C4. It is worth noting that the mixture model of the NE and HN detection functions all lie between curve forms 2 and 5 in Fig. C5.

The exponential quadratic model is less flexible than the exponential power series. That lesser flexibility, however, is not a detriment and may in fact lead to better properties of the exponential quadratic as a model for real data.

The Reversed Logistic Model

Fig. C5 shows the range of detection function shapes for the reversed logistic model (Eberhardt 1968, 1978a):

$$g(x) = \frac{(1+b)e^{-ax}}{1+be^{-ax}} \quad \begin{matrix} 0 < x < w \\ 0 < b \\ 0 < a \end{matrix}$$

In that model, a is a scale parameter, and b entirely determines the shape of the curves. In the limit as b goes to zero, we have $g(x)$ as the NE model (curve 1 in

Fig. C5 has $b = 0.1$ and is virtually indistinguishable from the NE model). No curves in the family fall below the NE curve shape; hence, there is a limit to the possible spiked nature of the curves. As b goes to infinity, $g(x)$ develops a more and more pronounced shoulder. The values of the shape parameter b for curves 1 to 6, respectively, in Fig. C5 are 0.1, 4, 10, 50, 500, and 10,000.

The Incomplete Gamma Model

Fig. C6 shows the range of detection function shapes for the incomplete gamma model:

$$g(x) = 1 - \int_0^x \frac{y^{b-1} \exp(-y/a)}{\Gamma(b)a^b} dy \quad \begin{matrix} 0 < x < \infty \\ 0 < a \\ 0 < b \end{matrix}$$

(see, e.g., Sen et al. 1974). Although the model is particularly complex mathematically, its range of curve shapes is virtually the same as that of the exponential power series or reversed logistic models.

In the incomplete gamma model, a is a scale parameter and b is a shape param-

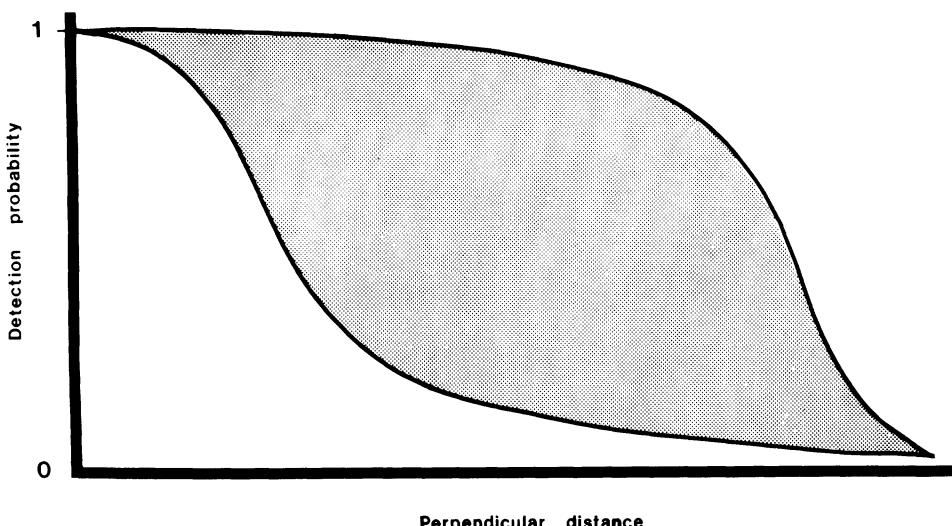


FIG. C7. The upper and lower curves are limits for what we consider to be the reasonable detection function shapes. We believe the true detection function should fall within the shaded area.

eter. For $b = 1$ we have the NE model (curve 2 of Fig. C6). For values of b less than 1, $g(x)$ is more spiked in shape, but as b gets larger than 1, $g(x)$ develops more of a shoulder. Curves 1 through 6 of Fig. C6 have values of b , respectively, 0.5, 1, 2, 4, 10, and 50.

A Comment

The generalized parametric models for $g(x)$ illustrated in Figs. C3–C6 all have similar curve forms (although the exponential quadratic model does not have the extreme shapes present in the other models). In particular, they cover, or tend to cover, the range of detection functions we consider realistic. Some of the models include forms we consider undesirable, if not in fact unrealistic ($g(x)$ shapes spiked at $x = 0$). A reasonable line transect study should have some distance in the immediate vicinity of the transect center line where the detection probability stays virtually 1. The alternative is a very rapidly falling detection probability, which leads to smaller sample sizes and more variable estimates of $f(0)$.

In Fig. C7, we presented our judgment

of the range of real world detection function shapes which either do, or should, apply given the scale to which the detection functions are plotted. We suggest that the true $g(x)$ should fall in the shaded area of Fig. C7. Only the lower curve is critical as a lower bound on $g(x)$. In particular, we rule out curves that are spiked at the origin. The upper curve may be realistic for "belt" or "strip" transects where w may be set so close to $x = 0$ (relative to the fall off of detection) that within the belt $g(x) = 1$. Yet if one did not have a finite w , $g(x)$ would have to eventually fall off to zero.

Some Fourier Series Models

Figs. C8–C10 show 3 probability density function curves, $f(x)$; note that they are not detection curves $g(x)$. Because $f(x)$ has the same shape as $g(x)$, and because the scale of the $y (= f(x))$ axis is not shown, the curves could just as well represent $g(x)$. However, we wish to illustrate some theoretical FS models, and for that purpose it is more logical to proceed in terms of $f(x)$ (the FS model is not intended to model $g(x)$). These 3 figures are

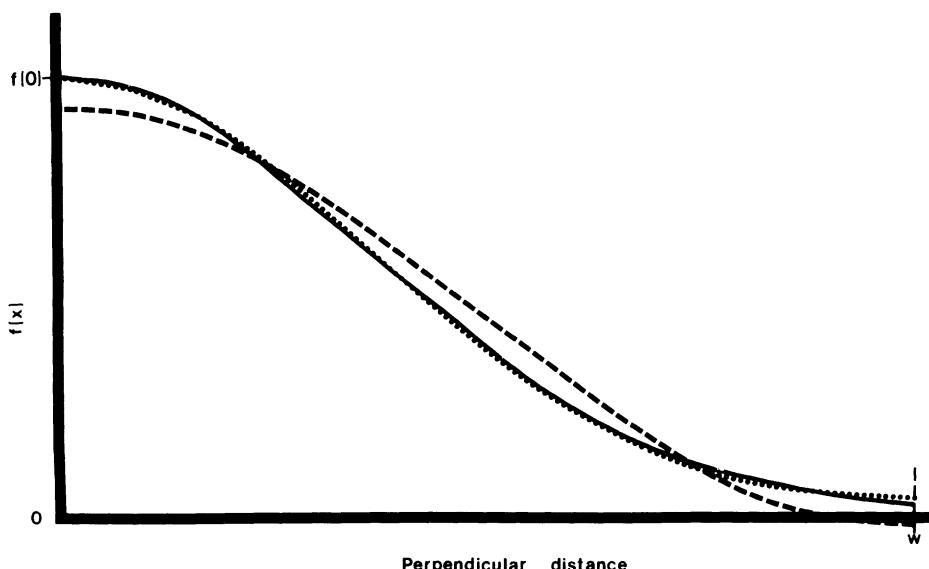


FIG. C8. The half-normal probability density function (—) and the theoretical best fitting Fourier series models with 1 term (----) and 2 terms (····).

to the same scale as the others; in essence we equated $f(0)$ to 1 and $f(w)$ to $0.02f(0)$.

In each figure, the true, underlying model for $f(x)$ is shown as a solid line. The theoretical optimal coefficients in the FS model are given by

$$a_j = \frac{2}{w} \int_0^w \cos\left(\frac{j\pi x}{w}\right) f(x) dx.$$

We wish to illustrate how the FS model with only a few terms (e.g., $m \leq 3$) can provide a very good approximation to a

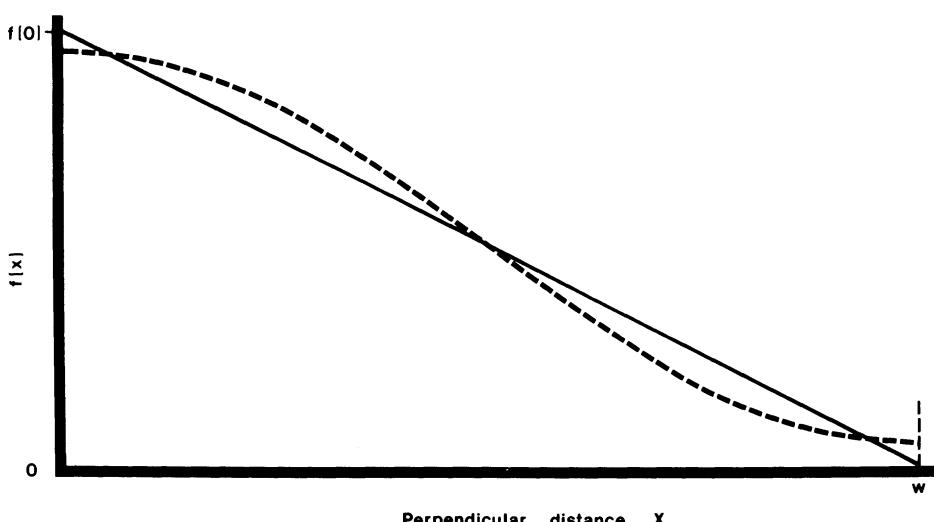


FIG. C9. The modified beta probability density function (—) for $a=1$ (i.e., the linear case) and the theoretical best fitting 3-term Fourier series model (----).

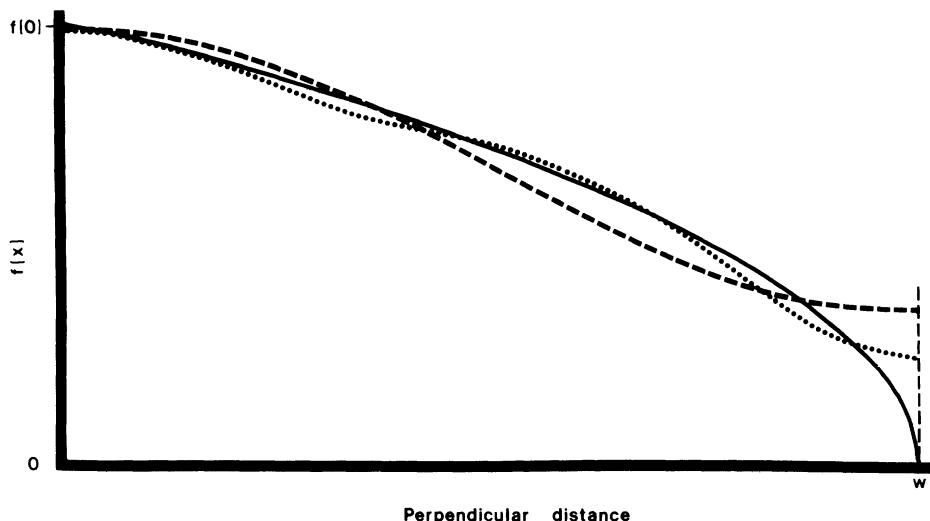


FIG. C10. The modified beta probability density function (—) for $a=0.5$ and the theoretical best fitting Fourier series models with 1 term (----) and 3 terms (····).

range of true, underlying models $f(x)$. Therefore, for a select number of terms (m) we plotted the theoretically best fitting FS model as

$$f_m(x) = \frac{1}{w} + \sum_{j=1}^m a_j \cos\left(\frac{j\pi x}{w}\right),$$

(shown as dashed and dotted lines in Figs. C8–C10). We emphasize that these “fitted” FS curves are theoretical, rather than from simulation results.

Fourier Series Fit to the Half-Normal Model

In Fig. C8, the true $f(x)$ is the HN pdf. The 1- and 2-term FS models (represented as ---- and ····, respectively) have been “fit” to this true pdf (that is, the true a_j have been computed using Eq. (C.1)). The 1-term FS model has negative values for x near w ; that causes no problems in estimating $f(0)$. The fit of the FS model depends only slightly on the value of w used. With the value of w used in Fig. C8 the PRB of the 1-term FS model is -7.7 percent. From Fig. C8, the 2-term FS model provides an excellent fit to $f(x)$ over the entire range of $0 \leq x \leq w$. Also,

the PRB of the FS estimator is only -0.5 percent for this 2-term model. If more terms were added to the FS model, the fit to the true $f(x)$ would become even better.

Fourier Series Fit to a Linearly Decreasing $f(0)$

In Fig. C9, the true $f(x)$ (the solid line) is the modified beta pdf with $a = 1$. One might not expect the FS model to fit that linear $f(x)$ very well for a small number of terms; but, it does quite well. The 3-term FS model (-----) for this true pdf has PRB of only -5 percent. The example illustrates that one can get good results in terms of estimating $f(0)$ with the FS model even if the FS model $f_m(x)$ does not closely match the true pdf over its entire range.

Fourier Series Fit to a Strictly Concave True pdf

The true pdf $f(x)$ in Fig. C10 is the modified beta, pdf with $a = 0.5$. The 1- and 3-term FS models have been fitted to this true pdf. The PRB of $\hat{f}(0)$ using those models is -1.2 percent in both cases; the

2 models are virtually identical at $x = 0$. However, it is clear from Fig. C10 that the 3-term FS model provides a better overall fit to the true pdf than does the 1-term model. Because our goal in the analysis of line transect data is to estimate $f(0)$, there is no advantage in this situation to using the 3-term model even though it is a better approximation to the true pdf over the whole range of $0 < x < w$. It is also worth noting that the 3-term FS model provides a very good approximation to the true pdf here except near $x = w$.

Comment

The important point about Figs. C8–C10 is that a FS model with only a few terms (1 to 3) provides a good fit to a variety of true pdf shapes (hence true detection functions) if $f(x)$ is not strongly spiked (the nearer $f'(0)$ is to zero the better the FS model performs). The results discussed, and those figures do not depend on the specific parametric models used to plot $f(x)$. It is the shapes of the true pdfs that govern how well the FS model can fit them.

APPENDIX D

COMMENTS ON ADDITIONAL LINE TRANSECT TOPICS AND RELATED SUBJECTS

Line transects have been a relatively unexplored methodology until recently, especially in terms of a well-developed, general theoretical foundation. Many areas of theory have not been discussed in this monograph; in part, this is for consideration of space. For example, the problem of animals that occur in clusters is an important one, and we have investigated it to some extent. More could be said about it than we intend, both because of space constraints and because we have seen only one set of line transect data with animals in clusters and for which cluster size had any demonstrable

effect on detection probability. Rather than deal with that topic (clustered animals) and others in detail in the body of the monograph we touch on them briefly here. This material is intended for people interested in further research topics in line transect theory, hence is mathematical. (Also, the notation used here is extended, and partly separate, from the rest of this monograph.)

Clustered Objects

In biological populations, it is quite common for individuals to be clustered (coveys of quail, herds of antelope, schools of porpoise). In such cases, it is necessary to distinguish the density of clusters from the density of individual animals. In the applications of line transect theory that we have presented, the clusters must be considered to be the objects of interest, and distances should be measured to the geometric center of the cluster. Estimation of the average cluster size, and of the density of individuals (as well as the density of cluster) is complicated only if detection probability is related to cluster size. We know of no substantial published results for the case when objects are clustered; however, Pollock and Burnham (pers. comm.) have worked on that problem and we present a few of their key results below.

All the basic results and definitions given throughout this monograph are applicable if the data are stratified by cluster size. However, to apply those results and extend them to the clustered case requires an elaboration of notation. We define the following:

S = the size of a randomly chosen cluster, $S \in \{1, 2, \dots\}$,

$D(s)$ = the density of clusters of size s ,

D_e = total density of clusters =

$$\sum_{s=1}^{\infty} D(s),$$

D_I = density of individuals =

$$\sum_{s=1}^{\infty} sD(s),$$

$g(x, s)$ = probability of detecting a cluster

of size s at perpendicular distance x ,

$f(x|s) = \text{pdf of observed perpendicular distances of clusters of size } s,$

$n(s) = \text{number of detected clusters of size } s,$

$n = \sum_{s=1}^{\infty} n(s) = \text{total sample size},$

S^* = size of a cluster detected by line transect sampling. The actual cluster size data will be denoted simply as s_1, \dots, s_n . The random variables S and S^* do not necessarily have the same distribution,

x_{sj} = observed perpendicular distances, by cluster sizes, $j = 1, \dots, n(s)$, $s = 1, 2, \dots$

The 4 assumptions given in PART 1 (*Assumptions*) apply here. In particular, we assume $g(0,s) = 1$ for all s . Also, it should be clear that

$$f(x|s) = \frac{g(x,s)}{\mu(s)} \quad 0 < x < w,$$

$$\mu(s) = \int_0^w g(x,s)dx,$$

and that w may be finite or infinite. Another basic result is $\Pr\{S = s\} = p(s) = D(s)/D_c$, therefore,

$$E(S) = \frac{\sum_{s=1}^{\infty} sD(s)}{D_c} = \frac{D_I}{D_c},$$

and hence, $D_I = E(S) D_c$.

Another important result is the probability distribution of observed cluster sizes, $p^*(s) = \Pr\{S^* = s\}$. This is not hard to derive once one realizes that $p^*(s) = E(n(s))/E(n)$. It is a basic result that $E(n(s)) = 2LD(s)/f(0|s)$ and of course $E(n) = 2LD_c/f(0)$, therefore

$$p^*(s) = \frac{D(s)}{D_c} \cdot \frac{f(0)}{f(0|s)} = p(s) \cdot \frac{f(0)}{f(0|s)}.$$

This result can be written in different ways. For example one can use

$$g(x) = \sum_{s=1}^{\infty} p(s)g(x,s) = \sum_{s=1}^{\infty} \frac{D(s)}{D_c} g(x,s),$$

hence,

$$(f(0))^{-1} = \sum_{s=1}^{\infty} \frac{D(s)}{D_c} \int_0^w g(x,s)dx$$

and finally,

$$p^*(s) = \frac{D(s) \int_0^w g(x,s)dx}{\sum_{s=1}^{\infty} D(s) \int_0^w g(x,s)dx}.$$

This verifies the heuristically evident result that S and S^* have the same distribution if, and only if, $g(x,s)$ is in fact independent of cluster size. Given that independence, the observed cluster sizes s_1, \dots, s_n are a simple random sample from the distribution of S thus $E(\bar{s}) = E(S)$ and $\hat{D}_I = \bar{s}\hat{D}_c$ (\hat{D}_c derives from the data “pooled” over cluster size).

If detection probability depends on cluster size, estimation of $E(S)$ is not so easy. A key result in this regard is derived from

$$p(s) = \frac{D(s)}{D_c} = \frac{E(n(s))f(0|s)}{\sum_{s=1}^{\infty} E(n(s))f(0|s)},$$

therefore,

$$E(S) = \frac{\sum_{s=1}^{\infty} E(n(s))f(0|s)s}{\sum_{s=1}^{\infty} E(n(s))f(0|s)},$$

which is equivalent to

$$E(S) = \frac{\sum_{i=1}^n E(f(0|s_i)s_i)}{\sum_{i=1}^n E(f(0|s_i))}.$$

An estimator of $E(S)$ is

$$\hat{E}(S) = \frac{\sum_{i=1}^n \hat{f}(0|s_i)s_i}{\sum_{i=1}^n \hat{f}(0|s_i)},$$

which can be used if one can model $g(x,s)$ (or $f(x|s)$). It is intuitively appealing to compute $\hat{E}(S)$ as a weighted average of the observed cluster sizes when detec-

tion depends upon cluster size. The above result shows it is in fact true and clarifies what the weights should be.

Those results demonstrate that the estimation of $E(S)$ requires modeling only $f(x|s)$. However, the joint distribution of perpendicular distance and cluster size is worth giving. It can be derived from results given above:

$$f(x,s) = \frac{g(x,s)p(s)}{\mu} = f(x|s)p^*(s),$$

where

$$\mu = \int_0^w \left(\sum_{s=1}^{\infty} g(x,s)p(s) \right) dx.$$

Note that the joint distribution depends upon the probability distribution of cluster size, a quantity which is not known. Consequently, it is important that inference can be made about D_c and $E(S)$ without having to know $p(s)$.

A great deal more can be and has been done with this clustered case. For our purposes here, we simply emphasize that in a real problem it is critical to test the null hypothesis of independence between cluster size and perpendicular distance data (based on such things as the correlation of s_i and x_i , and on analysis of variance type tests based on the distance data, x_{sj}). If operationally acceptable, the independence assumption makes inference relatively simple. If detection probability is judged to be dependent on cluster size, a weighted estimate of $E(S)$ is required; determination of the weights probably will be largely an empirical exercise.

Other Ancillary Data

Cluster size may be thought of as an ancillary variable that can be measured, especially if the density of clusters is the basic parameter of interest. Parameters related to the ancillary variable are, for example, average cluster size in the area sampled and the density of individuals. Many such types of information, other than distances and angles, can be record-

ed, but such information is ancillary to the basic distance data used to estimate density. For example, with individuals as the basic object of interest, one might record gender, age, and species. In those cases, one would presumably be interested in quantities such as sex ratios, age ratios, or species composition. With plants, some measure of size might also be taken.

In any large, complex line transect study, it is to be expected that some such ancillary data will be taken, with the objective of estimating the expected values of those variables for the entire population being sampled. Let z be such an ancillary variable. Let $E(z)$ be its average over the population. If the detection probability is independent of z , then, the sample z_1, \dots, z_n is a random sample from the entire population and \bar{z} is an unbiased estimator of $E(z)$. If detection probability depends upon z , estimation of $E(z)$ must take account of the nonrandom nature of the data. A theoretical framework can be developed for that general case, just as for the clustered case. It becomes more complex, however, if objects are clustered and the ancillary variable is a property of the cluster (such as species composition or sex ratio); the modeling problem in that case is quite complex.

Plotless Sampling

Two basic ways of estimating the density (numbers per unit area) of a biological population are either to count the number of objects in a sample of quadrats or to select a sample of points and at each point measure the distance to the k^{th} nearest object. In the first case, objects per area are sampled and, in the second, the area per objects is sampled. In line transects, the area per objects is sampled (i.e., it can be thought of as a plotless method). That is especially evident if $w = \infty$, but it is true even if w is finite because density cannot be estimated validly as $n/(2wL)$ if objects away from the center line

are missed. We describe here 2 plotless sampling methods related to line transects. Heuristically, they can be thought of as limiting cases where line length goes to zero.

Point Transects

Ramsey and Scott (1979) modified line transect theory so that data are taken only at points along the line, rather than being recorded continuously as one travels down the line. That modification of the theory was motivated by field work in Hawaii where it was impossible (or at least very unsafe) to walk and search for birds at the same time. Instead, data were recorded only at points systematically spaced along the transect. An observer would stand at each point for a fixed time and record the distances to all birds detected, without respect to angular orientation. With line transects, objects are detected only to the left or right of the line. For point transect sampling the distance from the observer to the object is the analogue of perpendicular distance and there is no analogue of sighting distance or angle. The basic line transect theory is easily adapted to this situation.

We define the following:

- k = the total number of sampling points (this quantity replaces line length),
- r = the distance to an observed object,
- w = the maximum distance to which observations are taken,
- $z = \pi r^2$ = the area sampled out to the observed object, and
- $g(r)$ = the probability of detecting an object given it is at distance r from the observer. We assume $g(0) = 1$.

It is convenient to develop results for finite w and then let $w \rightarrow \infty$ for the unbounded case; results can be pursued on the basis of distance, r . However, the results are not analogous to our results for line transects. Instead, one must realize that it is area which is being sampled, not simply linear distances. (Ramsey and

Scott 1979, developed their theory based on r , not z).

By assumption, objects are uniformly distributed in the area πw^2 . Let y represent the distance of an object from the sampling point. The cdf of y is $(\pi y^2)/(\pi w^2)$, or

$$H(y) = \left(\frac{y}{w} \right)^2 \quad 0 < y < w$$

and the pdf of y is $h(y) = 2y/w^2$. The unconditional probability of detecting an object is

$$P_w = \int_0^w g(y)h(y)dy = \int_0^w g(y) \frac{2y}{w^2} dy.$$

Transforming to area $z = \pi y^2$, and defining

$$\mu_w = \int_0^{\pi w^2} g(\sqrt{z/\pi}) dz,$$

we have $P_w = \mu_w/(\pi w^2)$. The pdf of r is derived by methods similar to those of Burnham and Anderson (1976),

$$f_r(r) = \frac{2rg(r)}{\int_0^w 2rg(r) dr}.$$

Note that $f_r(0) = 0$, but transforming to area, $z = \pi r^2$, gives the pdf of the area sampled as

$$f_z(z) = \frac{g(\sqrt{z/\pi})}{\mu_w},$$

and $f_z(0) = 1/\mu_w$.

The expected number of objects detected at the i^{th} sampling point is $E(n_i) = \pi w^2 D P_w = D \mu_w$. For k sampling points we have $n = n_1 + \dots + n_k$ and $E(n) = kD\mu_w$, and an estimator of D is based on $D = E(n)/(k\mu_w)$. This equation can be written as

$$D = \frac{E(n)f_z(0)}{k}.$$

This result is exactly analogous to $D = E(n)f(0)/(2L)$ in line transect sampling. If we work with the data as area, $z = \pi r^2$, all of our results on line transects are applicable to point transect sampling. In par-

ticular, we recommend using the FS estimator for such data. There is also an analogy between replicate transect lines and the k^{th} sampling points. Thus, the sampling variance of n can be estimated empirically.

Point to Plant Distance Sampling

One method of sampling plant communities is to select random points and measure the distance from each point to the k^{th} nearest plant of interest (often trees). Recently, S. A. Patil et al. (1979) presented a basis for nonparametric estimation of plant density when using data on the nearest plant.

We define the following:

r = the distance from a random point to the nearest plant,

$z = \pi r^2$ = the area sampled to the nearest plant,

$f(z)$ = the pdf of z , and

D = the average density of plants in the area sampled.

S. A. Patil et al. (1979) proved that under very mild conditions, $D = f(0)$. Thus, an estimator of $f(0)$ provides an estimate of plant density. As in the case of point transect sampling, if one develops results in terms of distance, r , no such result as $D = f(0)$ exists; rather the pdf of r at zero is zero. There are differences between line transect and point to plant sampling; in particular, in point to plant sampling we probably should assume $f'(0) < 0$, not $f'(0) = 0$ as in line transects. The reader is referred to S. A. Patil et al. (1979) for details, proofs, and an estimator for this situation.

Random Line Lengths and Interobservational Distances

All of our developments on line transects are based on a fixed (predetermined) line length, L . Then the number of objects detected, n , is a random variable. A reverse type of sampling is possible: fix n in advance and traverse a predetermined straight distance or set of

such "lines" until n objects are detected. In such a scheme, line length is random and sample size, n , is fixed. Seber (1979) has done some work on transects of random length, but not in the spirit of fixing n before sampling and letting L be random by virtue of sampling until n objects are detected.

Sampling with fixed n probably is not practical for line transects. That is especially clear in the case of aerial line transects. (Imagine telling a pilot the distance you want to fly is a random variable.) However, it would be inconvenient in any case. One aspect of that approach is appealing: one is led to consider the interobservational distances and whether anything can be done with them.

Let n be fixed and \bar{L} be the random line length. To get at the properties of \bar{L} we may consider it as the sum of n independent random variables. Let ℓ_1 be the distance from the start of the line to the point on the line perpendicular to the first detection. For $i = 2, \dots, n$, let ℓ_i be the distance along the line between the points perpendicular to the $(i - 1)^{\text{th}}$ and i^{th} detected object. Thus,

$$\bar{L} = \sum_{i=1}^n \ell_i.$$

The assumption of independence of detections suggests that we may treat the ℓ_1, \dots, ℓ_n as independent random variables (this point needs more work, however).

What information do those ℓ_i convey? That question is applicable even in the random n , fixed L case (where there is a segment of line beyond ℓ_n). The answer is that although they convey no information about object density, D , they do contain information about the spatial distribution of objects. Thus, the ℓ_i could be recorded and used to examine the hypothesis of a random distribution of objects over the sampled area. We present a few results on this idea, but investigation and use of interobservational distances is a virtually untapped subject.

The sampling distribution of objects in the area is

$\Pr\{N=i|A\}$ = the probability of i objects in a randomly selected plot of size A .

Let that probability be represented as $p(i|A)$. Let $F(\ell)$ represent the cdf of ℓ . To develop some theory, it is again appropriate to start with a finite width w . Then the area searched corresponding to ℓ is $2\ell w$. Conditional on there being N points in the area of size $2\ell w$, the number of points detected in the area by line transect sampling is a binomial $(N, p(\ell))$ random variable where $p(\ell) = 1/(wf(0))$.

Next, we write the conditional probability

$$\begin{aligned}\Pr\{\ell > \ell | N\} &= \Pr\{n = 0 | \ell = \ell\} \\ &= (1 - p(\ell))^N \\ &= \left(1 - \frac{1}{wf(0)}\right)^N.\end{aligned}$$

The unconditional value of $1 - F(\ell)$ is

$$\begin{aligned}1 - F(\ell) &= \sum_{i=0}^{\infty} \Pr\{\ell > \ell | N = i\} \\ &\quad \cdot \Pr\{N = i | A = 2\ell w\} \\ &= \sum_{i=0}^{\infty} \left(1 - \frac{1}{wf(0)}\right)^i p(i | 2\ell w).\end{aligned}$$

Without specifying the spatial distribution of objects (i.e., $p(i|A)$), further progress does not seem possible.

The result given above can be used as a starting point to derive a test of the hypothesis that objects have a given sampling distribution. For example, if a random distribution is assumed, then $p(i|A)$ is Poisson,

$$p(i|n) = \frac{e^{-DA}(DA)^i}{i!} \quad i = 0, 1, \dots$$

Under that assumption,

$$\begin{aligned}1 - F(\ell) &= \sum_{i=0}^{\infty} \left(1 - \frac{1}{wf(0)}\right)^i \frac{(2\ell w D)^i}{i!} e^{-2\ell w D} \\ &= \exp\left[\left(1 - \frac{1}{wf(0)}\right)2\ell w D\right] e^{-2\ell w D} \\ &= \exp\left(-\frac{2\ell D}{f(0)}\right).\end{aligned}$$

Thus, $\tilde{\ell}$ is a negative exponential random variable, with expected value $f(0)/(2D)$. It follows that $E(\tilde{L}) = nf(0)/(2D)$, or $D = nf(0)/(2E(\tilde{L}))$. None of those results are especially surprising. They illustrate that $f(0)$ and D are confounded in the sampling distribution of ℓ , and, therefore, that the interobservational distances provide no information about density D . But testing that those ℓ_1, \dots, ℓ_n have a negative exponential distribution does test whether the objects are randomly distributed in the area sampled (Epstein 1960 gave such tests).

Results for the assumption of objects having a negative binomial distribution are also readily obtained. Let

$$\Pr\{N = i | A\} = \binom{i + v - 1}{v - 1} (1 - p)^v p^i \quad i = 0, 1, 2, \dots$$

where $v > 1$ and $p \in [0, 1]$. From basic results, $E(N|A) = AD$, and from the negative binomial,

$$E(N|A) = v \frac{p}{(1 - p)},$$

hence, we must have

$$2\ell w D = v \frac{p}{(1 - p)}.$$

Applying the above formula for $1 - F(\ell)$ gives

$$1 - F(\ell) = \left(1 + \frac{p}{1 - p} \frac{1}{wf(0)}\right)^{-v},$$

thus,

$$1 - F(\ell) = \left(1 + \frac{2\ell D}{vf(0)}\right)^{-v}.$$

Again, D and $f(0)$ are confounded, so that the interobservational distances are useful only for testing whether the objects have a negative binomial distribution (for that purpose some test would have to be devised).

We recommend recording the interobservational distances if the spatial distribution of objects is to be investigated. We also recommend that a theory be devel-

oped to test for a variety of possible spatial distributions of objects.

Nonuniform Distribution of Objects About the Line

Under the assumptions of random line placement and no movement of objects in response to the observer, the population has a uniform (expected) distribution about the line. To clarify, let $D(x)$ be the density of objects at perpendicular distance x . If lines are run along roads, trails, or other physical features, rather than being placed at random, density is likely to be affected. Either $D(x)$ will be near zero for $x = 0$, and increase as x increases, or a concentration of objects could occur near the line and $D(x)$ would then be higher at $x = 0$ than it is further away from the transect center line.

The case of animal movement will also render $D(x)$ nonuniform (see ANALYSIS AND INFERENCE PROCEDURES FOR MOBILE POPULATIONS for some normalized plots of $D(x)$). If evasive movement occurs, we expect $D(x)$ to be low near $x = 0$ and increase to a maximum, then fall off again to D = the average density before movement. At least in the movement case, there is a meaningful parameter to estimate. In the case of nonrandom line placement, one may not even be able to define uniquely a density parameter of interest.

We develop here some basic results on this case of nonuniform distribution of objects about the line. To do so, we assume a finite strip width and a density of objects $D(x)$, $0 < x < w$, in that strip (results are for the left and right side of the line folded over). In the case of movement, $D(x)$ represents the animal density after movement. There is no need to have finite w to envision $D(x)$; in fact, we will have $D(x) \rightarrow D$ as $x \rightarrow \infty$ because there will be no effect on animals far enough from the line. We only need $w < \infty$ to normalize $D(x)$ to the pdf of perpendicular distances x , before detection,

$$h(x|w) = \frac{D(x)}{\int_0^w D(x)dx}.$$

With random line placement and no (evasive) movement, we have $h(x|w) = 1/w$. In the movement case, animals are displaced from only the immediate vicinity of the transect center line; hence, for large enough w , we will have

$$\int_0^w D(x)dx = wD.$$

However, in general we define the average density in $[0, w]$ as

$$\bar{D}_w = \frac{1}{w} \int_0^w D(x)dx.$$

We will have $\bar{D}_w = D$ (the parameter of interest) if w is large enough. For small w , $\bar{D}_w < D$.

The unconditional probability of detecting an object is

$$P_w = \int_0^w g(x)h(x|w)dx.$$

The expected number of objects in the area searched is $N_w = (2Lw)\bar{D}_w$. Define

$$\mu_w = w \int_0^w g(x)h(x|w)dx.$$

Note that μ_w has a finite limit as $w \rightarrow \infty$. Next, $E(n) = N_w P_w$, or

$$E(n) = 2Lw\bar{D}_w P_w = 2L\bar{D}_w \mu_w, \quad (D.1)$$

thus,

$$\bar{D}_w = \frac{E(n)}{2L\mu_w},$$

and for large enough w (certainly as $w \rightarrow \infty$), this gives

$$D = \frac{E(n)}{2L\mu_w}.$$

Define $f_m(0) = 1/\mu_w$, and we can write, as a formal equation,

$$\bar{D}_w = \frac{E(n)f_m(0)}{2L}, \quad (D.2)$$

where the subscript m denotes move-

ment. As explained below, this $f_m(0)$ is not the density function at zero of the perpendicular distances in the case of movement. We regard Eq. (D.2) as the appropriate basis for estimating density in the movement case. It shows that if w is large enough, $\bar{D}_w = D$ is the parameter of interest. Thus, in the case of movement, we want to find a good estimator of

$$f_m(0) = \frac{1}{w \int_0^w g(x)h(x|w)dx} = \frac{1}{\mu_w}. \quad (\text{D.3})$$

Estimation of $f_m(0)$ was investigated under ANALYSIS AND INFERENCE PROCEDURES FOR MOBILE POPULATIONS.

The theory in the nonrandom case can be developed in another way. The pdf of perpendicular distance data under a non-uniform distribution of objects about the line is

$$f(x) = \frac{g(x)h(x|w)}{\int_0^w g(x)h(x|w)dx} = \frac{wg(x)h(x|w)}{\mu_w}. \quad (\text{D.4})$$

Greater insight into Eq. (D.4) is obtained if we express $f(x)$ in terms of $g(x)$ and $D(x)$

$$f(x) = \frac{g(x)D(x)}{\int_0^w g(x)D(x)dx}. \quad (\text{D.5})$$

Starting with Eq. (D.1) and reducing it to the simplest possible form in terms of $g(x)$ and $D(x)$, we have

$$E(n) = 2L \int_0^w g(x)D(x)dx. \quad (\text{D.6})$$

Combining this with Eq. (D.5), we can derive a key result

$$D(x)g(x) = \frac{E(n)f(x)}{2L}. \quad (\text{D.7})$$

The right-hand side of Eq. (D.7) can be estimated, hence the left-hand side is estimable, but only as a product $D(x)g(x)$.

By assumption $g(0) = 1$, hence, a special case of Eq. (D.7) is

$$D(0) = \frac{E(n)f(0)}{2L}. \quad (\text{D.8})$$

If there is no movement and transects are randomly placed, then $D(0) \equiv D$ and Eq. (D.8) reduces to the result of Burnham and Anderson (1976).

When there is movement, we are likely to have $D(0) = 0$, in which case $f(0) = 0$. However, if $D(0) \neq 0$, Eq. (D.8) can be trivially rearranged as follows:

$$D = \frac{E(n)}{2L} \frac{f(0)}{(D(0)/D)}. \quad (\text{D.9})$$

This is a result derived (in a nonintuitive way) by Smith (1979). Smith's notation is $f(0) = H'(0)$ and $M'(0) = (D(0)/D)$. We see 2 problems with that: first $D(0)/D$ will not be known, and second, it is very difficult to estimate $f(0)$ in that movement case; it is actually easier to estimate what we called $f_m(0)$ of Eq. (D.3) (which is equal to $f(0)/(D(0)/D)$ if $D(0) \neq 0$). In particular, if $D(0) = 0$, $f(0) = 0$, and Eq. (D.9) makes no sense, but $f_m(0)$ of Eq. (D.3) is still meaningful as is Eq. (D.2).

Modeling Animal Density over the Area Sampled

The methods we recommend do not require objects to be distributed homogeneously over the area of interest. Variations in density and even distinct gradients can occur (e.g., Fig. 7). Given a proper sampling design, it is average density over the area that is estimated. We have made no attempt to model the density of objects in space, nor to incorporate it in a line transect model. Such models can be developed, in particular Schweder (unpublished dissertation) has provided a general framework for line transects wherein the distribution of animals is modeled as a first order stationary stochastic process. That allows movement independent of the observer. Schweder (1977) then derived a point

process model for line transect sampling; under his model, there are neither gradients (pronounced, systematic variations in object density over the area), nor other identified variations in density.

Schweder's model is quite general; however, when estimation of density is considered, the problem reduces to exactly what we have considered in this monograph, namely, an estimate of $f(0)$ based on the perpendicular distances x_1, \dots, x_n . Schweder explicitly recognized that, and stated that "... it is not important that the model fits well at the right tail, it is the density at zero we want to estimate!" (Schweder 1977:235).

No real advantage is apparent to modeling the distribution of objects in space unless gradients are introduced and an effort is made to estimate those gradients. Burdick (pers. comm.) has investigated the problem with special reference to porpoise populations in the eastern tropical Pacific Ocean. Burdick has meshed the idea of a nonhomogeneous spatial distribution of objects and estimation of that spatial distribution with the line transect sampling scheme. In particular, it is interesting that Burdick showed that the usual line transect estimator really estimates the average density of objects along the total transect path. This justifies the heuristically apparent need for a sampling design such that the transect lines cover the area representatively, thereby enabling valid extrapolation of the density estimate from the transect data to the entire area of interest.

We expect there will be more research to incorporate the spatial distribution of objects into line transect sampling. However, it is not necessary to estimate average density over the area. Rather, it is necessary to estimate density gradients. Readers interested in this subject are referred to Schweder (1977) and Burdick (in prep.).

Pooling Robustness and Area Stratification

We have not emphasized stratification as a study design because it is often not

practical. In particular, after stratification there will often be too few data points per stratum to compute separate estimates by strata. However, if sample size is not a problem, stratification is a good procedure for improving the precision of the overall estimate of density. Consideration of spatial stratification and the criterion of the *pooling robustness* strongly suggest that sampling effort over the area of interest should be uniform if stratification is not possible.

In principle, it may be possible to stratify the total area, A , into k subareas A_1, \dots, A_k such that the density varies by subarea and is fairly homogeneous within subareas. If it were possible to identify such strata and to sample each intensively enough to estimate density by strata, one would then have separate estimates $\hat{D}_1, \dots, \hat{D}_k$. The (true) average density over the whole area would be

$$\bar{D} = \frac{\sum_{i=1}^k A_i D_i}{A},$$

where $A = \sum_{i=1}^k A_i$; thus, one would want to pool the estimates \hat{D}_i as

$$\hat{D} = \frac{\sum_{i=1}^k A_i \hat{D}_i}{A}.$$

Let the total line length in the i^{th} stratum be ℓ_i , $i = 1, \dots, k$ and let $L = \sum_{i=1}^k \ell_i$. If strata were known, one would not weigh by line length, as this would give

$$\bar{D} = \frac{\sum_{i=1}^k \ell_i \hat{D}_i}{L},$$

and the 2 estimators, \hat{D} and \bar{D} , are not the same *unless* the following condition holds,

$$\frac{\ell_i}{L} = \frac{A_i}{A}. \quad (\text{D.10})$$

If information is insufficient to allow stratification by subareas (which is often

the case), then, the study design should have the expected value of the ratio ℓ_i/L equal to A_i/A . The ratios A_i/A (and the locations of the areas A_i) are not known however, and the only way to achieve that equality is by the use of a proper study design. Appropriate designs would be random, restricted random, or systematic placement of replicate lines. A systematic design would be best to achieve $(\ell_i/L) = (A_i/A)$.

The use of a *pooling robust* estimator (i.e., $n\hat{f}(0) = \sum_{i=1}^k n_i\hat{f}_i(0)$) in conjunction with condition (D.10) will give a valid estimated average density, \hat{D} , over the whole area, even if the density varies over this area. This is easily shown:

$$\begin{aligned}\hat{D} &= \frac{n\hat{f}(0)}{2L} \\ &= \frac{\sum_{i=1}^k n_i\hat{f}_i(0)}{2L} \\ &= \sum_{i=1}^k \frac{\ell_i}{L} \hat{D}_i \\ &= \frac{\sum_{i=1}^k A_i \hat{D}_i}{A}.\end{aligned}$$

A biased estimator of D could arise if stratification was a priori and lines were assigned to the strata in such a way that $(\ell_i/L) = (A_i/A)$ was seriously violated, and then the data were too sparse to allow separate estimates by strata. The data would then have to be pooled, but that would give the same result as computing separate estimates by strata and weighting them by (ℓ_i/L) , which are the wrong weights, and would therefore lead to a biased estimator of D .

We emphasize that there are 2 points here: stratification is a good procedure but it will rarely be practical. Without stratification, a reliable estimator of D (average density over the entire area) requires both a *pooling robust* estimator and a design that gives uniform coverage of the entire area.

Finite versus Infinite Population Sampling Theory Approaches

Our entire approach to modeling line transect sampling is based on the concept of the distance and angle data being random variables in the infinite population sense. Thus, we regard x , r , and θ as arising from a conceptually infinite population of responses characterized by their underlying pdfs. An alternative approach is to assume finite population sampling theory (see e.g., Overton and Davis 1969). That alternative apparently derives from the idea that only a finite number of objects exist in the area to be sampled. We do not find this sufficient argument to adopt the finite population approach.

The finite population approach has previously been applied for modeling and estimation based on sighting distance data. In particular, Hayne's estimator can be derived by using that approach. The approach can also be applied in the case of the elliptic model for sighting distances and angles.

We will comment upon those 2 approaches in the context that they provide competing philosophies to develop the elliptic flushing model. Two bases for detection have been recognized in the literature (see Eberhardt 1968, 1978a; Hayes unpublished thesis): fixed flushing radius models and variable detection distance models. Hayes elaborated on those 2 conceptual models, showed their proper relationship, and stated (Hayes unpublished thesis:8), with reference to Hayne's method, "... it is possible to regard the fixed flushing radius model as a special case of the probability model ..." We reject the fixed flushing radius model (finite population sampling approach) for several reasons:

- (1) The finite population approach (i.e., flushing models) implies that animals cannot flush behind the observer, yet this commonly occurs in the field.
- (2) The finite population approach assumes that sampling probabilities are known; this is not true in line transect sampling.

(3) Conceptually, the finite population approach assumes there is a fixed distance at which an animal will flush, hence be seen, rather than allowing that flushing distance to be influenced by continuously generated cues of the approaching observer (including rate and direction of approach).

(4) The finite population approach does not provide a unified theoretical basis for line transects. Such a unified approach is provided, however, by treating r and θ as random variables. For example, the fixed flushing radius model (finite population

approach) is applicable only when detection is due to an active response (flushing) by the animals sought, and is therefore not applicable to inanimate objects. The variable detection distance approach, however, is applicable whatever the reason for detection (e.g., either because the animal flushed, or because you spotted the animal before it was aware of you). This is especially important because real data, even for animals that will flush, represents a mix of those active and passive detection modes.