AN ABSTRACT OF THE THESIS OF

Valerie Jean Wildman for the degree of Doctor of Philosophy in Statistics presented on May 2, 1983.

Title: A New Estimator of Effective Area Surveyed in Wildlife Studies

Abstract approved: Redacted for privacy

Fred L. Ramsey

A new estimator of effective area surveyed is proposed for use in estimating population sizes or densities. The estimator is based on the cumulative distribution function for the observed detection areas and can be used with the data arising from either a line transect or a variable circular plot survey. Some sampling properties of the estimator are investigated and computer simulation results of comparisons with two previously proposed estimators are included. In addition, a proof of the consistency of the new estimator is given.

A NEW ESTIMATOR OF EFFECTIVE AREA SURVEYED IN WILDLIFE STUDIES

by

Valerie Jean Wildman

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed May 2, 1983

Commencement June 1983

APPROVED:

Redacted for privacy

Associate Professor of Statistics in charge of major

Redacted for privacy

Professor of Statistics in charge of first minor

Redacted for privacy

Associate Professor of Statistics in charge of second minor

Redacted for privacy

Chairman of department of Statistics

Redacted for privacy

Dean of Graduate School

Date thesis is presented May 2, 1983

Typed by researcher for Valerie J. Wildman

ACKNOWLEDGEMENTS

I would like to acknowledge the helpful suggestions made by Jeffrey L. Arthur, David S. Birkes, Justus F. Seely and Susan Stafford.

I would also like to acknowledge J. Michael Scott and the U. S. Fish and Wildlife Service for financial support during the course of this research.

And most of all, I want to thank James C. Reading at the University of Utah for giving me the encouragement to begin, my husband and major professor Fred L. Ramsey for giving me the encouragement to finish, and my father George W. Wildman for giving me encouragement all the way through.

TABLE OF CONTENTS

LIST OF FIGURES

Figure | Page

LIST OF TABLES

A NEW ESTIMATOR OF EFFECTIVE AREA SURVEYED IN WILDLIFE STUDIES

I. INTRODUCTION

A. Review of Literature and Goals of the Thesis

The Endangered Species Act and similar legislation have emphasized the need for accurate estimates of the population sizes and habitat requirements of various species of plants and animals. Although many techniques for obtaining such estimates have been proposed, they often have little theoretical basis or are designed for a particular species. Seber (1973) and Burnham and Anderson (1976) developed a general statistical model for the estimation problem which provided a basis for the development of nonparametric inference procedures as well as for the parametric approach taken by Ramsey (1979). A different field procedure proposed by Reynolds, Scott and Nussbaum (1980) led to estimation procedures based on a similar model by Ramsey and Scott (1979). The development of our new estimator was prompted by a desire for an estimator with a more intuitive basis than those already in use but with similar statistical properties. Thus, the goals of this thesis were established as:

(1) Describe the new estimator and examine its theoretical basis.

(2) Develop a proof of the consistency of the new estimator.

(3) Using computer simulations, compare the sampling properties of the new estimator to those of the Fourier Series

estimator proposed by Burnham and Anderson (1980) and a maximum likelihood estimator based on the exponential power model considered by Pollock (1978) and Ramsey (1979).

(4) Investigate the sensitivity of each of the three estimators to violations of various assumptions.

(5) Suggest possible modifications of the new estimator to use when we have a "donut", or nonperfect detectability at the origin.

B. Estimating Animal Abundance - the Problem in the Field

To formalize the statistical problem in estimating animal abundance, we assume that there is a known region R inhabited by some species of animal for which an estimate of population size or density is desired. This density D is defined as the ratio of the number of animals in the region, N, to the area of the region. Unless R is quite small or the animals are extremely cooperative, it is generally not possible to perform a complete census of the region, locating and counting every animal to determine N. Instead, a sample S of the region must be surveyed and the resulting information used to estimate the density D. This estimator will usually take the form $\hat{D} = n/A$, where n is the number of animals found in the subregion and A is the amount of area surveyed. If S is quite representative of R, the estimator $\hat{D}$ should provide a reasonable estimate of D.

One possible was to proceed in conducting a survey of the region R is to conduct a fixed area survey. In a survey of this type, the subregion S is clearly defined and marked off by flagging, fencing or other means. All animals in S are then located and counted. The number of animals found, n, can then be divided by the area of S to give $\hat{D}$. This method, although intuitively appealing, can be very expensive and time-consuming, allowing the surveying of only very small regions. For this reason, it is not adequate when the animals are rare or scattered, or inhabit a region such as the ocean which is not conducive to the marking of boundaries.

Another common approach to the problem is to conduct a variable area survey, in which an observer goes thorugh or over the region R and records every animal detected, rather than just those in some predefined subregion. Although some animals will be missed, the number found can be used as the numerator of $\hat{D}$. All that is then needed is some measure of the amount of area occupied by the animals that were detected, i.e. the amount of area surveyed. This area measurement is obtained by recording, for each animal detected, its distance from the observer at the time of detection. The density estimation problem in this situation then becomes one of using these distances to estimate the amount of area surveyed, which is then used for the denominator of $\hat{D}$. It is the collection of these detection distances that distinguishes the variable area survey methods from those in which no measure of occupied area is obtained, as well as from the fixed area survey methods.

C, Types of Surveys Commonly Used

Of the variable area survey methods, the line transect method

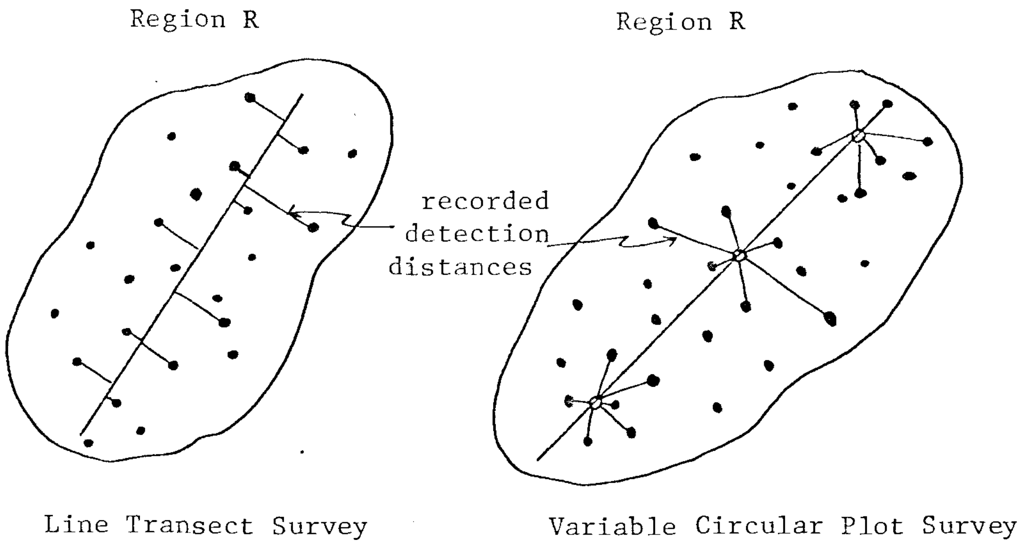


Figure 1. Detection distances recorded for line transect and variable circular plot surveys.

is probably the oldest and most commonly used. It is usually simple and inexpensive, and useful for surveying a wide variety of plant and animal species. In a survey of this type, one or more transects are placed at random in the region R and are then traversed by one or more observers using any appropriate means of locomotion. For a given transect of length L, the detection distance recorded for each animal is given by the perpendicular distance from the animal to the transect at the time of its detection. This is illustrated in Figure 1. The lengths and numbers of transect to be used, and their placement, will not be considered here; see for example Burnham and Anderson (1980).

Another variable area survey method was proposed by Reynolds et al. (1980) and has been shown to be effective in situations where the simultaneous recording of animals and traversing of the transect would be difficult or dangerous. This method, called the variable circular plot method, also uses randomly placed transects in R, but the actual recording of the detected animals is done only at fixed stations located at regular intervals along the transect. The observer arrives at a station, records animals for a predetermined length of time and then moves on to the next station. The detection distance which is recorded with this type of survey is the straight line distance from the observer to the detected animal; this is illustrated in Figure 1 also.

Although line transect and variable circular plot methods are the most widely used, additional methods have occasionally been suggested in which the detection areas (to be defined in Section D) have some form other than that of a rectangle. An example is given by Patil, Taillie and Wigley (1979) in which the detection areas are formed by trapezoids. As long as these areas can be viewed as a random sample of possible detection areas, the methods to be discussed here are applicable.

D. A Model for the Data from Variable Area Surveys

To analyze the data from variable area surveys, the general statistical model developed by Seber (1973), Burnham and Anderson (1976) and Ramsey and Scott (1979) is based on the following set of assumptions about the placement and movement of animals during the survey:

(1) the number of animals N in the region R is a random variable with $E(N) = D\mathcal{A}$, where $\mathcal{A}$ is the known area of the region R

and D is the unknown density.

(2) Conditional on N, the animals are distributed uniformly over the region R independently of D.

(3) The locations of different animals are determined independently.

(4) Animals occupy fixed positions during the survey period, and no animal is counted more than once.

(5) Detections of different animals are independent events.

(6) Distances to detected animals are determined without error.

Several authors have noted that these assumptions are not necessarily independent of each other, and that random placement of the transects within the region can eliminate the need for assumption number 2.

To get a model for the data, it has been customary to use a detectability function $g(y)$, defined to be the condition probability of detection given that an animal is at a distance y from the observer. This approach leads to different forms for the unconditional probability of detection in the line transect survey, which involves the distance y, and the variable circular plot survey, which involves y^2. This problem can be eliminated and the results applied to these as well as other types of variable area surveys if we instead let $g(a)$ be the conditional probability of detection for an animal whose detection area relative to the observer is a. For a line transect survey, this area is given by $a = 2Ly$, where y is the perpendicular distance from the animal to the transect and L is the length of the transect. For a variable circular plot survey, this

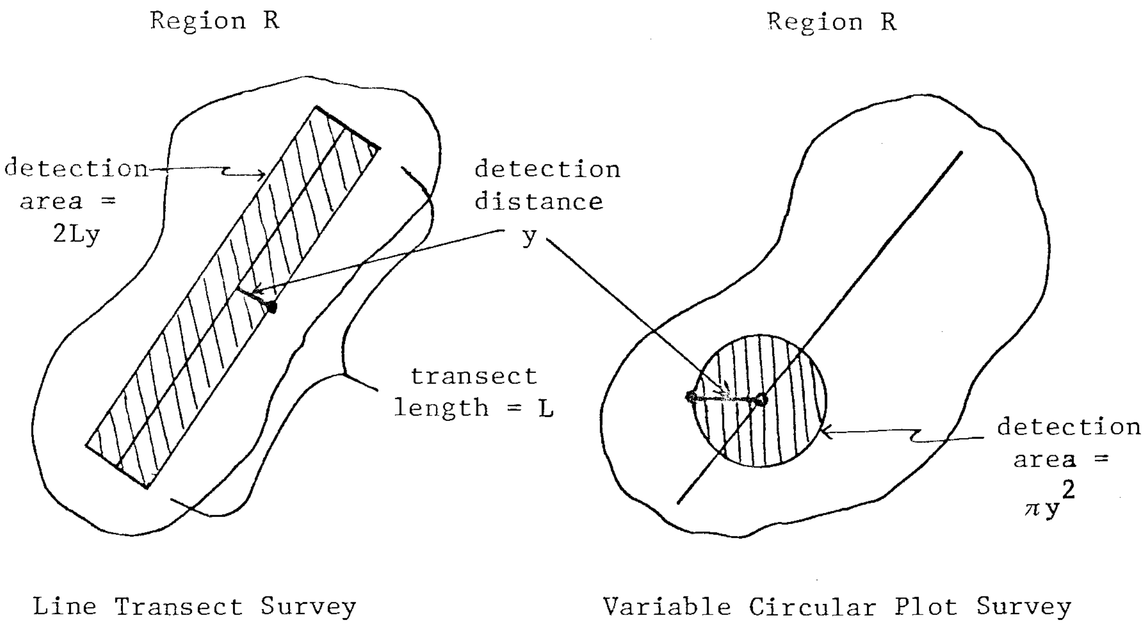


Figure 2. Detection areas recorded for line transect and variable circular plot surveys.

area is given by $a = \pi y^2$, where y is the straight line distance from the animal to the observer. For a general type of variable area survey, this area would be defined as the area interior to a contour of constant probability of detection relative to an observer. These detection areas are shown in Figure 2 for the line transect and variable circular plot surveys.

Now letting g(a) = Pr(detection given the detection area is a) we assume further that

(7) $g(0) = 1$, i.e. animals at zero distance from the observer are detected with probability one.

(8) $\int_0^\infty g(y)dy = A < \infty$.

The results of Burnham and Anderson (1976) and Ramsey (1979) can then be extended to show that, conditional on detection, the detection area of an animal has its probability density function given by $f(a) = g(a)/A$, and also that $E(n) = DA$. Thus, A is that quantity which, when it divides the number of detections, n, provides an unbiased estimate of D. If we were to suppose that the n detected animals occupied some subregion S in which all of the animals were detected and outside of which no animals were detected (such as in the fixed area survey), then A would be the area of the subregion S. Although this region generally does not exist - there will usually be some undetected animals which are closer to the observer than some of those which are detected - this interpretation has led Ramsey (1979) to call A the Effective Area Surveyed. The statistical problem thus becomes one of estimating A, conditional on n, and then estimating D by $\hat{D} = n/\hat{A}$. The estimation of A can be done by estimating the curve g and evaluating its integral; it can be done by estimating the density function F and evaluating it at zero; or it can be done by estimating the slope of the cumulative distribution function F of the observed detection areas at zero. These three approaches to the estimation of A will be discussed further in Chapter II in the context of the specific estimators which we are comparing.

II. THE ESTIMATORS

A. The Cum-D Estimator

The new nonparametric estimator of A we are proposing has been named the Cum-D estimator. The theoretical basis for this estimator was suggested by Ramsey and Scott (1981a); they note that if we consider, not the density function f for the observed detection areas but the cumulative distribution function F, then $F'(0) = 1/A$ and so we can estimate A by estimating the slope of F at the origin. This is the approach used with the Cum-D estimator. For a fixed sample size of n, the sample cumulative distribution function is defined in the standard way by $F_n(a) = (\text{number of } a_i \leq a)/n$, where $a_1, a_2, \ldots, a_n$ are the observed detection areas from a variable area survey. The relationship between F and A, and between F_n and $\hat{A}_{CD}$, are shown in Figure 3 below.

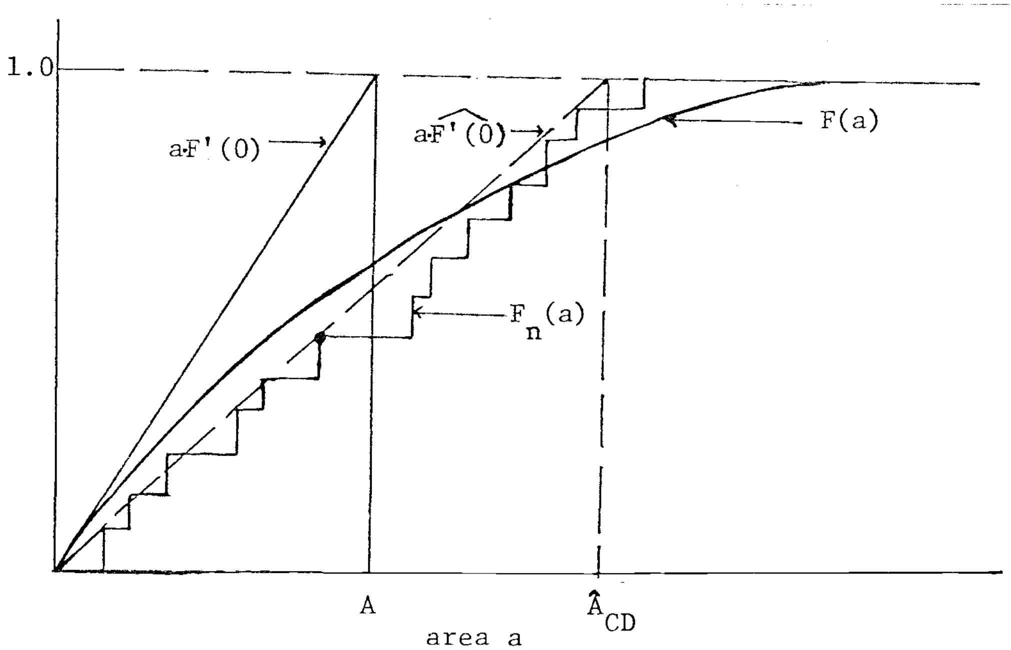


Figure 3. Theoretical and observed cumulative distribution function.

From Figure 3, we see that the Cum-D estimate $\hat{A}_{CD}$ of A is arrived at by first estimating the slope of the cumulative distribution function at zero and then extending a straight line with this estimated slope up to one and dropping a perpendicular line down to the horizontal axis to find $\hat{A}_{CD}$. For estimating the slope of F at the origin, we consider estimators of the form $F_n(a_k)/a_k$, where a_k is the k^{th} ordered detection area in the sample and $F_n(a_k) = k/n$. Thus, $\hat{A}_{CD} = (k/n)/a_k$ is the slope of a straight line from the origin to the point $(a_k, k/n)$ and $\hat{A}_{CD} = na_k/k$. For appropriate choice of k, this provides an intuitive estimate of F'(0). In Figure 3, k = 5 so that $\widehat{F'(0)} = (5/n)/a_5$.

The only remaining problem lies in the choice of an appropriate value of k to use. If k is very small, i.e. a_k is close to zero, the estimate of A will be based on only a small fraction of the data values, since the magnitudes of the observed detection areas greater than a_k will have no influence on $\hat{A}_{CD}$. This would generally lead to an estimator with a large variance. Using a large value of k will give an estimator with a smaller variance but with a larger bias, since the slope of F at a point near the origin will in general be larger than the slope of F at a large area value. The Cum-D estimator finds a value of k which is as large as possible while still insuring that there is no significant decrease in the slope between zero and a_k. We proceed as follows;

Step 1. Define the envelope function $\bar{F}_n(a)$ as a series of straight line segments with decreasing slopes in such a way that $F_n(a)$ always lies below $\bar{F}_n(a)$ for all $F_n(a) \geq 1/\sqrt{n}$. (The justifi-

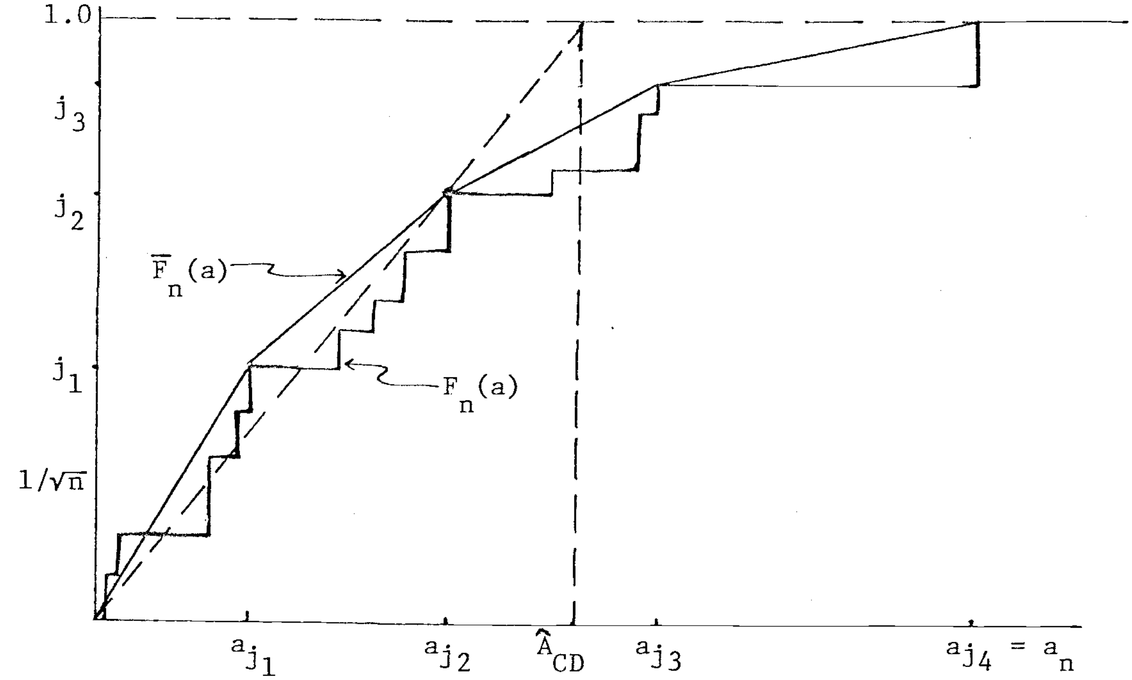


Figure 4. The envelope function $\bar{F}_n(a)$ with Cum-D estimate.

cation for the $\sqrt{n}$ rule will appear in Chapter IV.) Specifically,

$$\bar{F}_n(a) = \begin{cases} \left[\dfrac{j_1/n}{a_{j_1}}\right] \cdot a & 0 \leq a \leq a_{j_1} \\ \left[\dfrac{j_2/n - j_1/n}{a_{j_2} - a_{j_1}}\right] \cdot a + \left[\dfrac{j_1 a_{j2} - j_2 a_{j1}}{n(a_{j2} - a_{j1})}\right] & a_{j_1} \leq a \leq a_{j_2} \\ \vdots & \\ \left[\dfrac{1 - j_{m-1}/n}{a_n - a_{m-1}}\right] \cdot a + \left[\dfrac{j_{m-1} a_n - n a_{j_{m-1}}}{n(a_n - a_{j_{m-1}})}\right] & a_{j_{m-1}} \leq a \leq a_{j_m} = a_n \\ 1 & a \geq a_n \end{cases}$$

where $\left[\dfrac{j_1}{a_{j_1}}\right] = \underset{\sqrt{n} \leq j \leq n}{\text{maximum}} \left[\dfrac{j}{a_j}\right]$, $\left[\dfrac{j_2 - j_1}{a_{j_2} - a_{j_1}}\right] = \underset{j_1 < j \leq n}{\text{maximum}} \left[\dfrac{j - j_1}{a_j - a_{j_1}}\right]$, etc.

$\bar{F}_n(a)$ is illustrated in Figure 4 above.

It should be noted here that the slope of the line segment from $(a_{j_r}, j_r/n)$ to $(a_{j_{r+1}}, j_{r+1}/n)$ is equal to the density of detected animals whose detection areas are between a_{j_r} and $a_{j_{r+1}}$. Because we want to find the maximum order statistic a_k to use while still insuring that there is no significant decline in the slope of F, we proceed to

Step 2: Test the equality of the slope of the first line segment of the envelope function and the slope of the second line segment. If there is a difference, use $k = j_1$ and estimate A by $\hat{A}_{CD} = na_{j_1}/j_1$. If there is no difference, test the equality of the slope of the first line segment and the slope of the line segment joining $(a_{j_1}, j_1/n)$ and $(a_{j_3}, j_3/n)$. If there is a significant difference, let $k = j_2$; if not, test the equality of the first slope and the slope of the line segment joining $(a_{j_4}, j_4/n)$ and $(a_{j_1}, j_1/n)$. Continue in this manner until a point $(a_{j_r}, j_r/n)$ is found such that the slope of the line segment from $(a_{j_1}, j_1/n)$ to $(a_{j_r}, j_r/n)$ is significantly different from the slope of the first line segment of the envelope function, in which case let $k = j_{r-1}$, or until all points have been tested, in which case let $k = n$ and $\hat{A}_{CD} = a_n$.

There are two remaining points which should be mentioned. First, it should be noted that the subset of the data on which the estimate $\hat{A}_{CD}$ is based will have a minimum sample size of $\sqrt{n}$. This value was chosen arbitrarily as one possible value which will satisfy the necessary conditions for consistency of the estimator; this is discussed further in Chapter IV. Secondly, to test the hypothesis of the equality of the slopes of two line segments, we view the detected animals as Poisson events occurring in space and treat j_1 as a single observation from a Poisson distribution with parameter $\lambda_1 = D_1 a_{j_1}$ and $j_r - j_1$ as another independent observation from a Poisson distribution with parameter $\lambda_2 = D_2(a_{j_r} - a_{j_1})$. Then the generalized likelihood ratio test of H_o: $D_1 = D_2$ tells us to reject the hypothesis when

$$(1) \qquad j_1 \log(a_{j_1}/j_1) + (j_r - j_1)\log((a_{j_r} - a_{j_1})/(j_r - j_1)) - j_r \log(a_{j_r}/j_r) = \Lambda$$

is greater than 2.0 (Mood, Graybill and Boes, 1974).

B. The Fourier Series Estimator

The Fourier series estimation procedure, like the Cum-D procedure, is a nonparametric one. Work on the estimator was begun by Crain (1979) and expanded by Burnham, Anderson and Laake (1980). The approach makes use of a Fourier series expansion of the probability density function f(a), estimating the necessary coefficients with a finite number of terms m. The estimated function $\hat{f}$ is then evaluated at zero to give $\hat{A} = 1/\widehat{f(0)}$. Specifically, it is assumed that

$$f(a) = 1/w^* + \sum_{k=1}^{\infty} \left[b_k \cos(k\pi a/w^*) + c_k \sin(k\pi a/w^*) \right],$$

where w* is the half-period of the fundamental harmonic. Because sin(0) is zero, however, only the cosine terms are used in the approximation of f(0). Using the first m terms in the model, the estimates of the coefficients b_k are given by

$$\hat{b}_k = (2/nw^*) \sum_{i=1}^{n} \cos(k\pi a_i/w^*) \quad , k = 1, 2, \ldots , m$$

where the a_i are the observed detection areas from a variable area survey and all a_i are less than or equal to w*. Any detection areas greater than w* are eliminated from the analysis. Burnham et al. (1980) state that they have had success in using a w* value which eliminates between one and three percent of the largest detection areas, so for our use we have ordered the a_i and used $w^* = a_{25}$ when we have a sample of size 25, $w^* = a_{49}$ for a sample of size 50, $w^* = a_{98}$ for a sample of size 100 and $w^* = a_{196}$ for a sample of size 200. Although w* may be chosen large enough so as to include all of the detection areas, there are problems associated with choosing w* in this manner; this will be discussed more fully in Chapter VI.

The final decision to be made when using this estimator is in the choice of the number of terms to be included in the approximation. This choice presents the usual problem of trade-off between minimizing the bias, which can be made arbitrarily small by the inclusion of a sufficient number of terms, and minimizing the variance, which increases as the number of terms does. The stopping rule suggested by Burnham et al. (1980) is to choose the first value of m which satisfies

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

Then the estimate of A is given by $\hat{A} = \left[(1/w^*) + \sum_{k=1}^{m} \hat{b}_k\right]^{-1}.$

C. The Exponential Power - Maximum Likelihood Estimator

The exponential power-maximum likelihood estimator (EP-MLE) described by Pollock (1978) and Ramsey (1979) is a parametric procedure, in contrast to the Cum-D and Fourier series estimators. Ramsey (1979) defines a detectability curve kernel h(u) as any function which is continuous, nondecreasing on $(0,\infty)$ and has the properties that (i) $0 \leq h(u) \leq 1$ for all $u > 0$; (ii) $h(0) = 1$; and (iii) $\int_0^\infty h(u)\,du = 1$. He then shows that the detectability curve g(a) can be expressed as $g(a) = h(a/A)$, so that A can be viewed as a scale parameter in the distribution of the observed detection areas. The choice of the kernel function determines the shape of the detectability curve g independently of the parameter A.

One possibility for a detectability curve kernel that was used by Ramsey (1979) is the exponential power kernel, represented by the function

$$h(u) = \exp\left\{-(\Gamma(1 + 1/s)u)^s\right\}, \qquad 0 \leq u < \infty$$

for some shape parameter $s > 0$. If the shape parameter is known, the maximum likelihood estimator of A is given by $\hat{A} = s^s \Gamma(1 + 1/s) T_s^{\,1/s}$, where $T_s = \sum_{i=1}^{n} (a_i)^s$ and $a_1, a_2, \ldots, a_n$ are the observed detection areas. If s is unknown, it may be estimated by maximum likelihood procedures; this requires solving iteratively for s the equation

$$V_s/T_s = \left\{\log(s) + \log(T_s) + \Psi(1 + 1/s)\right\}/s,$$

where $V_s = (1/n)\sum_{i=1}^{n} (a_i)^s \log(a_i)$ and $\Psi(\cdot)$ is the digamma function, i.e. the derivative of the log of the gamma function $\Gamma(\cdot)$.

Andersen (1973) suggests that because of the possibility of obtaining nonconsistent estimators with maximum likelihood procedures, it is sometimes advantageous to use the distribution of a maximal invariant statistic under scale changes to estimate the shape parameter s. This is the approach we have used here. We let $z_i = a_i/a_n$, i = 1, 2, ... , n-1, where a_i is the i^{th} ordered observed detection area. Then $(z_1, z_2, \ldots, z_{n-1})$ is a maximal invariant statistic under scale changes and its probability density function is given by

$$f(z_1, z_2, \ldots, z_{n-1}) = (n!/s)\Gamma(n/s)\{\Gamma(1 + 1/s)\}^{-n} \cdot \left[1 + \sum_{i=1}^{n-1} z_i^s\right]^{n/s},$$

which does not depend on the scale parameter A. The maximum likelihood estimate of s using this approach is then given by the solution to the equation

$$V_s/T_s - (1/s)\cdot\left[\log(T_s) + \Psi(1 + 1/s) - \Psi(n/s)\right] + 1/n = 0.$$

The estimated value of s can then be used as before to estimate A.

The only remaining problem in using this estimator involves the calculation of $\Psi(1 + 1/s)$ and $\Psi(n/s)$. To get a numerical approximation for these values, we note that $\Gamma(t+1) = t\Gamma(t)$, and hence, $\Gamma(t) = \{\Gamma(t+1)\}/t$. By applying this recursively we see that

$$\Gamma(t) = \{\Gamma(t+k)\}/\{t(t+1)(t+2)\ldots(t+k-1)\}.$$

Now we can use Stirling's formula to approximate $\Gamma(t)$ as

$$\Gamma(t) \doteq \frac{\{(2\pi)^{\frac{1}{2}}\}\{\exp[-(t+k-1)]\}\{(t+k-1)^{t+k-\frac{1}{2}}\}}{t(t+1)(t+2)\ldots(t+k-1)}$$

so that

$$\log\{\Gamma(t)\} \doteq \tfrac{1}{2}\log(2\pi) - (t+k-1) + (t+k-\tfrac{1}{2})\log(t+k-1)$$
$$- \log\{t(t+1)(t+2)\ldots(t+k-1)\}$$

and
$$\Psi(t) \doteq -1 + \log(t+k-1) + \{(t+k-\tfrac{1}{2})/(t+k-1)\}$$
$$- 1/t - 1/(t+1) - \ldots - 1/(t+k-1).$$

Since Stirling's formula gives an approximation to $\Gamma(k+k)$ that is in error by less than 0.4% when (t+k) is greater than 20, we calculate the estimated values of $\Psi(t)$ by finding an integer k such that (t+k) is greater than 20 and using the above formula for the approximation. The resulting estimator will hereafter be referred to as the "Exponential Power - Maximum Likelihood Estimator", or EP-MLE, a nomenclature for the family of distributions that was suggested by Pollock (1978).

III. THE SIMULATION MODELING

A. The Models

To investigate the sampling properties of the three estimators, five different models for the detectability curve g were considered. The first model, which we have called the Ramsey-Scott model after a similar one used by Ramsey and Scott (1981a, 1981b), was included

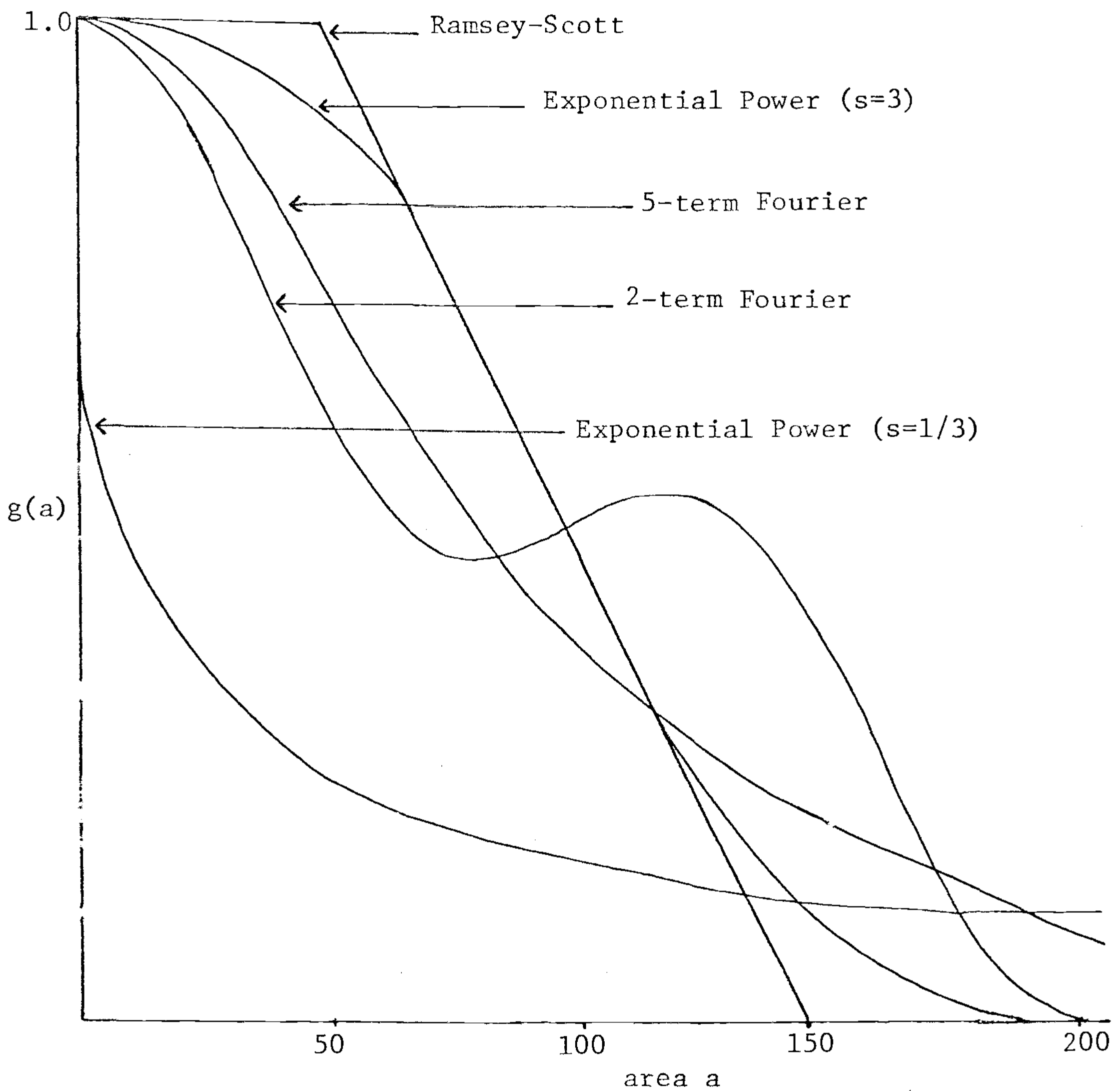


Figure 5. Detectability curves for the five simulation models.

as a model under which the Cum-D estimator was expected to perform well. It has a constant probability of detection for all areas from zero to 50 so that the slope of the cumulative distribution function F is constant on this range. The next two models were included to evaluate the performance of the Fourier series estimator under the assumptions of the estimation procedure. The 5-term Fourier model expresses the density curve f as the sum of five consecutive terms in the Fourier series expansion; the 2-term Fourier model uses an expansion with three terms but omits an intermediate term. The last two models we considered were generated by the exponential power detectability curve kernels; one has a shape parameter $s = 3$ while the other has a shape parameter $s = 1/3$. All five of the models were scaled so as to have an effective area surveyed of $A = 100$ units. The equations for each detectability curve g and cumulative distribution function F are given in Table 1, and the detectability curves are shown in graphical form in Figure 5.

B. The Data Generation

Random observations from each of the five models were generated using the probability integral transfrom and the IMSL random number generator GGUBFS. This generator is described by Lewis, Goodman and Miller (1969). An initial seed s_o is given and then a Uniform(0,1) deviate U_j is obtained by the algorithm

$$U_j = 2^{-31} s_j \text{ , where } s_j = 7^5 s_{j-1} (\text{mod } 2^{31} - 1).$$

Table 1. Detectability curve and cumulative distribution functions for the five simulation models.

Ramsey-Scott Model	$g(a) = \begin{cases} 1 \\ (150 - a)/100 \end{cases}$	$0 \le a \le 50$ $50 \le a \le 150$
	$F(a) = \begin{cases} a/100 \\ 1 - \frac{1}{2}((150 - a)/100)^2 \end{cases}$	$0 \le a \le 50$ $50 \le a \le 150$
2-term Fourier Model	$g(a) = \begin{cases} 1/2 + (1/3)\cos(\pi a/200) \\ \quad + (1/6)\cos(3\pi a/200) \end{cases}$	$0 \le a \le 200$
	$F(a) = \begin{cases} a/200 + (2/3\pi)\sin(\pi a/200) \\ \quad + (1/9\pi)\sin(3\pi a/200) \end{cases}$	$0 \le a \le 200$
5-term Fourier Model	$g(a) = .34016 + \sum_{j=1}^{5} \left[.4^j \cos(j\pi a/294)\right]$	$0 \le a \le 294$
	$F(a) = .34016a + \sum_{j=1}^{5} \left[.4^j (294/\pi j)\sin(j\pi a/294)\right]$	$0 \le a \le 294$
Exponential Power(s=3) Model	$g(a) = \exp\left[-\left(\Gamma(4/3)a/100\right)^3\right]$	$0 \le a < \infty$
	$F(a) =$ no closed form	
Exponential Power(s=1/3) Model	$g(a) = \exp\left[-\left(\Gamma(4)a/100\right)^{1/3}\right]$	$0 \le a < \infty$
	$F(a) = \begin{cases} 1 - \left[\exp\left(-(6a/100)^{1/3}\right)\right]\left[\frac{1}{2}(6a/100)^{2/3} \\ \quad + (6a/100)^{1/3} + 1\right] \end{cases}$	$0 \le a < \infty$

The total number of random deviates generated was 37,500, which is well below the algorithm's period of 2^{31}.

Once a deviate U_j has been obtained, we let $a = F^{-1}(U_j)$ by solving iteratively for a the equation $F(a) = U_j$. The iteration was continued until we obtained an a-value which differed from the preceeding one by less than 0.01. The last two a-values are then averaged to give the simulated detection area a_j. It should be noted that each random deviate was used to generate a detection area for each of the five models, and the resulting simulated detection areas are not independent from one model to the next. Also, since there is not closed form solution for the cumulative distribution function of the exponential power model with s = 3, another IMSL subroutine, MDGAM, was used to find the necessary values for the incomplete gamma function. As noted in the IMSL manual (1980), a series expansion or a continued fraction expansion in used in the determination of these values.

IV. CONSISTENCY OF THE CUM-D ESTIMATOR

A. Summary of Results

As was shown in Chapter I, we can estimate the effective area surveyed A by estimating the derivative or slope of the cumulative distribution function at the origin, since $F'(0) = f(0) = 1/A$. If we write

$$F'(0) = \lim_{a \to 0} \left[\frac{F(a) - F(0)}{a - 0}\right] = \lim_{a \to 0} \left[\frac{F(a)}{a}\right],$$

it seems reasonable to estimate $1/F'(0)$ by $\tilde{A}_n(k)$, where

$$\tilde{A}_n(k) = \left[\frac{F_n(a_k) - F(0)}{a_k - 0}\right]^{-1} = \left[\frac{k/n}{a_k}\right]^{-1} = \frac{na_k}{k}$$

and a_k is an extreme order statistic, i.e. $a_k \to 0$ as $n \to \infty$. This general form is used for other estimators as well, different estimators using different rules for the determination of which a_k to use. Some of these estimators are discussed in Chapter VI.

To show the consistency of the Cum-D estimator, we will view the sample size n as being fixed and consider the behavior of the estimator as $n \to \infty$. First, we will show under what conditions the estimator $\tilde{A}_n(k)$ for a non-random k will converge to A in mean square. Then we will show that the k-value determined empirically by the Cum-D estimation procedure satisfies the necessary conditions with probability one. Since convergence of the mean square error to zero implies the consistency of the estimator, the arguement is then complete.

The theorems which will follow utilize the concept of a basal area a_B, defined by $a_B = \sup\{a;\ F(a)=f(0)a\}$; a_B is therefore the area of a region about the observer in which there is perfect detectability. If a_B is positive, the function F on the interval $(0, a_B)$ is a straight line through the origin with slope $F'(0) = F(a_B)/a_B$, and we can estimate $F'(0)$ by $F_n(a_B)/a_B$ using a central order statistic rather than an extreme order statistic. We will consider the case for the zero basal area first.

B. The Zero Basal Area Case With Non-Random k

Theorem 1: Let $a_1, a_2, \ldots, a_n$ be the ordered detection areas for a random sample of size n from a distribution with cumulative distribution function F(a). Assume that $a_B = \sup\{a: F(a)=f(0)\cdot a\}$ is zero and that the density function f has a derivative which is everywhere bounded and continuous at zero. If k is such that $k/n \to 0$ and $k \to \infty$ as $n \to \infty$, then the asymptotic mean square error of $\tilde{A}_n(k) = na_k/k$ is zero.

Proof: Let $U_k = F(a_k)$, $k = 1,2,\ldots,n$. Then $U_1, U_2, \ldots, U_n$ can be viewed as the order statistics for a random sample of size n from a Uniform(0,1) distribution and each U_k has a Beta distribution with parameters k and n-k+1. This gives

$$E(U_k) = \frac{k}{n+1} \quad \text{and} \quad Var(U_k) = \frac{k(n-k+1)}{(n+1)^2(n+2)} .$$

Using a Taylor series expansion about the point $u_o = E(U_k)$, we can write

$$a_k = F^{-1}(U_k) = F^{-1}(k/(n+1)) + \left[\frac{d}{du} F^{-1}(u)\Big|_{u=\frac{k}{n+1}}\right]\left[U_k - \frac{k}{n+1}\right]$$

$$+ \left[\frac{d^2}{du^2} F^{-1}(u)\Big|_{u=\frac{k}{n+1}}\right]\frac{\left[U_k - \frac{k}{n+1}\right]^2}{2} + R,$$

where R is the remainder term. Using the change of variable $x = F^{-1}(u)$, $u = F(x)$ and $du/dx = f(x)$, we see that

$$\left[\frac{d}{du} F^{-1}(u)\right] = \frac{1}{f(F^{-1}(u))} \quad \text{and} \quad \left[\frac{d^2}{du^2} F^{-1}(u)\right] = \frac{-f'(F^{-1}(u))}{\{f(F^{-1}(u))\}^3}$$

so that

$$(2) \qquad E(a_k) = F^{-1}(k/(n+1)) - \frac{f'\{F^{-1}(k/(n+1))\}}{2\{f[F^{-1}(k/(n+1))]\}^3} \cdot \frac{k(n-k+1)}{(n+1)^2(n+2)}$$

$$+ o(k/n^2).$$

The determination of the order of the remainder terms involves the standard central moments of the Beta distribution and is shown in detail in Appendix A.

Now letting $k/n \to 0$ and $k \to \infty$ as $n \to \infty$, and using a Taylor series expansion of the terms in the above expression for $E(a_k)$ about zero, we find that

$$F^{-1}(k/(n+1)) = F^{-1}(0) + \frac{k/(n+1)}{f(F^{-1}(0))} - \frac{(k/(n+1))^2 f'(F^{-1}(0))}{2\{f(F^{-1}(0))\}^3}$$

$$+ o(k^2/n^2)$$

and

$$\frac{f'\{F^{-1}(k/(n+1))\}}{2\{f[F^{-1}(k/(n+1))]\}^3}\cdot\frac{k(n-k+1)}{(n+1)^2(n+2)} = \frac{k(n-k+1)}{(n+1)^2(n+2)}\cdot\frac{f'(F^{-1}(0))}{\{f(F^{-1}(0))\}^3}$$

$$+ \; o(k/n^2).$$

Since $F^{-1}(0) = 0$ and $f(0) = 1/A$, we can substitute into expression (2) for $E(a_k)$ to get

$$E(a_k) = A\left[\frac{k}{n+1}\right] - \left[\frac{f'(0+)A^3}{2}\right]\left[\frac{k}{n+1}\right]^2 - \left[\frac{f'(0+)A^3}{2(n+2)}\right]\left[\frac{k}{n+1}\right]$$

$$+ \; o(k^2/n^2)$$

so that

$$E(\tilde{A}_n(k)) = A - \frac{f'(0+)A^3}{2}\cdot\left[\frac{k}{n}\right] + o(k/n).$$

To find the variance of a_k, we proceed in a similar fashion, using only the first-order term in the Taylor series expansion. We write

$$a_k = F^{-1}(k/(n+1)) + \frac{\left[U_k - \frac{k}{n+1}\right]}{f\{F^{-1}(k/(n+1))\}} + R$$

so that $\mathrm{Var}(a_k) = \mathrm{Var}(U_k)/\{f[F^{-1}(k/(n+1))]\}^2 + o(k^2/n^3)$

$$= \frac{1}{\{f[F^{-1}(k/(n+1))]\}^2}\cdot\frac{k(n-k+1)}{(n+1)^2(n+2)} + o(k^2/n^3).$$

Again letting $k/n \to 0$ and $k \to \infty$ as $n \to \infty$, a Taylor series expansion of the variance term about zero gives

$$\mathrm{Var}(a_k) = A^2\left[\frac{k}{(n+1)(n+2)}\right]\left[1 - \left(\frac{k}{n+1}\right)\left(1 + 2f'(0+)A^2\right)\right] +$$

$$o(k^2/n^3).$$

Thus, for the variance of $\tilde{A}_n(k)$ we get

$$\mathrm{Var}(\tilde{A}_n(k)) = A^2/k - A^2(1 + 2f'(0+)A^2)/(n+1) + o(1/n)$$

From this we can see that the asymptotic mean square error of $\tilde{A}_n(k)$ is given by

$$\mathrm{MSE}(\tilde{A}_n(k)) = A^2/k - A^2(1 + 2f'(0+)A^2)/(n+1) + (k^2/n^2)(f'(0+)A^3/2)^2 + o(k^2/n^2) + o(1/n).$$

Since $k/n \to 0$ and $k \to \infty$ as $n \to \infty$, the asymptotic mean square error of $\tilde{A}_n(k)$ is zero, and Theorem 1 is proved.

We note that if $f'(0+)$ is zero, we want to have k increase rapidly with n so that $k/n \to 0$ slowly; this will control the variance. If, however, $f'(0+)$ is a large negative number, we want to increase k slowly to control the bias. For large n, we can show that the optimal k-value is one which is proportional to $n^{2/3}$.

Corollary: For large n, the mean square error of $\tilde{A}_n(k)$ is minimized by choosing

$$k = k^* = n^{2/3}\left[2/\{A^4(f'(0+))^2\}\right]^{1/3}.$$

Proof:

$$\frac{d}{dk}\mathrm{MSE}(A_n(k)) = \frac{d}{dk}\left[\frac{A^2}{k} + \left(\frac{A^3 f'(0+)}{2}\cdot\frac{k}{n}\right)^2\right]$$

$$= \frac{-A^2}{k^2} + \left[\frac{A^3 f'(0+)}{2n}\right]^2 \cdot 2k$$

$$= -A^2/k^2 + \left[A^6\{f'(0+)\}^2/2n^2\right]k$$

Setting this equation equal to zero and solving for k yields the k* value above. Also,

$$\frac{d^2}{dk^2} \; MSE(\tilde{A}_n(k)) = 2A^2/k^3 + A^6\{f'(0+)\}^2/2n^2,$$

which is always positive. This shows that the mean square error is a convex function of k so that k* is a unique global minimum. And, although we have treated k as continuous, the integer value which minimizes the mean square error will be one of the integers closest to k*, if k* is not an integer itself.

C. The Positive Basal Area Case With Non-Random k

In the case when the basal area a_B is positive, we can use a central order statistic to estimate F'(0). The proof of the consistency of $\tilde{A}_n(k)$ in this case is similar to the proof of Theorem 1.

Theorem 2: Let $a_1, a_2, \ldots, a_n$ be the ordered detection areas for a random sample of size n from a distribution with cumulative distribution function F(a). Assume that $a_B = \sup\{a: F(a) = f(0)\cdot a\}$ is positive and that the density function f has a derivative which is everywhere bounded and continuous at a_B. If k is such that $k/n \to F(a_B)$ and $k \to \infty$ as $n \to \infty$, then the asymptotic mean square error of $\tilde{A}_n(k) = na_k/k$ is zero.

Proof: The proof follows the same line of reasoning as did the proof of Theorem 1 to equation (2), at which point we had

$$E(a_k) = F^{-1}(k/(n+1)) - \frac{f'\{F^{-1}(k/(n+1))\}}{2\{f[F^{-1}(k/(n+1))]\}^3}\cdot\frac{k(n-k+1)}{(n+1)^2(n+2)}$$

$$+ \; o(k/n^2).$$

Again using Taylor series expansions of the terms about the point $F(a_B)$, and letting $k/n \to F(a_B)$ and $k \to \infty$ as $n \to \infty$, we get

$$F^{-1}(k/(n+1)) = F^{-1}(F(a_B)) + \frac{(k/(n+1)) - F(a_B)}{f\{F^{-1}(F(a_B))\}}$$

$$- \left[\frac{k}{n+1} - F(a_B)\right]^2 \cdot \frac{f'\{F^{-1}(F(a_B))\}}{2\{f[F^{-1}(F(a_B))]\}^3}$$

$$+ o\left[\left(\frac{k}{n+1} - F(a_B)\right)^2\right]$$

and

$$\frac{f'\{F^{-1}(k/(n+1))\}}{2\{f[F^{-1}(k/(n+1))]\}^3} \cdot \frac{k(n-k+1)}{(n+1)^2(n+2)} = \frac{[F(a_B)][1 - F(a_B)] \cdot f'\{F^{-1}(F(a_B))\}}{\{2\{f[F^{-1}(F(a_B))]\}^3\}(n+2)}$$

$$+ \frac{o\left[\frac{k}{n+1} - F(a_B)\right]}{(n+2)}.$$

Since $F(a_B) = f(0) \cdot a_B$ and $f(a_B) = f(0) = 1/A$, these equations can be simplified and substituted into equation (2) to give finally

$$E(a_k) = A\left(\frac{k}{n+1}\right) - \left[\frac{f'(a_B+)A^3}{2}\right]\left[\frac{k}{n+1} - F(a_B)\right]^2$$

$$- \frac{F(a_B)\{1 - F(a_B)\} f'(a_B+)A^3}{2(n+2)}$$

$$+ o\left[\left(\frac{k}{n+1} - F(a_B)\right)^2\right] + o(1/n).$$

From this we can then see that $E(\tilde{A}_n(k))$ is given by

$$E(\tilde{A}_n(k)) = A - \left[\frac{k}{n+1} - F(a_B)\right]^2 \cdot \left[\frac{f'(a_B+)A^3}{2F(a_B)} - \frac{F(a_B)\{1 - F(a_B)\} f'(a_B+)A^3}{2k}\right] + o\left[\left(\frac{k}{n+1} - F(a_B)\right)^2\right] + o(1/n)$$

To find the variance of $\tilde{A}_n(k)$, we proceed as before with the Taylor series expansion of a_k about the point $k/(n+1)$, and then the Taylor series expansions of the resulting terms about the point $F(a_B)$, letting $k/n \rightarrow F(a_B)$ and $k \rightarrow \infty$ as $n \rightarrow \infty$. This gives

$$\mathrm{Var}(a_k) = \frac{F(a_B)\{1 - F(a_B)\}A^2}{(n+2)} + \left[\frac{k}{n+1} - F(a_B)\right] \cdot \left[\{1 - 2F(a_B)\}A^2 + 2F(a_B)\{1-F(a_B)\}f'(a_B+)A^4\right] + \frac{O\left(\frac{k}{n+1} - F(a_B)\right)}{(n+2)} + o(k/n^2).$$

This then gives

$$\mathrm{Var}(\tilde{A}_n(k)) = A^2\{1 - F(a_B)\}/nF(a_B) + o(1/n),$$

so that the mean square error can be obtained as

$$\mathrm{MSE}(\tilde{A}_n(k)) = \frac{A^2\{1 - F(a_B)\}}{nF(a_B)} + \left(\left[\frac{\{1 - F(a_B)\}f'(a_B+)A^3}{2n}\right] + \left[\frac{k}{n+1} - F(a_B)\right]^2 \cdot \left[\frac{f'(a_B+)A^3}{2F(a_B)}\right]\right)^2 + o(1/n).$$

Since this is a sum of terms of order $O(1/n)$, $o(1/n)$ and $O(v^4)$, where $v = (k/(n+1)) - F(a_B)$, the asymptotic mean square error of $\tilde{A}_n(k)$ is zero, and Theorem 2 is proved.

It should be noted at this point that the order of the bias term for the case when a_B is zero is larger than the order of the bias term for the case when a_B is positive. We would therefore expect an estimator of this form to perform more poorly on a model such as the Fourier series or exponential power model than on one such as the Rasmey-Scott model.

D. Consistency With the Empirical k-Value

Although the two preceding theorems give simple conditions under which $\tilde{A}_n(k)$ is a consistent estimator of A, they generally are not applicable in practice because a_B is unknown, and we therefore do not know to what value k/n should converge. To solve this problem, the Cum-D estimation procedure determines k empirically from the data, essentially estimating the basal area to arrive at an estimate of A. The estimation of the basal area was an approach suggested by Emlen (1971); the Cum-D estimator can be viewed as a formalized, systematic extension of his "inspection" procedure. To argue the consistency of the Cum-D estimator, then, we want to show that the empirically-determined k-value satisfies the conditions of Theorem 1 or Theorem 2, as appropriate, i.e. that as $n \to \infty$, $k/n \to F(a_B)$ and $k \to \infty$, whether a_B is zero or positive.

We consider first the case when the basal area is positive, such as in the Ramsey-Scott model. From Theorem 2 we saw that

$$E(a_k) = F^{-1}(k/(n+1)) - \frac{f'\{F^{-1}(k/(n+1))\}}{2\{f[F^{-1}(k/(n+1))]\}^3} \cdot \frac{k(n-k+1)}{(n+1)^2(n+2)}$$

$$+ \ o(k/n^2).$$

Since $f'(a)$ is zero for all a less than or equal to a_B, and $F^{-1}(u)$ is equal to Au for all u less than $F(a_B)$, we can see that when $k/(n+1)$ is less than $F(a_B)$, $E(\tilde{A}_n(k)) = An/(n+1)$, which converges to A as $n \to \infty$. Thus, it is sufficient to show that the probability of $k/(n+1)$ being greater than $F(a_B)$ can be made small, i.e. that

$$\Pr\left[\limsup_{n\to\infty} (k/n) \le F(a_B)\right] = 1.$$

Then, since the minimum $\sqrt{n}$ rule guarantees that $k \to \infty$ as $n \to \infty$, we will have shown that the necessary conditions are satisfied.

<u>Theorem 3</u>: Let k be the index of the order statistic selected by the Cum-D estimation procedure for estimating A from a random sample of n ordered detection areas $a_1, a_2, \ldots, a_n$. Assume that the basal area a_B is positive and that $f(a)$, the probability density function for each of the a_i, is a monotone decreasing function for a greater than a_B. For $\Delta > 0$ and $\epsilon > 0$, we can find an N such that

$$\Pr\left[k/n < F(a_B) + \Delta \text{ for all } n \ge N\right] \ge 1 - \epsilon .$$

Proof: Define $a_\Delta = F^{-1}(F(a_B) + \tfrac{1}{2}\Delta)$ and $\varphi = \{a_\Delta \cdot f(0) - F(a_\Delta)\}$, as shown in Figure 6. Let δ be less than $(a_B/(a_B + a_\Delta))\cdot\varphi$, and define a δ-band about the true cumulative distribution function F. Since $F_n \to F$, almost surely, uniformly in a, we can find an N_1 such that

$$\Pr\left[|F_n(a) - F(a)| < \delta \text{ for all } n > N_1\right] \ge 1 - \epsilon;$$

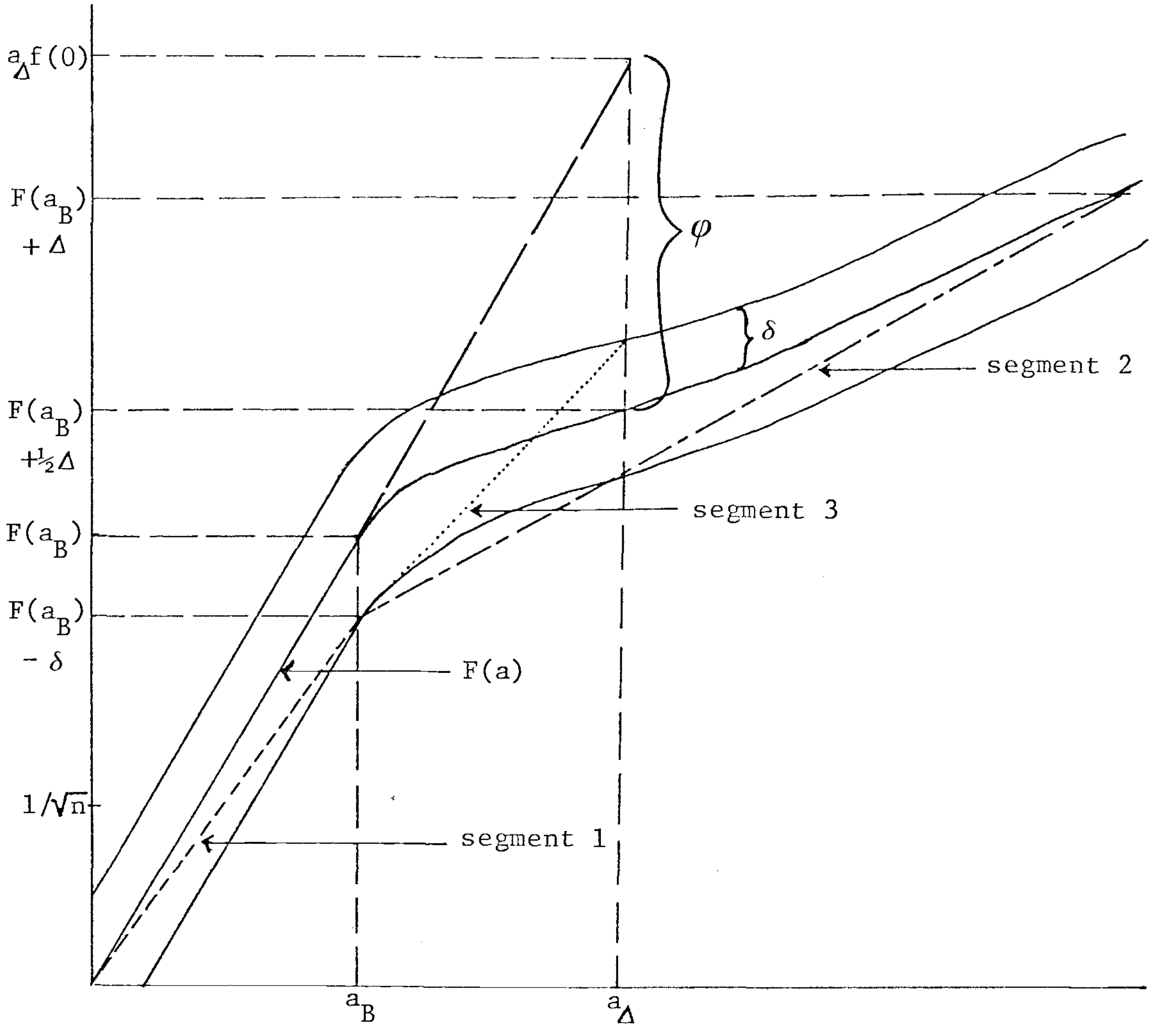


Figure 6. The cumulative distribution function with δ-band and least favorable envelope.

see, for example, Durbin (1971). Now k/n will be greater than $F(a_B) + \Delta$ only if, in step 2 of the Cum-D estimation procedure, we fail to reject the hypothesis of equal slopes using the first line segment of the envelope function and the line segment from $(a_{j_1}, j_1/n)$ to $(a_{j_r}, j_r/n)$ for some j_r/n which is greater than $F(a_B) + \Delta$. If

F_n lies within the δ-band about F, the minimum slope that the first line segment can have is $(F(a_B) - \delta)/a_B$, as illustrated by segment 1 in Figure 6. Conditional on the placement of this first line segment, segment 2 has the maximum slope possible while still insuring that j_r/n is greater than $F(a_B) + \Delta$. The slope of segment 2 is less than the slope of segment 3 in Figure 6, which is given by

$$\left\{\left(F(a_B) + \Delta/2 + \delta\right) - \left(F(a_B) - \delta\right)\right\}/(a_\Delta - a_B).$$

Our choice for δ also insures that the slope of segment 1 is greater than the slope of segment 3. Thus, if we reject the hypothesis of equal slopes for line segments 1 and 3 in Figure 6, we would also reject the hypothesis of equal slopes for any line segments arising from the envelope function of an F_n within the δ-band about F and having the property that k/n is greater than $F(a_B) + \Delta$.

To test the equality of the slopes of line segments 1 and 3, we get from equation (1) in Chapter II the following test statistic:

$$\Lambda = n_1 \log(n_1/a_B) + (n_2 - n_1)\log((n_2 - n_1)/(a \ - a_B)) - n_2 \log(n_2/a \ ),$$

where $n_1/n = F_{-}(a_B) = F(a_B) - \delta = a_B f(0) - \delta$ and $n_2/n = F_n(a_\Delta) = F(a_\Delta) + \delta = a_\Delta f(0) - \varphi + \delta$. Thus, for the least favorable envelope with segments 1 and 3, we get

$$n_1/a_B = n(a_B f(0) - \delta)/a_B = n(f(0) - \delta/a_B) = nf(0)\cdot\left[1 - \frac{\delta A}{a_B}\right],$$

$$\frac{n_2 - n_1}{a_\Delta - a_B} = \frac{n(a_\Delta f(0) - \varphi + \delta) - n(a_B f(0) - \delta)}{a_\Delta - a_B} = nf(0)\cdot\left[1 + \frac{(2\delta - \varphi)A}{a_\Delta - a_B}\right],$$

and $n_2/n = n(a_\Delta f(0) - \varphi + \delta)/a_\Delta = nf(0)\cdot\left[1 + \frac{(\varphi - \delta)A}{a_\Delta}\right].$

This then gives

$$\Lambda = n\left\{(a_B f(0)-\delta)\log\left\langle 1-\frac{\delta A}{a_B}\right\rangle + \left[(a_\Delta - a_B)f(0) + 2\delta - \varphi\right]\cdot\log\left\langle 1+\frac{(2\delta\,\varphi)A}{a_\Delta\ a_B}\right\rangle\right.$$

$$\left. - (a_\Delta f(0) + \delta - \varphi)\log\left\langle 1+\frac{(\delta-\varphi)A}{a_\Delta}\right\rangle\right\}.$$

Since Λ is a likelihood ratio test statistic, it always greater than or equal to zero, and our choice of δ insures that it is strictly positive. Thus, by increasing n, Λ can be made arbitrarily large. If we let N_2 be the sample size needed to have Λ greater than 2.0 and N_3 be such that $1/\sqrt{N_3} < F(a_B)$, by setting $N = \text{maximum}(N_1, N_2, N_3)$ we can insure that we reject the hypothesis of equal slopes, i.e. that $\Pr(k/n < F(a_B) + \Delta \text{ for all } n \geq N) \geq 1 - \epsilon$, and Theorem 3 is proved.

The case when the basal area a_B is zero can now be treated as a corollary to Theorem 3.

Corollary: Let k be the index of the order statistic selected by the Cum-D estimation procedure for estimating A from a random sample of n ordered detection areas. Assume that the basal area is positive and that f(a), the probability density function for the observed detection areas, is a monotone decreasing function of the area a. For $\Delta > 0$ and $\epsilon > 0$, we can find an N such that

$$\Pr(k/n < \Delta \text{ for all } n \geq N) \geq 1 - \epsilon .$$

Proof: We again define $a = F^{-1}(\tfrac{1}{2}\Delta)$ and $\varphi^* = a_\Delta f(0) - F(a_\Delta)$, then choose δ to be less than $(\varphi^*/2)(a^*/(a_\Delta + a^*))$, where a^* is such that $0 < a^* < a$ and $F(a^*)/a^* = (a_\Delta f(0) - \varphi^*/2)/a_\Delta$, as shown in Figure 7.

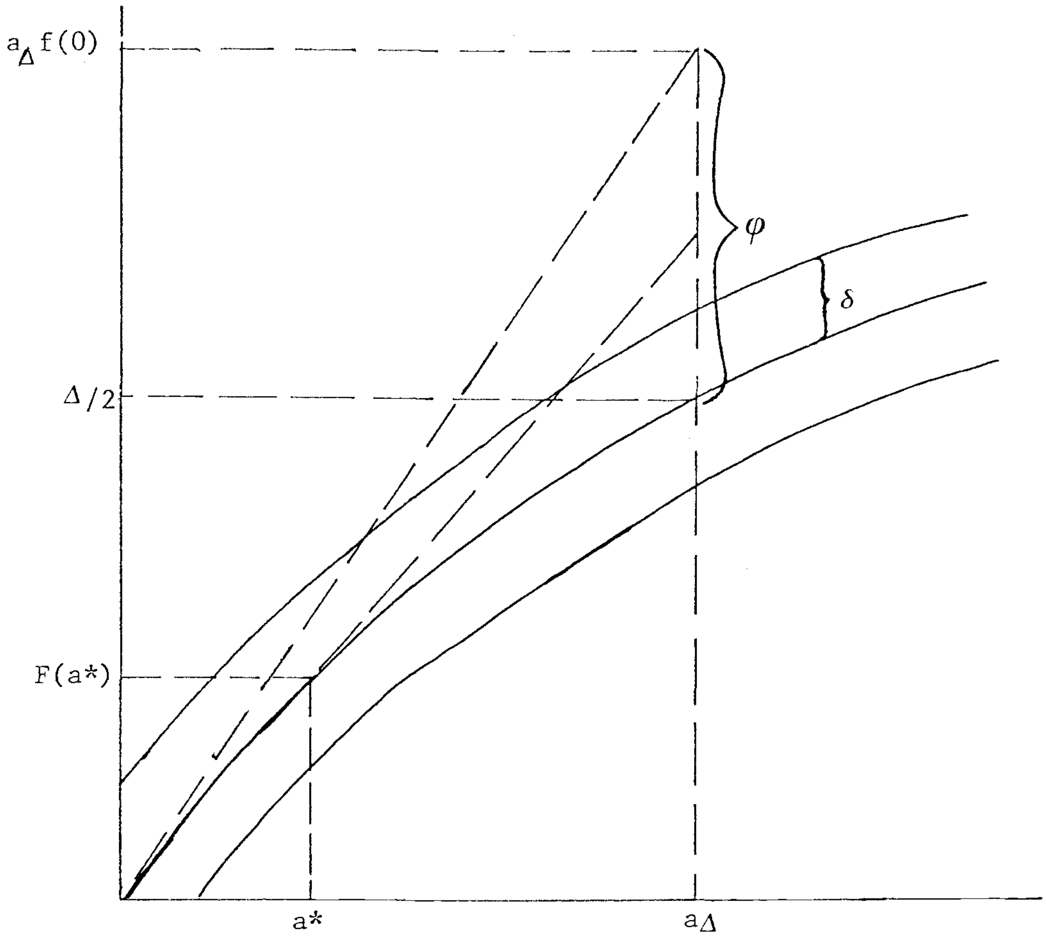


Figure 7. The cumulative distribution function with δ-band and the point a^*.

We now replace a_B in Theorem 3 by a^* and φ by $\varphi^*/2$ and proceed as before, finding an N such that $\Pr\{k/n < \Delta\}=1$ for all F_n within the the δ-band about F, which implies that

$$\Pr\{k/n < \Delta \text{ for all } n \geq N\} \geq 1-\epsilon,$$

and the proof is complete.

V. SIMULATION AND ANALYSIS RESULTS

A. Rates of Convergence of the Cum-D Estimator

Although the results of Chapter IV show that the Cum-D estimation procedure gives a consistent estimator of A, and that its asymptotic variance is zero, they provide no information about the rate of convergence of the estimates to A in practice. To estimate the rate of convergence of the Cum-D estimator as well as the Fourier series and EP-MLE estimators, simulation results were used. The Ramsey-Scott, the 5-term Fourier and the exponential power (s=3) models were each used to generate 100 data sets with samples of sizes 25, 50, 100 and 200. If we assume that $Y = c/n^b$ so that $\log(y) = \log(c) - b \cdot \log(n)$, the convergence rates can be estimated by using log-log regression. For the dependent variable Y, we considered the absolute value of the bias, the standard deviation and the root mean square. The results of these regressions for each model and each estimator are displayed in Table 2. Although a more concise statement and larger sample sizes would be obtained by considering a general linear model with one regression variable for log(n) and three indicator variables for the three different simulation models, F-tests indicate that the rates of convergence are moderately model-dependent for the Cum-D estimator ($p = .0725$) and the Fourier series estimator ($p = .0870$) and are very model-dependent for the EP-MLE estimator ($p = .0050$) when using the root mean square as the dependent variable. Although the root mean square of the Fourier series estimator does converge more quickly

Table 2. Log-log regressions of sample size on absolute value of the bias, standard deviation and root mean square.

Estimator	Model		Sample size n 25	50	100	200	Regression Coefficients β_0	β_1
Cum-D	Ramsey-Scott	Bias	5.282	4.855	-0.370	-0.252	7.407	-1.688
		SD	23.038	18.847	15.641	14.094	3.875	-0.238
		RMS	23.636	19.114	15.465	14.097	3.957	-0.254
	5-term Fourier	Bias	22.729	19.684	9.215	8.092	5.182	-0.745
		SD	43.098	29.186	22.367	19.893	4.897	-0.373
		RMS	48.724	35.203	24.191	21.476	5.166	-0.409
	Expon. Power	Bias	7.414	6.831	3.029	1.445	4.864	-0.825
		SD	23.757	22.092	17.216	14.973	3.961	-0.237
		RMS	24.887	23.124	17.480	15.042	4.082	-0.258
Fourier Series	Ramsey-Scott	Bias	*12.561	12.369	1.862	1.676	6.422	-1.145
		SD	*71.920	47.353	28.795	17.559	6.494	-0.682
		RMS	*73.009	48.942	28.855	17.639	6.546	-0.691
	5-term Fourier	Bias	8.474	6.096	0.902	1.028	6.029	-1.188
		SD	30.403	20.655	16.974	12.243	4.742	-0.422
		RMS	31.562	21.536	16.988	12.287	4.850	-0.443
	Expon. Power	Bias	3.987	12.073	-1.353	1.877	3.934	-0.642
		SD	38.650	45.675	18.883	18.354	5.246	-0.450
		RMS	38.855	47.244	18.932	18.450	5.277	-0.454
EP-MLE	Ramsey-Scott	Bias	-4.761	1.007	0.402	2.590	2.109	-0.400
		SD	20.215	13.511	8.423	6.994	4.668	-0.528
		RMS	20.768	13.550	8.433	7.458	4.624	-0.512
	5-term Fourier	Bias	-3.711	0.561	-3.100	-0.689	2.406	-0.479
		SD	35.323	22.364	15.269	13.126	5.052	-0.484
		RMS	35.518	22.371	15.581	13.144	5.504	-0.482
	Expon. Power	Bias	-5.713	-1.353	-5.352	-4.040	1.031	-0.560
		SD	21.949	13.918	12.826	11.041	3.985	-0.309
		RMS	22.679	13.977	13.897	11.757	3.928	-0.285

*The bias, SD, and RMS figures for the Fourier series estimator on the Ramsey-Scott model were calculated after omitting one estimate of -816.75.

than does the root mean square of the Cum-D estimator, the Cum-D procedure starts out initially with smaller RMS values and therefore perform better than the Fourier series estimator on the range of sample sizes we considered. The EP-MLE, however, does better than either the Cum-D or the Fourier series procedures in almost every case - the Fourier series is slightly better than the EP-MLE estimator on the 5-term Fourier model. The regression lines for the root mean square are shown in Figure 8 to illustrate these relationships.

A graphical comparison of the effects of increasing sample size can also be made by considering Figures 9a, 9b and 9c. Here, the sample cumulative distribution functions of the 100 estimates obtained by each of the estimation procedures and for each of the four sample sizes are presented. From these, we see that increasing the sample size has the effect of reducing the upper tail of the sampling distribution for the Cum-D estimator while the lower tail remains relatively unchanged for increasing sample sizes. For the Fourier series estimator, the behavior is similar, although large sample sizes seem to have relatively little effect on the occurrence of very large overestimates which the Fourier series procedure occasionally produces. The EP-MLE estimator exhibits the general type of behavior that would be expected of a maximum likelihood estimator, pulling in both tails of the sampling distribution as the sample size increases, although there does not seem to be a large difference in the upper tail between samples sizes of 100 and 200, and the effect on the lower tail is a bit unclear.

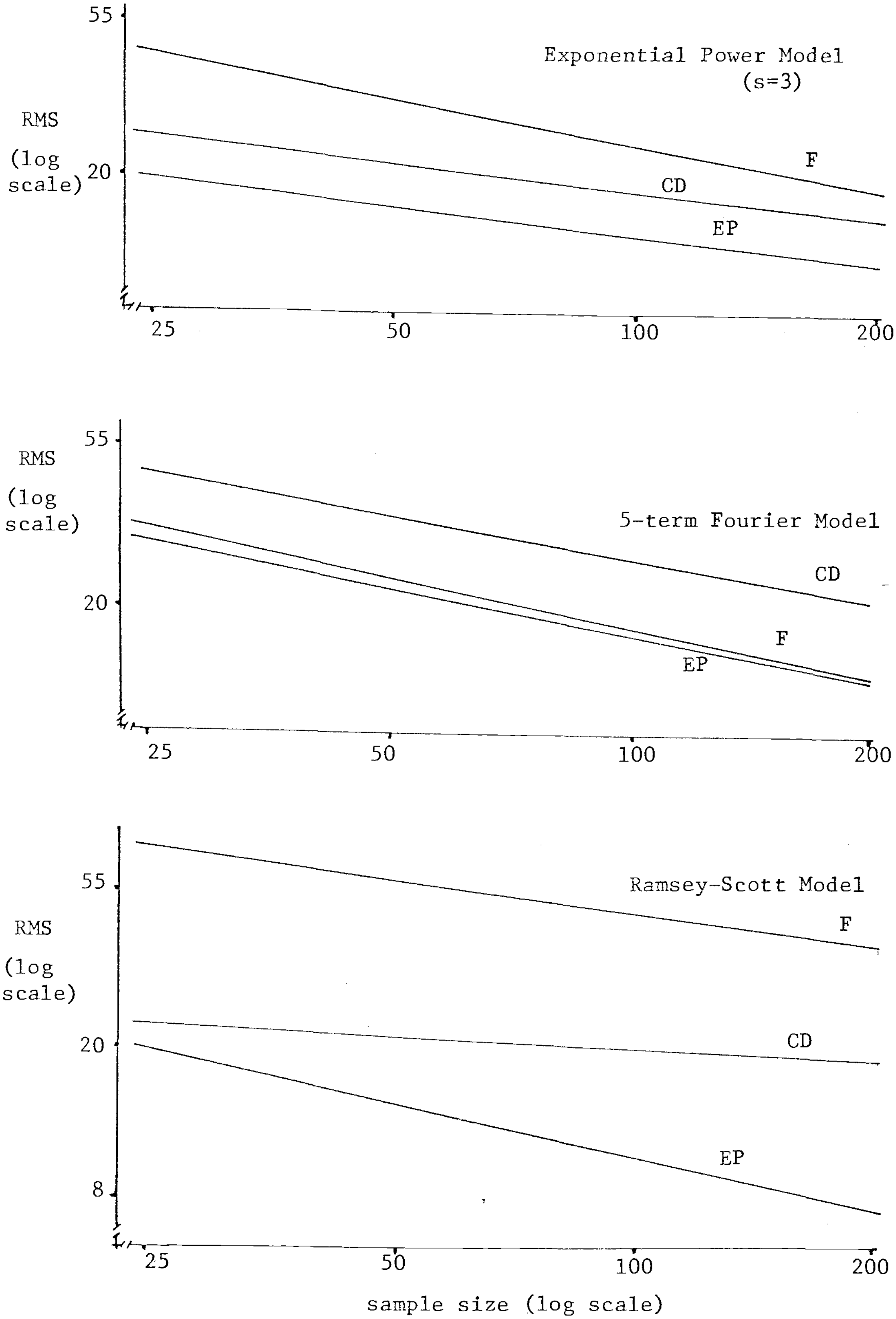


Figure 8. Log-log regression of the RMS on sample size.

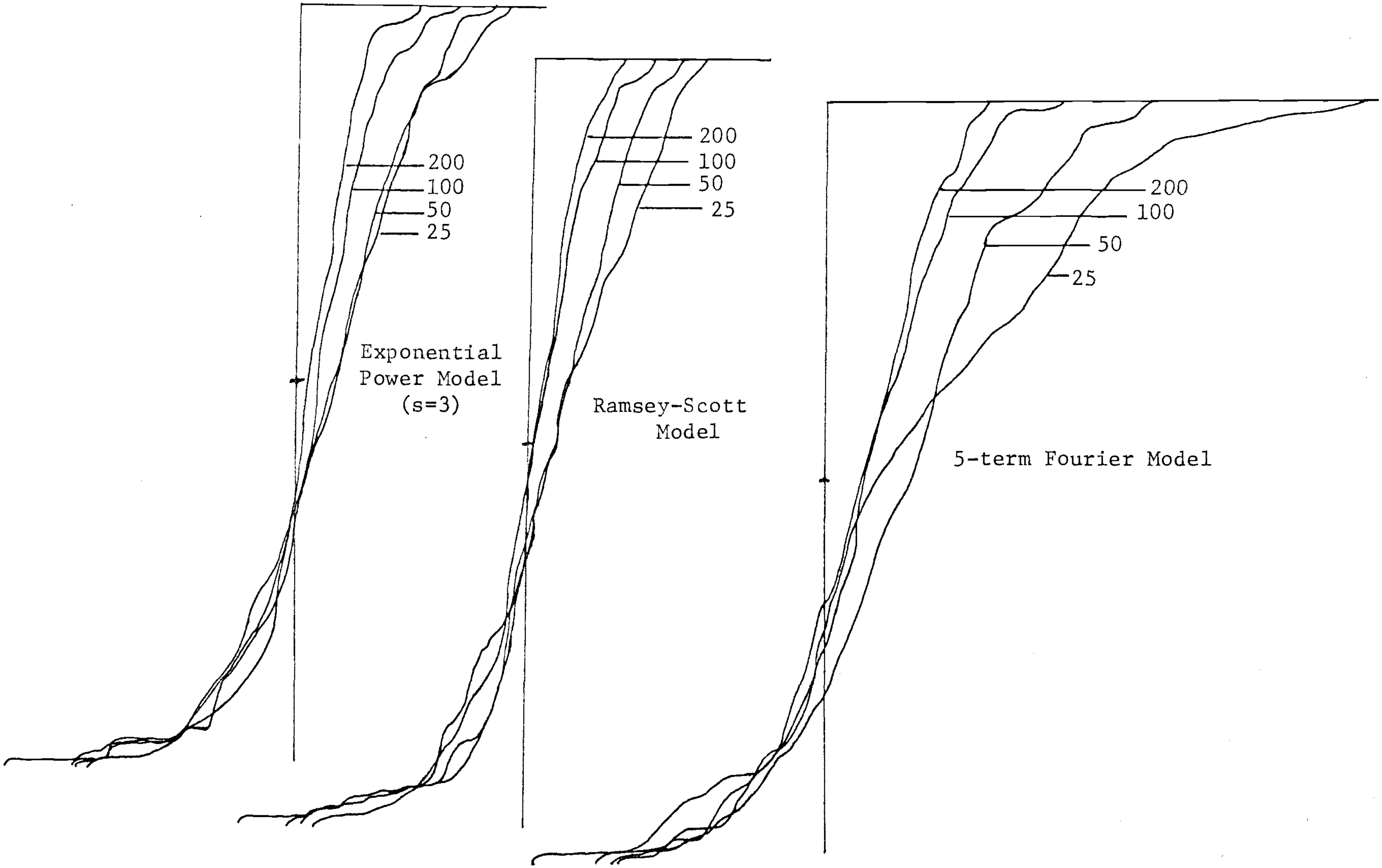


Figure 9a. Sample cdf curves for the Cum-D estimator.

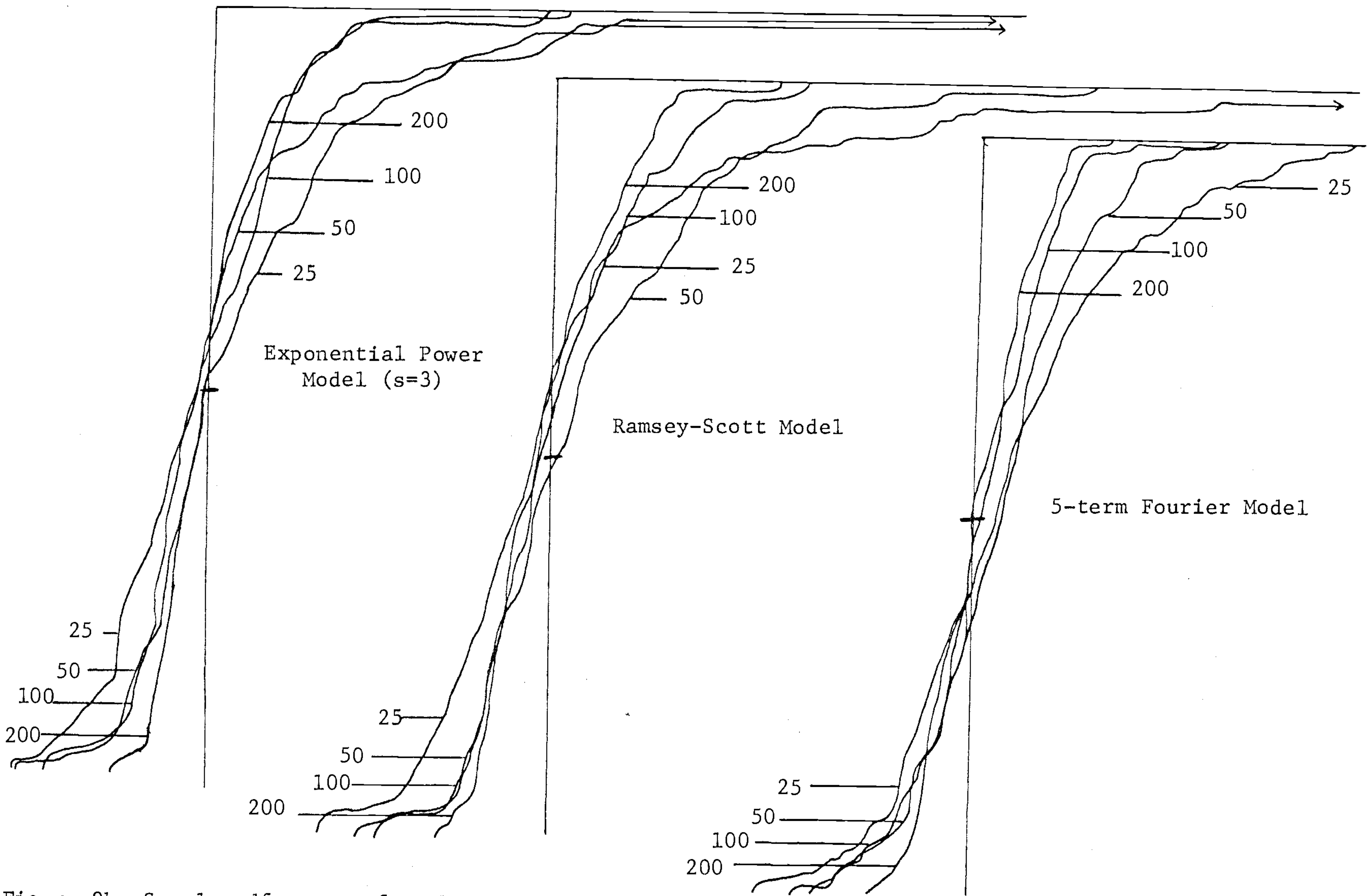


Figure 9b. Sample cdf curves for the Fourier series estimator.

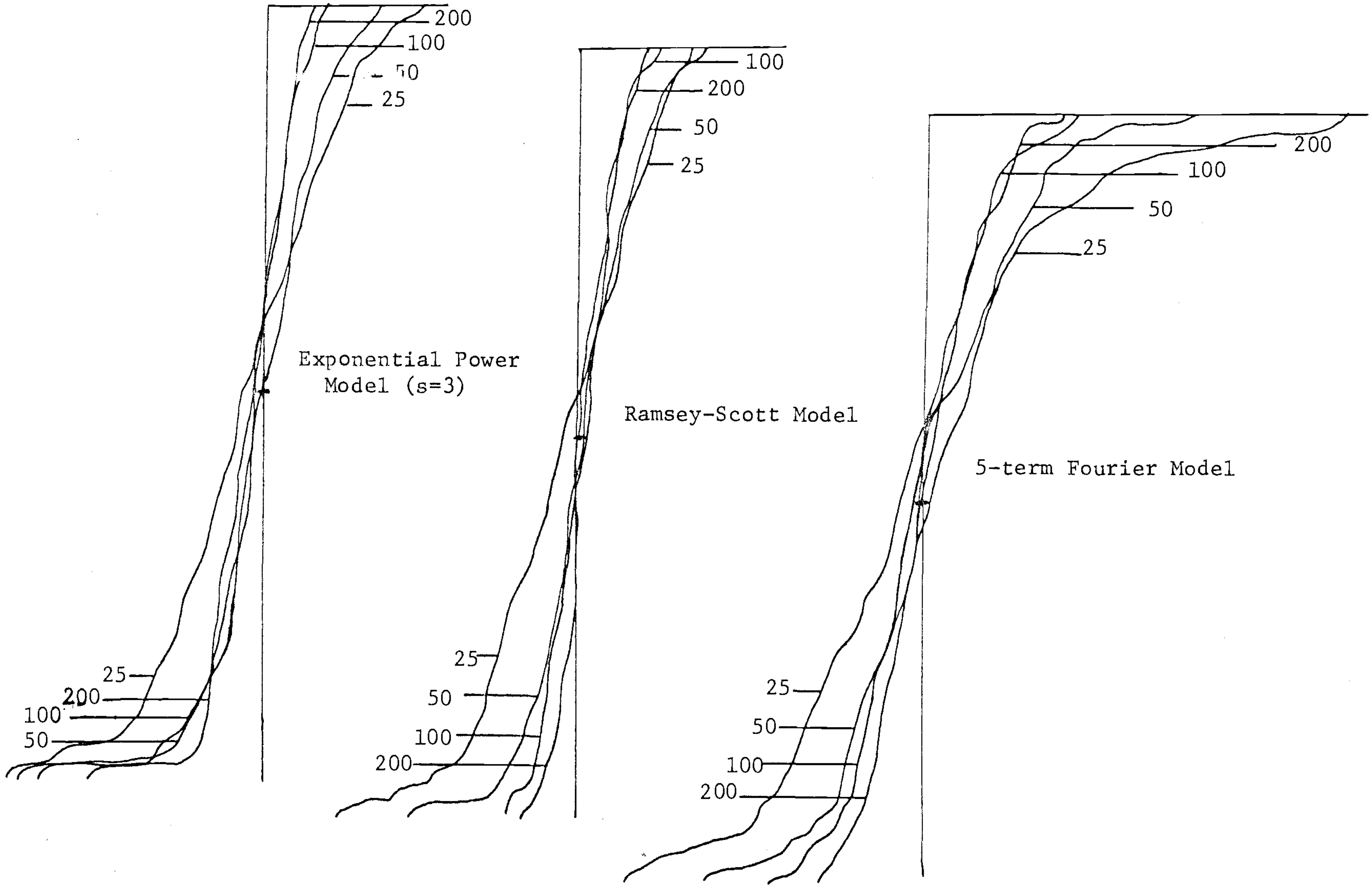


Figure 9c. Sample cdf curves for EP-MLE estimator.

Table 3. Regressions on the log sample size for the log of the ratio k/n from the Cum-D estimator.

Model	Sample Size 25	50	100	200	Regression Coefficients β_0	β_1
Ramsey-Scott	.9104	.8568	.7704	.7138	.1793	-.3222
5-term Fourier	.7464	.6814	.5709	.5062	.3451	-.1936
Exponential Power(s=3)	.8836	.8130	.7538	.6686	.3064	-.1316

One final measure of convergence we examined was that of the ratio k/n, or more precisely, the convergence rate of $|(k/n - F(a_B))|$, for the Cum-D estimator. We have shown that the convergence of k/n to $F(a_B)$ is sufficient to give the convergence of the Cum-D estimator, but no requirements on the rate of convergence are given. The Fourier series and exponential power models all have a basal area of zero so that $F(a_B)$ is zero, but the Ramsey-Scott model has a basal area a_B of 50 units so that $F(a_B)$ is 0.50. We were especially interested in determining if the rate of convergence of k/n to $F(a_B)$ would be different for models with zero basal area than for models with positive basal area, and if, when the basal area is positive, k/n converged to $F(a_B)$ or to some value less than $F(a_B)$.

The average of the ratio k/n for the 100 data sets with each sample size was calculated, and log-log regressions performed on the absolute value of the difference between k/n and $F(a_B)$; the results are shown in Table 3. The rates of convergence in this case are also model-dependent ($p = .0013$), being somewhat higher for the Ramsey-Scott model, in which k/n is converging to a positive value, than for the 5-term Fourier and exponential power ($s=3$) models, in which k/n is converging to zero. It does also appear that in the Ramsey-Scott model k/n is converging to $F(a_B)$ rather than to some value which is less than $F(a_B)$. This behavior was to be expected, although the proof of the convergence of the Cum-D estimator did not rely on this point.

One additional model was considered to examine the convergence rate of the ratio k/n to $F(a_B)$; this is the exponential model for which $F(a) = 1 - \exp(-a/A)$ and $f(a) = (1/A)\exp(-a/A)$ so that $F^{-1}(u) = -A \log(1-u)$ for u greater than or equal to zero. We considered the case in which, for a given sample size n,

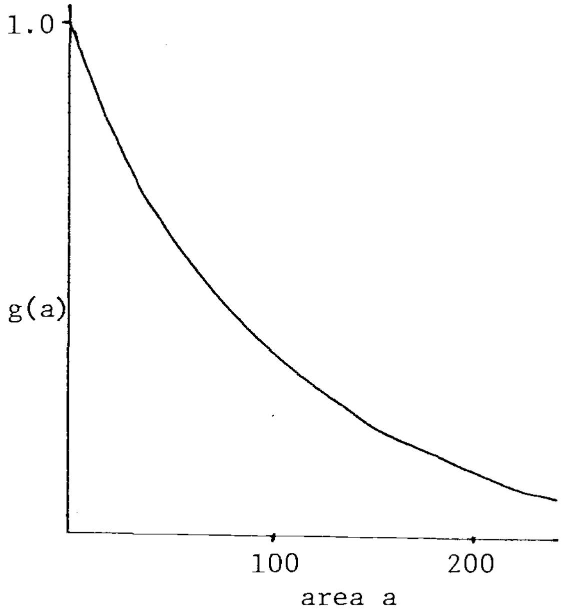
$$a_{j,n} = F^{-1}(j/(n+1)) = -A\cdot\log\left[1 - (j/(n+1))\right].$$

We then let

$$d_{m,n} = \frac{F(a_{m+1,n}) - F(a_{m,n})}{a_{m+1,n} - a_{m,n}} = nA\cdot\log\left[\frac{1 - m/(n+1)}{1-(m+1)/(n+1)}\right]$$

and note that $d_{m,n}$ is greater than $d_{m+1,n}$. Thus, the envelope function associated with the points $(a_{j,n}, j/(n+1))$ will consist of the straight line segment from the origin to the point $(a_{\sqrt{n},n}, \sqrt{n}/(n+1))$, which has slope $\{\sqrt{n}/(n+1)\}\{1/[-A\cdot\log[1-\sqrt{n}/(n+1)]]\}$, and then the straight line segments joining this and each of the consecutive points $(a_{j,n}, j/(n+1))$. The detectability curve $g(a)$ and the envelope function associated with these points are shown in Figure 10.

We now want to consider testing the hypothesis of equal slopes for the first segment of the envelope function and the segment from



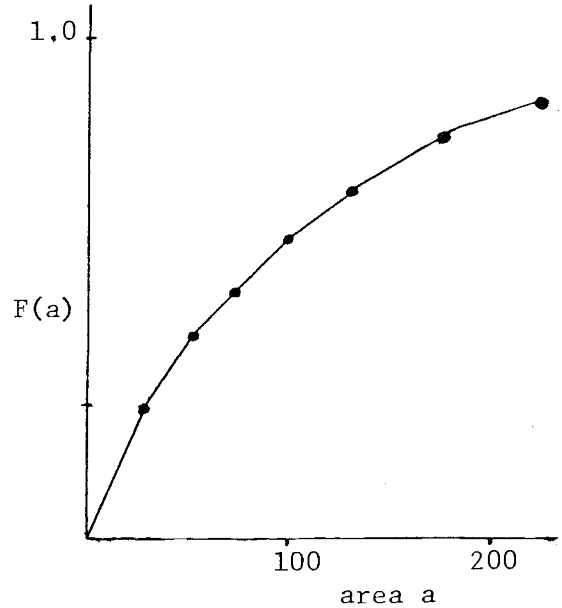


Figure 10a. The detectability curve g(a) for the exponential model.
b. The envelope function for the exponential model.

the first point of the envelope function to some point $(a_{k,n}, k/(n+1))$. From equation (1) we see that the test statistic in this case is given by

$$\Lambda = \sqrt{n}\cdot\log\left[\frac{\sqrt{n}}{a_{\sqrt{n},n}}\right] + (k-\sqrt{n})\log\left[\frac{k-\sqrt{n}}{a_{k,n}-a_{\sqrt{n},n}}\right] - k\cdot\log\left[\frac{k}{a_{k,n}}\right].$$

If we find the largest k-value for each n which would cause us to fail to reject the hypothesis of equal slopes, i.e. the k-value which would be determined by the Cum-D procedure with this envelope function, we would obtain the k-values which are shown in Table 4. Although k/n does converge to zero in this case, as we have shown that it must, it does so at a very slow rate. It is not until n is nearly 1000 that we can expect to estimate A to within 50% of its true value. This suggests that the Cum-D estimation procedure may

Table 4. Sample sizes and associated Cum-D estimator k-values for the exponential model.

n	k	k/n	$\hat{A}/A$
50	44	.8800	2.41
100	81	.8100	2.05
200	147	.7375	1.80
400	262	.6550	1.62
800	466	.5825	1.50
1,000	561	.5610	1.47
10,000	3,508	.3508	1.23
100,000	20,950	.2095	1.12
1,000,000	121,580	.1216	1.07
10,000,000	695,800	.0696	1.04
100,000,000	3,948,000	.0395	1.02
1,000,000,000	23,110,000	.0231	1.01

not be the best one to use for situation in which there is reason to expect that the detectability curve drops off very sharply from the origin. The results of the next section will further emphasize this point when we present the results of the estimations for the exponential power (s=1/3) model, since it also has this same general shape.

B. Comparisons of Bias and Variability

In addition to comparing the rates of convergence of the three estimators, we were interested in other sampling properties as well. To investigate some of these properties, 100 data sets with samples of size 50 were generated for the 2-term Fourier and exponential power (s=1/3) models; similar data sets had already been obtained for the Ramsey-Scott, the 5-term Fourier and the exponential power (s=3) models. The mean and standard deviation of the 100 estimates of effective area surveyed were calculated for each model and each estimator, along with the minimum and maximum estimates obtained for each. These results are presented in Table 5.

Table 5. Mean, standard deviation, minimum and maximum of 100 estimates of effective area surveyed A for the Cum-D, Fourier series and EP-MLE estimators. Note: The true effective area is A = 100.0 and the sample size is 50.

Estimator		Model: Ramsey-Scott	2-term Fourier	5-term Fourier	Expon. Power (s=1/3)	Expon. Power (s=3)
Cum-D	mean	104.9	128.0	119.7	391.1	106.8
	SD	18.5	35.9	29.2	181.8	22.1
	min	39.5	43.6	43.6	83.0	39.5
	max	137.5	185.9	186.4	1006.0	152.0
Fourier Series	mean	112.4	113.9	106.1	592.9	112.1
	SD	47.4	27.9	20.7	172.9	45.7
	min	50.6	49.7	52.0	287.6	48.1
	max	457.1	189.7	160.9	1257.4	382.9
EP-MLE	mean	101.0	140.4	100.6	139.9	98.6
	SD	13.5	23.5	22.4	88.2	13.9
	min	54.2	50.1	40.7	19.4	52.9
	max	127.8	172.3	166.8	510.6	128.2

An examination of Table 5 shows that the most obvious difference between the three estimators is in their ability to give accurate estimates for the exponential power (s=1/3) model. The detectability curve for this model, like that of the exponential model discussed in Section A, declines very sharply away from the origin, and only the EP-MLE estimator is flexible enough to give an average estimate of effective area surveyed that is within 50% of the true value. The Cum-D estimator, although giving an average that is almost four times the true value, does considerably better than the Fourier series estimator, which has a minimum value of 287.6 and an average that is nearly six times the true value. These problems in estimating A are also reflected in the standard deviations, which are four to eight times larger for this model than for the other models, again with the EP-MLE estimator's standard deviation only half as large as the other two. Because of the difficulties in obtaining accurate estimates for this model, it was not used for the other comparisons we made. We also eliminated the 2-term Fourier model because the detectability curve is not monotone decreasing, and because the results for this model were generally quite similar to those obtained for the 5-term Fourier model, although all three estimators did have a larger bias for this model than for the 5-term Fourier model.

Another characteristic worth noting is the tendency of the Fourier series estimator to occasionally give estimates that are very large. For the Ramsey-Scott model, the Fourier series estimator had 14 of 100 estimates which were greater than the maximum given by the Cum-D estimator and 25 of 100 estimates which were greater than the maximum given by the EP-MLE estimator. For the 5-term Fourier model, it makes a slightly better showing, having a maximum which is slightly smaller than those of the Cum-D and EP-MLE estimators. For the exponential power (s=3) model, 10 of the 100 Fourier series estimates are greater than the Cum-D maximum while 21 are greater than the EP-MLE maximum. This problem with the Fourier series estimator can be most clearly seen in Figure 11, where the sample cumulative distribution functions of the estimates are shown for each of the three models within each estimator. In all cases the Fourier series estimator has a distribution with longer tails than the Cum-D and EP-MLE estimators.

One final difference between these three estimators that is worth noting is in the computer costs of calculating the estimates of effective area surveyed. Although the EP-MLE estimator generally gives the best performance in terms of bias and variability, it required 115 CP seconds execution time for the calculation of the 500 estimates based on samples of size 50 while the Fourier series estimator required only 22 CP seconds execution time and the Cum-D estimator required only 13 CP seconds execution time for the calculation of the same number of estimates. The Cum-D estimator might therefore be recommended when there are a very large number of estimates to be calculated and cost is a major consideration.

C. Effects of Errors in Determination of Detection Areas

One of the assumptions given in Chapter I about the placement and movement of animals during the survey is that all distances to detected animals are determined without error. For many species of animals, however, it is unlikely that this assumption is satisfied

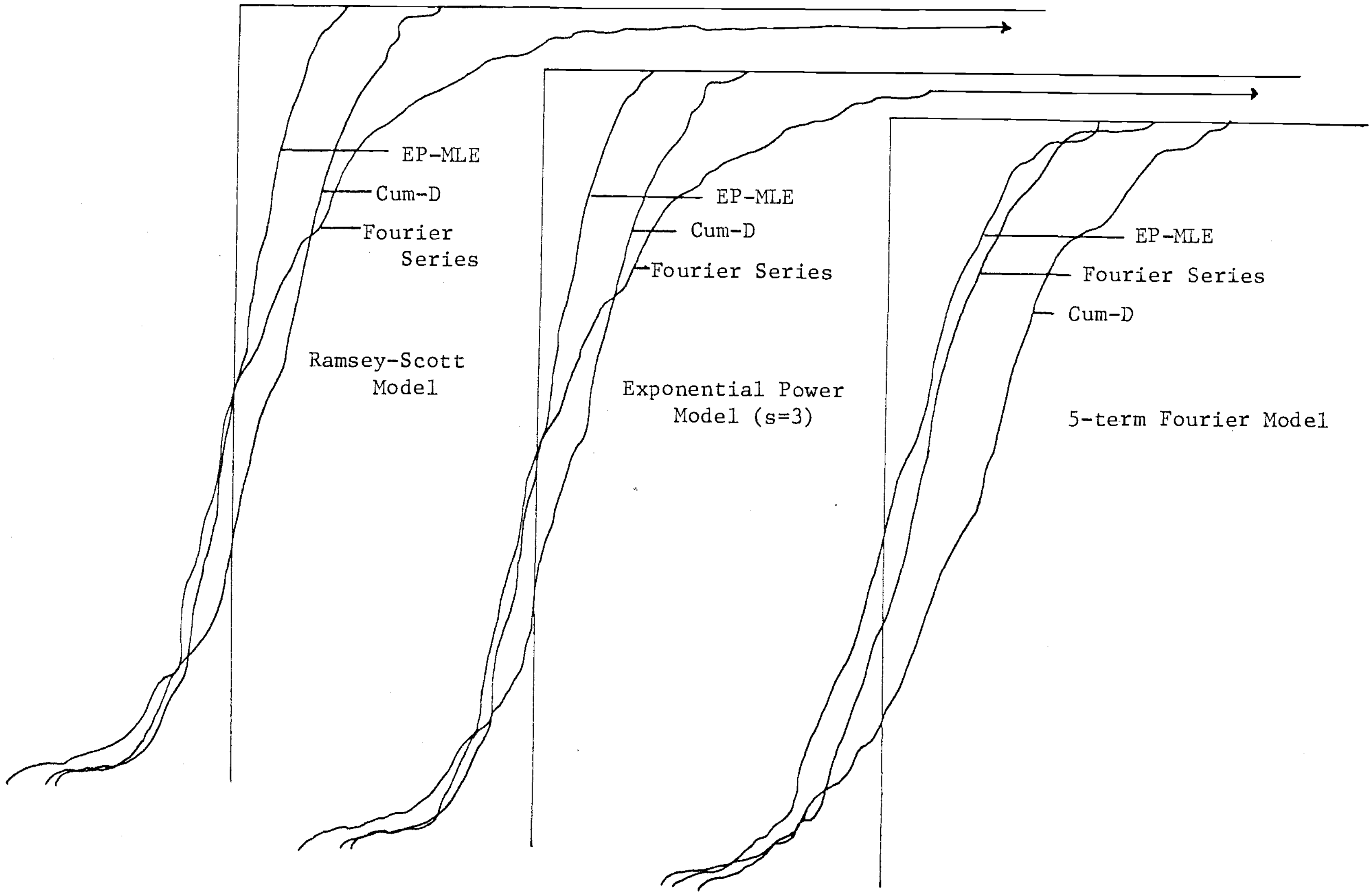


Figure 11. Sample cdf curves for Cum-D, EP-MLE and Fourier series estimators, each with n = 50.

in practice; distances are often estimated by the observer at the time of detection rather than being measured accurately. We were interested in determining what effect such errors in the determination of detection distances and areas would have on the estimates of effective area surveyed given by the three estimation procedures.

Our first approach to this problem was to consider a multiplicative model for the errors in estimated detection areas. Let Y be the true detection area for a detected anumal with $f_Y(y) = g(y)/A$ its probability density function. If we expect that small areas are estimated most accurately, with the magnitude of errors increasing as the true detection area increases, one possible model is to let the observed or estimated detection area W be given by $W = Y/X$, where X is an error random variable defined on $(0,\infty)$ with $E(X) > 0$. If X is less than one, the detection area will be overestimated; X greater than one will produce an underestimate of Y. It is easily shown that the probability density function for the observed detection area is then given by

$$f_W(w) = \int_0^\infty [f_Y(wx) \cdot f_X(x) \cdot x]\, dx$$

so that

$$f_W(0) = \int_0^\infty [f_Y(0) \cdot f_X(x) \cdot x]\, dx = E(X)/A.$$

By using the estimated detection areas we are therefore actually estimating $A/E(X)$ rather than estimating A as desired.

We now considered an error random variable X with a log-logistic distribution, i.e. $X = \exp(Z)$ where Z has a logistic distribution with $F_Z(z) = 1/\{1 + \exp(-z/b)\}$, for a parameter b between zero and one. With this error variable,

$$F_X(x) = 1/(1 + x^{-1/b}) \quad \text{and } E(X) = \Gamma(1-b)\Gamma(1+b),$$

so $A/E(X) = A/\left[\Gamma(1-b)\Gamma(1+b)\right]$. A random error was generated by the log-logistic distribution and each of the original detection areas was perturbed. The three estimation procedures were then used to estimate A. The results of these simulations are displayed in Table 6 for three different b-values as well as for the unperturbed data.

Table 6. Mean and standard deviation of estimates of effective area surveyed with log-logistic error model.

				b		
Estimator	Model		no error	.10	.25	.50
Cum-D	Ramsey-Scott	mean	104.9	103.5	102.6	86.7
		SD	18.5	18.9	24.4	23.7
	5-term Fourier	mean	119.7	117.5	113.5	93.7
		SD	29.2	30.1	29.3	31.6
	Expon. Power (s=3)	mean	106.8	106.4	102.9	83.9
		SD	22.1	20.4	23.4	23.6
Fourier Series	Ramsey-Scott	mean	112.4	108.5	94.1	82.5
		SD	47.4	50.6	22.8	18.2
	5-term Fourier	mean	106.1	105.5	100.6	93.4
		SD	20.7	21.3	19.6	22.8
	Expon. Power (s=3)	mean	112.2	220.5	98.1	81.5
		SD	45.7	53.0	61.7	19.2
EP-MLE	Ramsey-Scott	mean	101.0	95.7	83.2	49.6
		SD	13.5	13.4	18.0	18.8
	5-term Fourier	mean	100.6	96.7	82.2	48.0
		SD	22.4	21.7	20.9	20.5
	Expon. Power (s=3)	mean	98.7	95.3	80.0	41.0
		SD	13.9	14.4	17.8	19.9
		A/E(X)	100.0	98.4	90.0	63.7

From these results we can see that neither the Cum-D nor the Fourier series estimators are seriously affected by these types of errors; although they do underestimate A by approximately 10 - 15% when b is as large as 0.50, there is very little change when b is 0.25 or less. In contrast, the EP-MLE estimator gives slight underestimates when b is only 0.10 and, if b is as large as 0.50, this procedure gives estimates which are in error of the true value by more than 50%. All of the standard deviations are relatively stable, although there may be a slight tendency for them to increase as b increases. In their study, Scott and Ramsey (1981) estimated b in field tests on vocal detections of forest birds, finding it to be approximately 0.125; this suggests that the problem of errors in detection distance estimation is not a great one if a nonparametric estimator is used.

D. Grouping of Detection Areas

Another type of error model that we considered was a grouping, or heaping model such as that discussed by Burnham et al. (1980). This model represents the phenomena of recording detection distances or detection areas at certain convenient values. An example of this would be if detection distances were recorded as 0, 5, 10, 20, 50 and 100 meters only. Burnham et al. (1980) note that even when distances are measured with a tape measure, a fairly high degree of heaping at particular values still occurs. In practice, the degree of grouping often increases with the distance (or area) from the observer. To evaluate the sensitivity of each of the three estimation procedures to various types of grouping, we have used four different grouping models and applied them to the simulation data.

The proper theoretical approach to an analysis of grouped data involves the multinomial distribution. If we assume that the detection areas are recorded as being within one of the intervals $(0, c_1)$, (c_1, c_2), ... , (c_{m-1}, c_m), and record that n_k animals were detected at detection areas within the interval (c_{k-1}, c_k), where $c_o = 0$, then the multinomial cell probabilities are given by

$$p_k = \int_{c_{k-1}}^{c_k} f(y)\ dy.$$

Under the Fourier series model, these cell probabilities have closed forms given by

$$p_k = \frac{c_k - c_{k-1}}{w^*} + \sum_{k=1}^{\infty} \left[\frac{b_k w^*}{k}\right]\left[\sin\left(\frac{k \cdot c_k}{w^*}\right) - \sin\left(\frac{k \cdot c_{k-1}}{w^*}\right)\right]$$

Maximum likelihood estimates of the parameters b_1, b_2, ... , b_m for a finite number of terms m can then be obtained by iterative numerical methods. As Burnham et al. (1980) note, however, there is no longer a simple stopping rule for determining the number of terms in the model, since the estimates b_i change as each additional term is added. They propose using a general likelihood ratio test or a Chi-square goodness-of-fit test to determine the number of terms.

For the exponential power model, the problem is more complex. The cell probabilities in this case are given by

$$p_k = \int_{c_{k-1}}^{c_k} \{(1/A) \cdot \exp[-(\Gamma(1 + 1/s)(y/A))^s]\}\ dy.$$

The maximum likelihood estimates of A and s would be found by taking the partial derivatives of the multinomial likelihood function with respect to A and s and solving them simultaneously. Since the above integral does not have a closed form (unless 1/s is a integer so that successive integration by parts can be used), this involves complex iterative numerical integration procedures.

The Cum-D estimator, since it is based on the cumulative distribution function, can be used with grouped data with a slight modification. The envelope function is determined using the end-points of the intervals and the stopping rule for the determination of k is applied as usual. This avoids the problem of specifying the cell probabilities as a function of f and greatly simplifies the analysis.

Although the Fourier series and the EP-MLE estimators do have these modified forms for use with grouped data, they are so complex that in practice the procedures for ungrouped data are used even when the data is grouped. We were interested in determining what effect this would have on the estimates of effective area surveyed given by the different procedures. The simulation data from the Ramsey-Scott, the 5-term Fourier and the exponential power (s=3) models for samples of size 50 were grouped according to the four methods described below and the midpoints of the intervals used as the data to estimate effective area surveyed.

(1) Uniform grouping by area - moderate degree. For this grouping model, we defined the cutoff points of the intervals by increments of 10 units, i.e. $c_j = 10j$, $j = 0, 1, \ldots, 30$. The

detection areas will therefore be classified into fifteen categories for the Ramsey-Scott model and thirty categories for the 5-term Fourier and exponential power (s=3) models, with all detection areas greater than 290 going into the last class for the exponential power model.

(2) Uniform grouping by area - severe degree. For this grouping model the cutoff points of the intervals were determined by $c_j = 30j$, $j = 0, 1, \ldots, 10$. The Ramsey-Scott model will then have only three distinct data values while the 5-term Fourier and exponential power (s=3) models will have only ten distinct data values.

(3) Progressive grouping by area. For this grouping model, the degree of grouping is dependent on the magnitude of the detection area so that the lengths of the intervals become successively longer. The cutoff points used were:

$c_o = 0.0$, $c_1 = 1.0$, $c_2 = 3.0$, $c_3 = 5.0$, $c_4 = 7.0$, $c_5 = 9.0$, $c_6 = 15.0$, $c_7 = 21.0$, $c_8 = 27.0$, $c_9 = 33.0$, $c_{10} = 45.0$, $c_{11} = 57.0$, $c_{12} = 77.0$, $c_{13} = 97.0$, $c_{14} = 147.0$, $c_{15} = 197.0$, $c_{16} = 247.0$ and $c_{17} = 297.0$.

(4) Grouping by detection distances. For this model we treated our simulated data as having been generated from the detection distances in a variable circular plot survey. The distances were grouped by uniform increments of one unit, resulting in the following intervals for the detection areas;

$c_o = 0.0$	$c_1 = 0.79$	$c_2 = 7.07$	$c_3 = 19.63$
$c_4 = 38.48$	$c_5 = 63.62$	$c_6 = 95.03$	$c_7 = 132.73$
$c_8 = 176.71$	$c_9 = 226.98$	$c_{10} = 283.53$	$c_{11} = 300.0.$

The results of the simulations are shown in Table 7, along with the values for the ungrouped data.

Table 7. Bias and standard deviation of estimates of effective area surveyed for the four types of grouped data.

Estim.	Model		Uniform 10 area units	Uniform 30 area units	Grouping by distance	Progress. grouping by area	No Group.
			Grouping Method				
Cum-D	Ramsey-Scott	Bias	8.47	10.55	10.56	4.12	4.9
		SD	15.52	15.93	15.64	17.20	18.5
	5-term Fourier	Bias	26.34	28.71	30.13	21.30	19.7
		SD	29.20	27.52	27.64	28.13	29.2
	Expon. Power (s=3)	Bias	13.37	15.03	13.60	8.04	6.8
		SD	18.97	17.62	18.48	16.94	22.1
Fourier Series	Ramsey-Scott	Bias	15.03	1.97*	10.62	13.01*	12.4
		SD	56.46	36.72*	52.91	54.29*	47.4
	5-term Fourier	Bias	5.52	5.12	5.23	5.12*	6.1
		SD	20.96	20.57	19.67	24.22*	20.7
	Expon. Power (s=3)	Bias	12.37	1.81	9.03*	13.51*	12.2
		SD	61.17	36.73	41.92*	73.33*	45.7
EP-MLE	Ramsey-Scott	Bias	0.91	-0.19	0.85	2.70	1.0
		SD	13.30	13.08	12.95	15.01	13.5
	5-term Fourier	Bias	0.84	3.21	0.61	1.64	0.6
		SD	22.27	21.07	23.23	22.32	22.4
	Expon. Power (s=3)	Bias	-0.91	-1.36	-2.09	0.35	-1.3
		SD	13.98	14.45	13.82	15.25	13.9

* These values were calculated after omitting one value that was larger than 10,000,000.

Rather surprisingly, none of the grouping methods seems to have much of an effect on the bias and variability of the EP-MLE estimator. There is an increase in the variability of the Fourier series estimates for the Ramsey-Scott and exponential power (s=3) models but the bias appears relatively unchanged. The Cum-D estimator may produce estimates with slightly smaller variances and slightly larger biases. Overall, however, none of the estimators experience severe difficulty in any of the grouping model situations, suggesting that any one of the estimators could be used when the detection areas are recorded categorically.

VI. DISCUSSION AND CONCLUSIONS

A. Choice of w* for the Fourier Series Estimator

Although the Fourier series estimator involves relatively simple calculations, there are some problems associated with its use that are often overlooked. One of the major problems is in the choice of w*. This is often interpreted as a "remote horizon" beyond which no detections are made, but there is some danger in choosing w* to be an arbitrarily large value. Because w* is the half-period of the fundamental harmonic in the Fourier series representation of the density function f, the larger that w* is the higher frequency the function appears to be and the greater the number of terms required by the Fourier series expansion to give an adequate representation. The temptation to choose w* as a large upper limit on the detection areas should therefore be avoided as severe overestimates of effective area surveyed can result.

If w* is determined empirically from the data, similar problems can arise. Burnham et al. (1980) found that using a w* value which eliminated from one to three percent of the data worked well, but if the theoretical distribution has a very long right tail, this sample percentile may still be quite large and produce the same problems that a large fixed w* does. This may explain why some of the Fourier series estimates for the exponential power ($s=1/3$) model are so large.

Another problem associated with the use of the Fourier series estimator is the possibility of obtaining a negative value for the estimate of effective area surveyed. We had one occurrence of this problem with a simulated data set of sample size 25, which may suggest that it is primarily a small-sample problem. Fortunately, the error in such an estimate is immediately obvious to the researcher, while the error in a very large overestimate may not be.

B. Estimation of s for the Exponential Power Model

The EP-MLE estimation procedure is considerably more costly than either of the nonparametric procedures, and approximately 90% of this cost is associated with estimating the shape parameter s. Our estimates were accurate to within 0.01 units of the value which would give an exact solution to the likelihood equation and it is possible that using less stringent requirements for the accuracy of $\hat{s}$ would greatly reduce the costs of computation.

Another question of interest with respect to the estimation of s involves the difference between the estimates that would be obtained by using the EP-MLE estimator, which is based on a maximal invariant statistic under scale changes, and those that would result from the standard maximum likelihood method described in Chapter II. To examine this we considered the two exponential power models and estimated effective area surveyed by each of these methods, as well as by the maximum likelihood procedure using a known value of s. The results of these analyses are displayed in Table 8.

It is clear from these results that the procedure based on a maximal invariant statistic under scale changes gives slightly better estimates than does the standard maximum likelihood procedure,

Table 8. Bias and standard deviation of estimates of effective area surveyed using the EP-MLE procedure, the standard maximum likelihood procedure and the maximum likelihood procedure with known shape parameter s.

Model		EP-MLE	MLE	Known s
Exponential Power (s=3)	Bias	-1.4	-2.1	2.0
	SD	13.9	22.3	10.0
	$\hat{s}$	3.3	3.6	-
Exponential Power(s=1/3)	Bias	39.9	46.2	3.4
	SD	88.2	94.5	23.7
	$\hat{s}$	0.37	0.37	-

and that for the model which declines sharply from one at the origin, almost all of the bias and variability in the estimates of effective area surveyed are a result of needing to estimate s.

C. The Cum-D and Other Stopping Rules

The Cum-D estimation procedure is one of a general class of estimators of the form $\hat{A} = na_k/k$, where k is determined by some stopping rule for the inclusion of the area a_k into the basal area. One of the first such stopping rules was proposed by Emlen (1971) who suggested that the histogram of the observed detection distances be plotted and "inspected" to find a point where the density starts declining. This procedure is quite subjective, however, and is difficult to apply when a large number of estimates of effective area surveyed are required.

Reynolds et al. (1980) also used a plot of the histogram of observed detection distances to estimate the basal area, finding a distance d_k such that the average density of animals in all of the histogram classes below d_k was greater than twice the density in the next histogram class. This procedure, while less subjective than Emlen's(1971), doesn't take into account the sample size and the ability to detect even small changes in density of animals with large amounts of data.

Another estimator of this form is the isotonic regression estimator (Barlow et al., 1972). This procedure uses the maximum slope of a straight line from the origin to a point on the sample cumulative distribution function to determine k; this is identical to the first step of the Cum-D procedure without the $\sqrt{n}$ rule. The problem with this estimator is that it tends to select k-values which are too small, and the known erratic local behavior of the Brownian motion process suggests that the sequence of k-values produced may not increase quickly enough with n to provide a consistent estimator (Cramér and Leadbetter, 1967).

The estimation procedure proposed by Ramsey and Scott (1979) is the most similar to the Cum-D procedure. It is also based on the cumulative distribution function and its corresponding envelope function, determining k by the stopping rule $k = \min\{j;$ slope from the origin to $(a_{j_r}, j_r/n)$ on the envelope function is significantly different from the slope joining $(a_{j_{r+1}}, j_{r+1}/n)\}$, where significance is also determined by the generalized likelihood ratio test for Poisson scattering of animals in space. To see how this stopping rule compared to the Cum-D stopping rule, we used it on the exponential model discussed in Chapter V, since it was for this type of model we felt that the Ramsey and Scott (1979) procedure might have difficulty. We again calculated the value of k which would be selected by the estimation procedure for each of the sample sizes n between 50 and 1,000,000,000 shown in Table 4, and found that k/n converged to approximately 0.919 rather than to zero, suggesting that with a detectability curve such as this one the procedure may not give consistent estimates. When this stopping rule was applied to the simulation data for the Ramsey-Scott, the 5-term Fourier and the exponential power (s=3) models with samples of size 50, the results were uniformly worse than those of the Cum-D procedure, but not by very much; the average increase in the bias was only 3.45 units while the average increase in the standard deviation was 0.33 units.

D. Donuts and the Modification of the Cum-D Estimator

The assumption that there is perfect detectability at the origin may be quite reasonable in some situations but there are many in which it is suspect or obviously false. In surveys of some species of birds, for example, it is only the males defending territories which sing and move about so as to be detectable to an observer. In this situation, the shape of the histogram of the observed detection areas and the empirical cumulative distribution function may provide little evidence of the violation of the assumption of perfect detectability at the origin. What will be obtained with any of the estimation procedures we have discussed is an estimate of the density of detectable animals rather than the density of all animals, and it is only with the use of auxilliary information about the proportion of the population that is detectable that a researcher can modify the estimate of effective area surveyed to give a valid estimate of the total population size or density.

A second type of violation of the assumption of perfect detectability at the origin arises when animals react to the presence of an observer, either by making themselves completely undetectable or by moving away from the observer to points further away where they then become detectable. In this case, the histogram or cumulative distribution function may give an indication of the violation of the assumption of perfect detectability at the origin by exhibiting what we have termed a "donut". Such a detectability curve and cumulative distribution function are illustrated in Figure 12. If

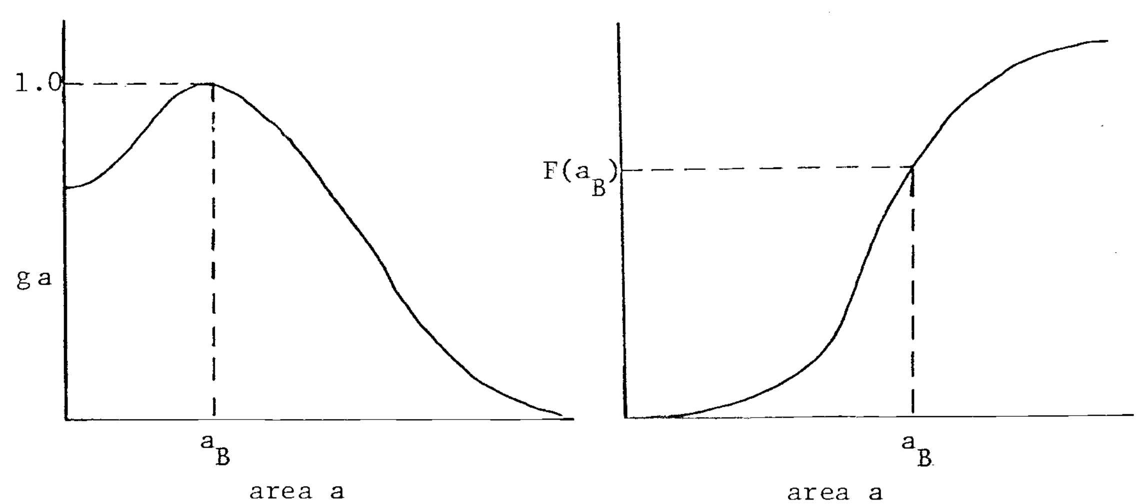


Figure 12. Donut-type detectability curve and cumulative distribution function.

the maximum height of the detectability curve is still one, such as when there is a positive basal area a_B and the range of the disturbance of the animals is within the basal area, an estimate of effective area surveyed may still be obtained, although the procedure for the two types of avoidance behavior will be different, as the second type will require a modification of the usual Cum-D procedure.

First, for the case when animals react to the observer by moving away to points still within the basal area and then becoming detectable, the total number of detections will remain the same, as illustrated in Figure 13a. In this situation we want the estimate of effective area surveyed to remain the same as if there was no disturbance; this requires us to estimate the maximum slope of the cumulative distribution function at a_B. This is easily done by using the maximum slope of a line from the origin to a point on the cumulative distribution function, as shown in Figure 13a.

For the second case, in which animals react to the presence of an observer by becoming completely undetectable near the origin, the total number of detections will be smaller than if there had been no observer avoidance and the estimate of effective area surveyed must correspondingly be reduced. This estimate of A is obtained by finding the maximum slope of the cumulative distribution function, not necessarily through the origin, and extending extending a straight line with this slope to both of the horizontal axes, as shown in Figure 13b. The region between the points on the horizontal axis is then the basal area which should be used for estimating density with the number of animals detected.

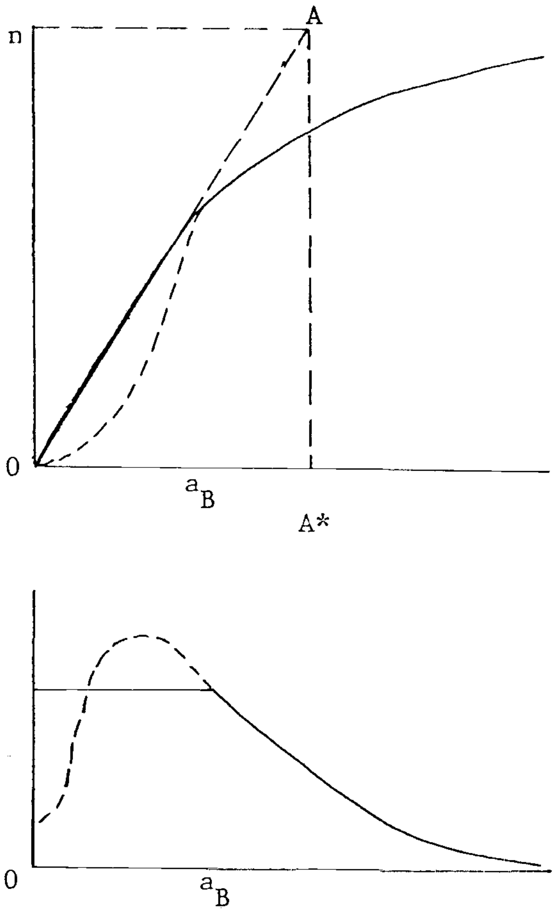


Figure 13a. Cumulative distribution function and detectability curve when effective area surveyed with observer avoidance (A*) equals A.

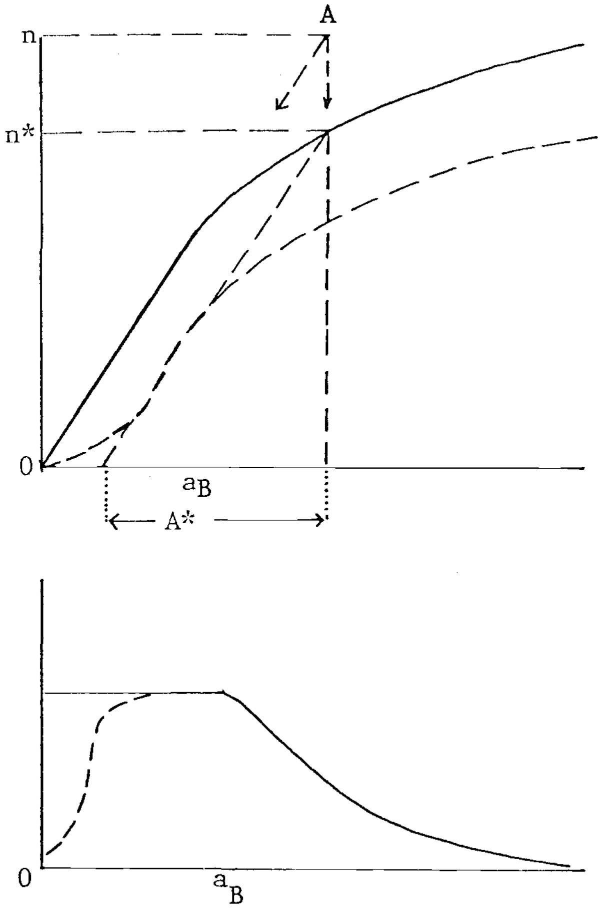


Figure 13b. Cumulative distribution function and detectability curve when effective area surveyed with observer avoidance (A*) is less than A.

It should be noted that in Figures 13a and 13b, the cumulative distribution functions have been scaled to the total number of detections, n, rather than to the fraction of detections, so as to emphasize that in case b the total number of detected animals is smaller. If both of the dashed cumulative distribution functions had been scaled to the fraction of the total detections, they would be identical, although the estimates of effective area surveyed should be different. It is again the case that only a knowledge of the behavioral characteristics of the animals being surveyed will enable the researcher to decide which of these avoidance curves and corresponding modifications to the Cum-D procedure should be used to provide the estimate of A.

As an example of this type of avoidance behavior and the estimates of effective area surveyed that would be obtained by using each of the three estimation procedures, we have shown the cumulative distribution function for one species of bird studied by the U.S Fish and Wildlife Service during the Hawaiian Forest Bird Survey. The bird is the Omao, or Hawaiian Thrush, and the cumulative distribution function for 263 detections is shown in Figure 14. As we can see, the Cum-D estimator (in unmodified form) gives an estimate of effective area surveyed which would correspond to the type of avoidance behavior illustrated in Figure 13a, i.e. that of the birds moving away and then becoming detectable. The EP-MLE estimator gives an estimate which is nearly identical to that which we would obtain by assuming that the animals react to the observer by becoming completely undetectable and applying the modified Cum-D procedure. The Fourier series estimator gives an estimate which is very much larger than the others, corresponding to the slope of the empirical cumulative distribution function very near the origin. This procedure clearly cannot be expected to give reliable estimates of A for the situation in which the density function f(a) is still given by g(a)/A but in which g(0) is not equal to one. It is therefore extremely important to plot the cumulative distribution functions or the detectability

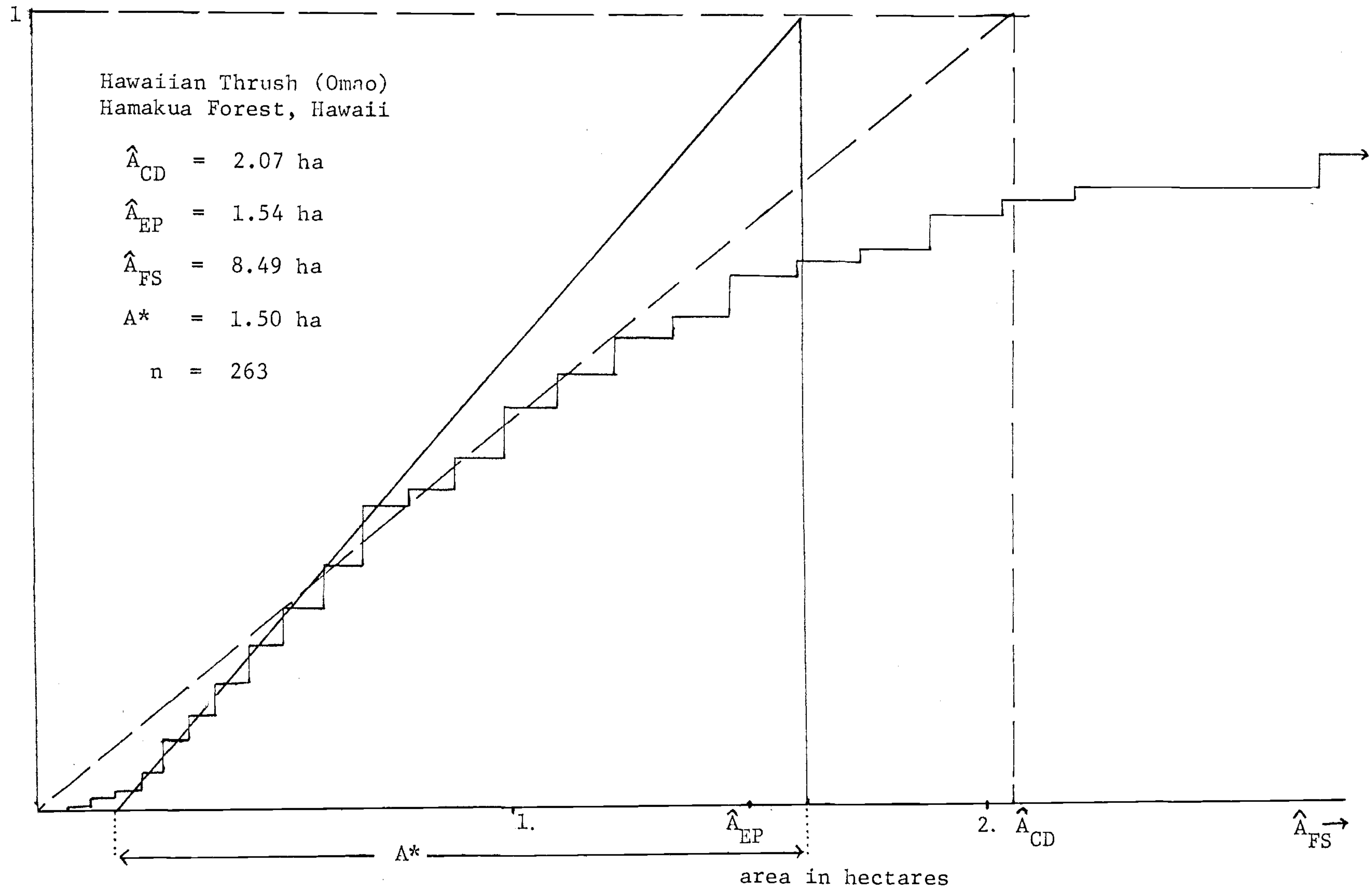


Figure 14. Sample cumulative distribution function for the Omao with estimates of effective area surveyed.

curves (i.e. density curves or histograms) before routinely using the Fourier series estimator to get estimates of effective area surveyed with a species of animal which may exhibit either of these types of avoidance behavior.

E. Conclusions

We can draw three major conclusions from the results of these analyses. The first conclusion is that the Cum-D estimation procedure is a viable alternative to the other nonparametric procedure, the Fourier series estimator. The Cum-D estimator has as good or better sampling properties in almost every situation and is less sensitive to the various departures from assumptions that we have considered. The second conclusion that we can draw is that the the maximum likelihood estimation procedure and the exponential power family of detectability curves has enough flexibility to adequately model most of the detectability curves that we normally envision using for animal surveys. The cost of using the nonparametric procedures was clear in several instances when the EP-MLE estimator had smaller bias and standard deviation and faster rates of convergence than did either of the other two procedures, even on the Ramsey-Scott and Fourier models. The lack of sensitivity to most of the violations of assumptions we considered also suggests that the EP-MLE estimator has wider application than previously expected.

The final conclusion that we can draw from these results is that there is a great deal of value in plotting the sample cumulative distribution functions and their associated estimates of effective area surveyed, regardless of which estimation procedure is being used. It is only by such visual inspections that we can really determine whether a parametric or nonparametric estimator would give the best results, and evaluate the reasonableness of the estimates provided by the various estimators. Although this may be very difficult to do in large surveys such as the Hawaiian Forest Bird Survey, the advantages should not be overlooked whenever possible.

BIBLIOGRAPHY

Andersen, E. B. 1973. Conditional Inference and Models for Measuring. Copenhagen, Mentalhygiejnisk Forlag.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and H. D. Brunk. 1972. Statistical Inference Under Order Restrictions. John Wiley & Sons, Inc., New York.

Burnham, K. P. and D. R. Anderson. 1976. Mathematical models for nonparametric inferences from line transect data. Biometrics 32:325-336.

Burnham, K. P., Anderson, D. R. and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. Wildlife monograph no. 72. Suppl. to Journal of Wildlife Manag. 44.

Crain, B. R., Burnham, K. P., Anderson, D. R. and J. L. Laake. 1979. Nonparametric estimation of population density for line transect sampling using Fourier series. Biometrical Journal 21:731-748.

Cramér, H. and M. R. Leadbetter. 1967. Stationary and Related Stochastic Processes. John Wiley & Sons, Inc., New York.

Durbin, J. 1971. Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. Journal of Appl. Prob. 8:431-453.

Emlen, J. T. 1971. Population densities of birds derived from transect counts. Auk 88:332-342.

IMSL Library Reference Manual. 1980. Edition 8, Vol. 2. IMSL, Inc.

Lewis, P. A. W., Goodman, A. S. and J. M. Miller. 1969. Pseudorandom number generator for the system/360. IBM Systems Journal 8(2):136-146.

Mood, A. M., Graybill, F. A. and D. C. Boes. 1974. Introduction to the Theory of Statistics. McGraw-Hill, Inc., New York.

Patil, G. P., Taillie, C. and R. L. Wigley. 1979. Transect sampling methods and their application to the deep-sea Red Crab. pp 51-76 in J. Cairns, Jr., G. P. Patil and W. E. Waters (eds), Environmental Biomonitering, Assessment, Prediction and Management. Stat. Ecol. Ser., Vol. 11. Int. Co-op. Publ. House. Fairland, Md.

Pollock, K. H. 1978. A family of density estimators for line transect sampling. Biometrics 34:475-478.

Ramsey, F. L. 1979. Parametric models for line transect surveys. Biometrika 66:505-512.

Ramsey, F. L. and J. M. Scott. 1979. Estimating population densities from variable circular plot surveys. pp. 155-181 in R. M. Cormack, G. P. Patil and D. S. Robson (eds), Sampling Biological Populations. Stat. Ecol. Ser., Vol. 5. Int. Co-op Publ. House. Fairland, Md.

Ramsey, F. L. and J. M. Scott. 1981a. Analysis of bird survey data using a modification of Emlen's methods. pp. 483-487 in C. J. Ralph and J. M. Scott (eds), Estimating the Numbers of Terrestrial Birds. Stud. Avian Biol. 6.

Ramsey, F. L. and J. M. Scott. 1981b. Tests of hearing ability. pp. 341-345 in C. J. Ralph and J. M. Scott (eds), Estimating the Numbers of Terrestrial Birds. Stud. Avian Biol. 6.

Reynolds, R. T., Scott, J. M. and R. A. Nussbaum. 1980. A variable circular-plot method for estimating bird numbers. Condor 82: 309-313.

Scott. J. M. and F. L. Ramsey. 1981. Distance estimation as a variable in estimating bird numbers from vocalizations. pp. 334-340 in C. J. Ralph and J. M. Scott (eds), Estimating the Numbers of Terrestrial Birds. Stud. Avian Biol. 6.

Seber, G. A. F. 1973. The Estimation of Animal Abundance. Hafner, New York, and Griffin, London.

APPENDIX

APPENDIX A

Determination of Order of Remainder Terms in Taylor Series Expansion of $F^{-1}(U_k)$.

Using a Taylor series expansion of $F^{-1}(U_k)$ about $E(U_k)=k/(n+1)$, we can write

$$a_k = F^{-1}(U_k) = F^{-1}\left(\frac{k}{n+1}\right) + \left[\frac{d}{du} F^{-1}(u)\Big|_{u=\frac{k}{n+1}}\right]\left[U_k - \frac{k}{n+1}\right]$$

$$+ \left[\frac{d^2}{du^2} F^{-1}(u)\Big|_{u=\frac{k}{n+1}}\right]\frac{\left[U_k - \frac{k}{n+1}\right]^2}{2!} + \left[\frac{\left[U_k - \frac{k}{n+1}\right]^2}{2!}\cdot\right.$$

$$\left.\left\{\left[\frac{d^2}{du^2} F^{-1}(u)\Big|_{u=\xi}\right] - \left[\frac{d^2}{du^2} F^{-1}(u)\Big|_{u=\frac{k}{n+1}}\right]\right\}\right]$$

for some ξ between U_k and $k/(n+1)$. Since the second derivative of F^{-1} is given by

$$\frac{d^2}{du^2} F^{-1}(u) = \frac{-f'(F^{-1}(u))}{\left[f(F^{-1}(u))\right]^3} ,$$

we see that

$$E(a_k) = F^{-1}(k/(n+1)) + \frac{f'\left[F^{-1}(k/(n+1))\right]}{\left\{f\left[F^{-1}(k/(n+1))\right]\right\}^3} \cdot \mathrm{Var}(U_k)$$

$$+ E\left\{\left[\frac{-f'(F^{-1}(\xi))}{\left[f(F^{-1}(\xi))\right]^3} + \frac{f'\left[F^{-1}(k/(n+1))\right]}{\left\{f\left[F^{-1}(k/(n+1))\right]\right\}^3}\right]\cdot\frac{\left[U_k - \frac{k}{n+1}\right]^2}{2!}\right\}.$$

Applying the Cauchy-Schwartz inequality, we can then see that

$$\left| E\left\{\left[\frac{-f'(F^{-1}(\xi))}{[f(F^{-1}(\xi))]^3} + \frac{f'[F^{-1}(k/(n+1))]}{\{f[F^{-1}(k/(n+1))]\}^3}\right]\cdot\left[\frac{\left[U_k - \frac{k}{n+1}\right]^2}{2!}\right]\right\}\right| \leq$$

$$\sqrt{E\left\{\left[\frac{1}{4}\right]\left[\frac{-f'(F^{-1}(\xi))}{[f(F^{-1}(\xi))]^3} + \frac{f'[F^{-1}(k/(n+1))]}{\{f[F^{-1}(k/(n+1))]\}^3}\right]^2\right\} E\left\{\left[U_k - \frac{k}{n+1}\right]^4\right\}}$$

Because $U_k \xrightarrow[P]{} F(a_B)$, and f' is everywhere bounded and continuous at a_B,

$$\lim_{n\to\infty, k\to\infty, k/n \to F(a_B)} \sqrt{E\left\{\left[\frac{1}{4}\right]\cdot\left[\frac{-f'(F^{-1}(\xi))}{[f(F^{-1}(\xi))]^3} + \frac{f'[F^{-1}(k/(n+1))]}{\{f[F^{-1}(k/(n+1))]\}^3}\right]^2\right\}}$$

is zero, i.e. this term is $o(1)$. And, since

$$E(U_k^{\,j}) = \frac{(k)(k+1)(k+2)\ldots(k+j-1)}{(n+1)(n+2)\ldots(n+j)} ,$$

$$E\left[\left(U_k - \frac{k}{n+1}\right)^4\right] = \frac{k(k+1)(k+2)(k+3)}{(n+1)(n+2)(n+3)(n+4)} - \frac{4k^2(k+1)(k+2)}{(n+1)^2(n+2)(n+3)}$$

$$+ \frac{6k^3(k+1)}{(n+1)^3(n+2)} - \frac{4k^4}{(n+1)^4} + \frac{k^4}{(n+1)^4}$$

$$= \left[\frac{k}{n+1}\right]\cdot\left[\frac{1}{n+2}\right]\left(\left[\frac{k}{n+1}\right]^3\left[\frac{3(n-5)}{(n+3)(n+4)}\right] - \left[\frac{k}{n+1}\right]^2\left[\frac{6(n-5)}{(n+3)(n+4)}\right] + \right.$$

$$\left. \left[\frac{k}{n+1}\right]\left[\frac{6(n-7)}{(n+3)(n+4)}\right] + \frac{6}{(n+3)(n+4)}\right)$$

From this we see that

$$\lim_{n\to\infty, k\to\infty, k/n \to F(a_B)} \left\{\frac{\sqrt{E\left[\left(U_k - \frac{k}{n+1}\right)^4\right]}}{k/n^2}\right\} = 6$$

which implies that the remainder term is $o(1)\cdot O(k/n^2) = o(k/n^2)$.

To find the order of the remainder term for the variance, we proceed similarly, writing $a_k = F^{-1}(k/(n+1)) + T + R$, where

$$T = \frac{U_k - \frac{k}{n+1}}{f\{F^{-1}(k/(n+1))\}} \quad \text{and} \quad R = \frac{\left[U_k - \frac{k}{n+1}\right]^2}{2} \cdot \frac{-f'[F^{-1}(\xi)]}{\{f[F^{-1}(\xi)]\}^3}$$

for some ξ between U_k and $k/(n+1)$. Then $\mathrm{Var}(a_k) = \mathrm{Var}(T + R)$ and $\mathrm{Var}(T + R) = \mathrm{Var}(T) + \mathrm{Var}(R) + 2\mathrm{Cov}(T,R)$.

$$\mathrm{Var}(T) = \mathrm{Var}\left[U_k - \frac{k}{n+1}\right] \Big/ \{f[F^{-1}(k/(n+1))]\}^2 ,$$

$$\mathrm{Var}(R) \le E\left[\left(\left[U_k - \frac{k}{n+1}\right]^4 \cdot \left[-f'[F^{-1}(\xi)] \big/ \{f[F^{-1}(\xi)]\}^3\right]\right)^2\right]$$

$$\le M \cdot E\left\{\left[U_k - \frac{k}{n+1}\right]^4\right\} = O(k^2/n^4) \quad \text{for some } M,$$

since we have assumed that f' is everywhere bounded. Also,

$|\mathrm{Cov}(T,R)| = |E(TR)| \le \{E(T^2) \cdot E(R^2)\}^{\frac{1}{2}}$ by the Cauchy-Schwartz inequality. But $E(T^2) = \mathrm{Var}(T) = O(k/n^2)$ and $E(R^2) = O(k^2/n^4)$ as was shown before, so $|\mathrm{Cov}(T,R)| = O(k^{3/2}/n^3)$.

Combining these results gives

$\mathrm{Var}(T + R) = \mathrm{Var}(T) + O(k^2/n^4) + O(k^{3/2}/n^3)$, and since $k \to \infty$, the remainder term is $o(k^2/n^3)$.