PROPERTIES OF TWO MODIFIED MOMENT ESTIMATORS FOR PARAMETERS OF THE NEGATIVE BINOMIAL DISTRIBUTION

by

J. Richard Hebel, B.S.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

December 1965

Blacksburg, Virginia

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

CHAPTER I

INTRODUCTION

(a) Preliminary Remarks.

The negative binomial distribution (NBD) is a widely used two-parameter distribution with applications in the fields of biology, psychology, and others. The earliest known derivation of the NBD is due to Montmort in 1714 (Todhunter [21]), and it was discussed by Pascal and Fermat (Todhunter [21]).

The problem of estimating the parameters of the NBD has proved to be an interesting and challenging one. As yet, no consistent method of estimation has been developed for which such properties as bias, variance, etc., of both estimators can be determined exactly. The estimation problem is further complicated by the many different parametric forms of the negative binomial distribution, none of which is used exclusively.

The maximum likelihood approach to the joint estimation of the two parameters has two serious drawbacks. First, the computation of one of the estimators is always difficult. Several iterative and "trial and error" computational techniques have been proposed but these are tedious and require the aid of an electronic computer. In some instances, no admissible solution to the maximum likelihood equation exists. Second, while the asymptotic properties of the maximum likelihood estimators are well known, the small-sample properties are quite difficult to determine.

The difficulties associated with the maximum likelihood method of estimation in the case of the NBD have caused many workers to turn to other methods in search of estimators whose large-sample properties approach those of the maximum likelihood estimators. Among the many estimators which have been proposed, the moment estimators are the easiest to compute. The simplest moment estimators, i.e., those obtained by equating the first two sample moments to the first two population moments, have sometimes been rejected on the basis of poor asymptotic efficiency. However, recent work has shown that large-sample properties do not always provide a sound means of comparing the various estimators. There is evidence, in the case of the negative binomial distribution, that the size of a sample required to justify use of the asymptotic results is often far beyond practical limits. It thus seems important that one should, in this case, be concerned with small-sample properties.

An infinite number of moment estimators can be constructed for the parameters of the NBD. It is also possible to alter the construction method of fitting moments so as to produce <u>modified moment estimators</u>. Two such estimation procedures are developed and discussed at length in this dissertation.

(b) <u>Purpose</u>.

The primary purpose of this thesis is to give results for the properties of two modified moment estimators by means of asymptotic expansions of the bias, variance, determinant of the variance-covariance matrix, and higher moments of these estimators. Several terms of the expansions have been obtained in order to approximate the small-sample properties of the estimators.

Certain background material is presented in this dissertation. The various forms of the negative binomial distribution are discussed. The estimation problem, for the parameters of the NBD, is considered and some properties of estimators, studied by other investigators, are given. The practical applications for the NBD are mentioned and some examples given. Since orthogonal statistics (Shenton and Myers [15]) play an important role in the development of modified moment estimators and in the investigation of their properties, they are discussed in detail. The actual construction of modified moment estimators is also described.

A special computational technique was used which avoids much of the tedious algebraic work ordinarily involved with asymptotic expansions. This technique is described and an example is given. The numerical terms in the asymptotic expansions of the properties for the two modified moment estimators are tabulated. The properties of these estimators are then compared to similar properties of the maximum likelihood and simple moment estimators. Finally, the NBD is fitted to some experimental data, using the method of modified moments and others.

CHAPTER II

FORMS AND ESTIMATORS FOR THE NEGATIVE BINOMIAL DISTRIBUTION

(a) Forms.

There are a number of parametric forms for the negative binomial distribution. Three forms which are commonly used are as follows:

$$P_x = \binom{\alpha+x-1}{x} \frac{\lambda^x \alpha^\alpha}{(\lambda+\alpha)^{\alpha+x}} \quad , \tag{2-1}$$

where $\alpha, \lambda > 0$, $x = 0,1,\ldots$;

$$P_x = \binom{k+x-1}{x} \frac{p^x}{(1+p)^{k+x}} \quad , \tag{2-2}$$

where $k,p > 0$, $x = 0,1,\ldots$;

$$P_x = \binom{m/a+x-1}{x} \frac{a^x}{(1+a)^{x+m/a}} \quad , \tag{2-3}$$

where $m,a > 0$, $x = 0,1,\ldots$, $P_x = \Pr(X=x)$.

In each case the combinatorial term is to be taken as unity when $x=0$. The parametric forms (2-1), (2-2), and (2-3) were introduced by Anscombe [1], Fisher [8], and Evans [6] respectively. The mean, variance, and factorial moment generating function for the NBD in each of the three parametric forms above are given in Table I. Other

Table I.

Properties of the Negative Binomial Distribution

Form	Mean	Variance	F.M.G.F. (parameter t)
(2-1)	λ	$\lambda+\lambda^2/\alpha$	$(1-\lambda t/\alpha)^{-\alpha}$
(2-2)	kp	kp(1+p)	$(1-pt)^{-k}$
(2-3)	m	m(a+1)	$(1-at)^{-m/a}$

parametric forms, such as the inverse binomial sampling or Pascal form (Todhunter [21]), have been given but (2-1), (2-2), and (2-3) have received the most attention. It might be noted here that in the inverse binomial sampling form, one parameter is known and the other parameter is easily estimated in terms of the known parameter. This is then actually a one-parameter situation, which does not concern us in this work.

(b) Estimators.

The problem of estimation, in the case of the two-parameter negative binomial distribution has received wide attention. Of course, this problem differs in nature according to the parametric form which is chosen. In this thesis the form to be considered is that given by (2-1).

1. <u>Maximum Likelihood Estimation</u>.

The maximum likelihood estimators of λ and α are:

$$\lambda^* = m_1' \quad ,$$

where m_1' is the arithmetic mean of a sample of size n, and α^*, which is a root of the equation

$$n \, \ell n(1+m_1'/\alpha) = \sum_{x=1}^{\infty} n_x \left(\frac{1}{\alpha} + \frac{1}{\alpha+1} + \ldots + \frac{1}{\alpha+x-1}\right) \quad , \qquad (2-4)$$

where n_x is the frequency of x. In the sequel, the notation

$$m_j' = \sum_{i=1}^{n} \frac{x_i^j}{n} \quad ; \qquad \mu_j' = E(X^j) \quad ;$$

$$m_j = \sum_{i=1}^{n} \frac{(x_i-m_1')^j}{n} \quad ; \qquad \mu_j = E[(X-\mu_1')^j] \quad ,$$

will be used for the sample moments and population moments.

The momental properties of α^* have not been determined exactly. The asymptotic results are well known but there is very little knowledge of the small-sample properties of α^*. In order to extend beyond the usual "large-sample" results, Bowman and Shenton [2] have considered the following asymptotic expansions:

$$E(\alpha^*-\alpha) = \frac{B_1^*}{n} + \frac{B_2^*}{n^2} + \ldots \quad , \tag{2-5}$$

$$\mathrm{Var}(\alpha^*) = \frac{V_1^*}{n} + \frac{V_2^*}{n^2} + \ldots \quad , \tag{2-6}$$

$$\begin{vmatrix} \mathrm{Var}(\lambda^*) & \mathrm{Cov}(\alpha^*,\lambda^*) \\ \mathrm{Cov}(\alpha^*,\lambda^*) & \mathrm{Var}(\alpha^*) \end{vmatrix} = \frac{C_2^*}{n^2} + \frac{C_3^*}{n^3} + \ldots \quad , \tag{2-7}$$

and have obtained numerical values for B_1^*, B_2^*, V_1^*, V_2^*, C_2^*, and C_3^* in the parameter subspace ($1 \leq \alpha \leq 100$, $1 \leq \lambda \leq 100$). The results not only give an indication of the small-sample properties of the estimator α^*, but also cast some doubt on the use of the "large-sample" term as a criterion in the evaluation of estimators. For example, at $\alpha=3$, $\lambda=1$:

$$\mathrm{Var}(\alpha^*) = \frac{357.6}{n} + \frac{97960}{n^2} + \ldots \quad .$$

Thus for a sample size of 100, the n^{-2} term is actually larger than the asymptotic (n^{-1}) term. Even for a sample size of 1000, the n^{-2} term contributes considerably to the variance.

For each of the expansions (2-5), (2-6), and (2-7), the second term dominates the asymptotic term when $\alpha > \lambda$. This effect is less striking as α approaches λ and is scarcely in evidence at all when $\lambda > \alpha$. Very large sample sizes are required in order that the asymptotic term be safely used as the momental value except when $\lambda > \alpha$. The behavior of the first two terms in the expansions seems to substantiate the claim that in order for (2-4) to have a real positive solution, m_2 must be greater than m_1'. Anscombe [1] conjectured that there is no admissible solution to the equation (2-4) when $m_2 < m_1'$. Shenton [18] was able to prove for a number of particular cases that (2-4) has no real solutions when $m_2 = m_1'$.

We find that if we hold λ constant and allow α to approach infinity, the negative binomial distribution approaches the Poisson distribution. The region of the parameter space where $\alpha \gg \lambda$ is then the "near Poisson" region. Therefore as the Poisson region is approached, the probability that $m_2 = m_1'$ becomes greater and, thus, we would expect the moments of α^* to become infinite.

Of course in order to better evaluate the small-sample properties of α^*, it would be helpful to obtain terms beyond those determined by Bowman and Shenton in the expansions (2-5), (2-6), and (2-7). The problem of obtaining these terms, however, has proved to be quite laborious, even with the aid of an electronic computer, and has not been attempted.

Practical determination of the maximum likelihood estimator (even where a solution actually exists) is also troublesome. The solution of (2-4) requires an iterative technique, such as that given by Haldane [10]. Of course without the use of an electronic computer, the experimentor is faced with a difficult computational problem. The difficulties presented by the maximum likelihood estimator α^* have led to the investigation of estimators which are more easily obtained and whose properties appear to be favorable.

2. Estimation by the Method of Moments.

Anscombe [1] suggests, as an alternative to the maximum likelihood estimation of α, the following procedure. Let f_x be any function of x. Hence, $E(f_x)$ will be some function of the parameters $F(\lambda,\alpha)$. Since m_1' is a fully efficient estimator of λ, the estimator for α is taken as the root of the equation

$$\frac{1}{n} \sum_{x=0}^{\infty} n_x f_x = E(f_x)\Big|_{\lambda=m_1'}$$

$$= F(m_1',\alpha) \quad . \tag{2-8}$$

This class of estimators $\{\hat{\alpha}\}$ will have the following asymptotic properties (Anscombe [1]):

$$\mathrm{Cov}(\hat{\alpha},\hat{\lambda}) \to 0$$

$$\mathrm{Var}(\hat{\alpha}) \to \frac{E(f_x^2) - [E(f_x)]^2 - (\lambda + \frac{\lambda^2}{\alpha})A_\lambda^2}{n\,A_\alpha^2} \quad ,$$

where

$$A_\lambda = \frac{\partial}{\partial\lambda} F(\lambda,\alpha) \quad ,$$

$$A_\alpha = \frac{\partial}{\partial\alpha} F(\lambda,\alpha) \quad .$$

If we let $f_x = x^2$ in (2-8), we construct the method of moments estimator or simple moment estimator, i.e.,

$$\hat{\alpha} = \frac{m_1'^2}{m_2 - m_1'} \quad .$$

As with the maximum likelihood estimator, it is not possible to determine the exact moments of $\hat{\alpha}$. Shenton and Myers [14] considered the asymptotic expansions:

$$E(\hat{\alpha}-\alpha) = \frac{B_1}{n} + \frac{B_2}{n^2} + \ldots \quad , \tag{2-9}$$

$$\mathrm{Var}(\hat{\alpha}) = \frac{V_1}{n} + \frac{V_2}{n^2} + \ldots \quad , \tag{2-10}$$

$$\begin{vmatrix} \mathrm{Var}(\hat{\lambda}) & \mathrm{Cov}(\hat{\alpha},\hat{\lambda}) \\ \mathrm{Cov}(\hat{\alpha},\hat{\lambda}) & \mathrm{Var}(\hat{\alpha}) \end{vmatrix} = \frac{C_2}{n^2} + \frac{C_3}{n^3} + \ldots \quad . \tag{2-11}$$

For the bias and variance expansions, Shenton and Myers derived algebraic expressions for the terms through n^{-4}. Terms through n^{-3} were found for the expansion of the covariance determinant.

The results indicate that the behavior of $\hat{\alpha}$ is very similar to that of α^*. For example, at $\alpha=3$, $\lambda=1$,

$$\mathrm{Var}(\hat{\alpha}) = \frac{384.0}{n} + \frac{1.055\times10^{5}}{n^{2}} + \frac{2.994\times10^{7}}{n^{3}} + \frac{9.565\times10^{9}}{n^{4}} + \dots \quad .$$

Again we see that the asymptotic term is not an adequate representation of the variance of $\hat{\alpha}$, even for what is often considered large samples. This condition is more pronounced as we approach the Poisson region, i.e., $\alpha \gg \lambda$. For at $\alpha=100$, $\lambda=1$,

$$\mathrm{Var}(\hat{\alpha}) = \frac{2.061\times10^{8}}{n} + \frac{3.379\times10^{13}}{n^{2}} + \frac{5.958\times10^{18}}{n^{3}} + \frac{1.227\times10^{24}}{n^{4}} + \dots \quad .$$

Obviously, $\hat{\alpha}$ becomes indefinitely large as m_2 approaches m_1', and thus, the moments of $\hat{\alpha}$ approach infinity in the Poisson region. It is not surprising that this is the region of similar behavior for the moments of the maximum likelihood estimator. Of particular importance, however, is that this behavior is not necessarily indicated by the asymptotic term, but is clearly displayed by the terms beyond the n^{-1} term. On the other hand, the asymptotic term, in the expansions (2-9), (2-10), and (2-11), does dominate when $\lambda > \alpha$ for reasonably large sample sizes, say n=250. At α=3, λ=5, for example,

$$\mathrm{Var}(\hat{\alpha}) = \frac{61.44}{n} + \frac{1835}{n^2} + \frac{72300}{n^3} + \frac{1030000}{n^4} + \cdots \quad .$$

Thus, in the region where $\lambda > \alpha$, we are able to safely approximate the moments of $\hat{\alpha}$ by use of the first four terms in the expansions for small sample sizes, say n=100.

The asymptotic efficiency of $\hat{\alpha}$, i.e.,

$$\bar{E} = \frac{C_2^*}{C_2} = \frac{V_1^*}{V_1} \tag{2-12}$$

is plotted in Figure 1 for various λ and α. Note that in the (α,λ) region where the asymptotic efficiency is low, the n^{-1} term in the expansion of $\mathrm{Var}(\hat{\alpha})$ dominates. Conversly, in the region of high asymptotic efficiency, the n^{-1} term of $\mathrm{Var}(\hat{\alpha})$ is least effective. Previous investigators, such as Fisher [8] and Anscombe [1], have stressed the importance of asymptotic efficiency in the choice of an estimator for α. They recommend the use of $\hat{\alpha}$ when $\bar{E} > 90\%$. We now see that one could easily be mislead by this criterion.

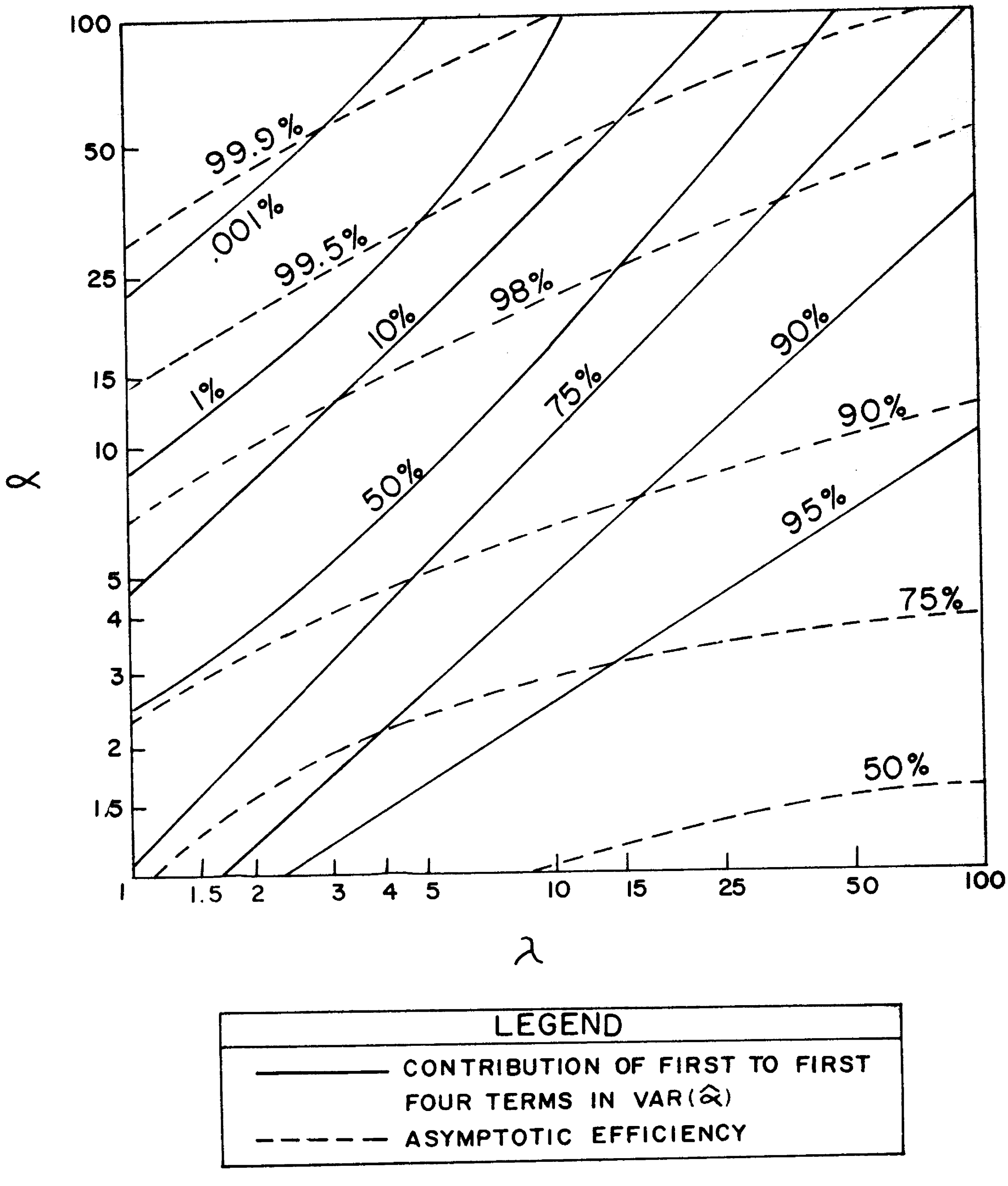

Figure 1. The asymptotic efficiency of $\hat{\alpha}$ and the contribution of the n^{-1} term in Var($\hat{\alpha}$) for n=250.

3. <u>Frequency Estimators</u>.

If we let $f_o = 1$ and $f_x = 0$ for $x \geq 1$ in (2-8), we construct the zeroth frequency estimator, which is given as the root of

$$\frac{n_o}{n} = (1 + \frac{m_1'}{\alpha})^{-\alpha} \quad .$$

Let us denote this estimator as $\hat{\alpha}_o$. Anscombe [1] gives the asymptotic efficiency of $\hat{\alpha}_o$, which is shown in Figure 2 for various α and λ. Only when both α and λ are small does the asymptotic efficiency exceed 90%.

Expansions of the bias and variance of $\hat{\alpha}_o$, through terms to n^{-2}, have been determined by Cassidy [3] in the parameter subspace $(1 \leq \alpha \leq 10,\ 1 \leq \lambda \leq 10)$. As with the expansions for α^* and $\hat{\alpha}$, the n^{-2} term is dominant when $\alpha > \lambda$. Using the same case of $\alpha=3$, $\lambda=1$, as an example,

$$\mathrm{Var}(\hat{\alpha}_o) = \frac{436.9}{n} + \frac{140293}{n^2} + \ldots \quad .$$

However, in the region where $\lambda > \alpha$, the behavior of $\hat{\alpha}_o$ differs from that of α^* and $\hat{\alpha}$. For example, at $\alpha=3$, $\lambda=5$,

$$\mathrm{Var}(\hat{\alpha}_o) = \frac{127.1}{n} + \frac{14715}{n^2} + \ldots \quad .$$

Thus, the n^{-2} term contributes heavily even when $\lambda > \alpha$ and the asymptotic properties of $\hat{\alpha}_o$ are then of questionable

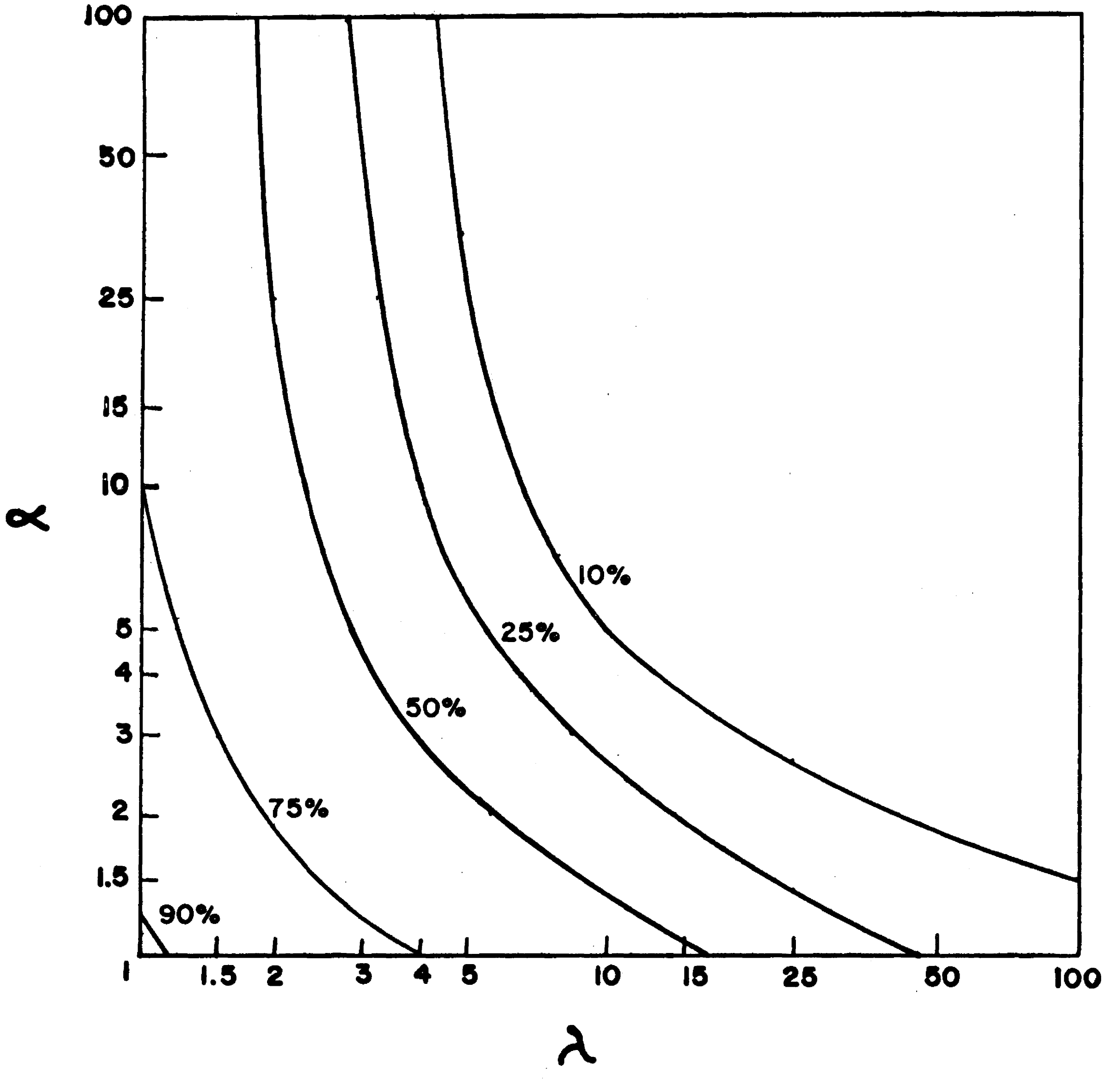

Figure 2. The asymptotic efficiency of $\hat{\alpha}_o$.

value. It should further be noted that the n^{-2} terms of the bias and variance expansions for $\hat{\alpha}_o$ are consistently larger than those for α^* and $\hat{\alpha}$. This would seem to indicate that $\hat{\alpha}_o$ does not present itself as particularly "attractive" for estimating α.

Of course, other frequency estimators for α could be constructed. However, none of the properties of these have been investigated.

4. Other Estimators.

Anscombe [1] proposed two other estimators for α arising from (2-8); these are the inverse moment estimator and the geometric moment estimator. The inverse moment estimator is obtained by letting $f_x = \frac{1}{x+1}$ and, hence, is the solution of the equation

$$\frac{1}{n} \sum_{x=0}^{\infty} \frac{n_x}{x+1} = \frac{(1-\hat{X}) - (1-\hat{X})^{\alpha}}{(\alpha-1)\hat{X}} \quad ,$$

where $\hat{X} = m_1'/(m_1'+\alpha)$. Similarly, by letting $f_x = C^x$, C being a positive constant not equal to unity, (2-8) becomes

$$\frac{1}{n} \sum_{x=0}^{\infty} n_x \, C^x = [1 + (1-C) \frac{m_1'}{\alpha}]^{-\alpha} \quad .$$

Solving for α we then obtain the geometric moment estimator. Anscombe quotes the asymptotic variance for both these estimators but, as yet, no further investigation of them has been undertaken.

Three other possibilities for f_x are given by Anscombe. They are:

i) $f_o = 0, \quad f_x = 1 + \frac{1}{2} + \ldots + \frac{1}{x} \quad \text{for } x \geq 1 \quad ;$

ii) $f_x = \ell n(x+1) \quad \text{for } x \geq 0 \quad ;$

iii) $f_x = \sqrt{x} \quad \text{for } x \geq 0 \quad .$

However, the difficulty of obtaining $E(f_x)$ in each of these cases has discouraged any further investigation.

In addition to the simple moment estimator discussed earlier, it is possible to construct a variety of other moment estimators by equating ratios of sample moments to the corresponding population quantities. For example, by using the ratios of third to second factorial moments, i.e., by letting

$$\frac{m_{(3)}}{m_{(2)}} = \frac{\mu_{(3)}}{\mu_{(2)}} \quad , \tag{2-13}$$

where,

$$m_{(k)} = \sum_{j=1}^{n} \frac{x_j(x_j-1) \ldots (x_j-k+1)}{n} \quad ,$$

$$\mu_{(k)} = E[X(X-1) \ldots (X-k+1)] \quad ,$$

and solving for α, we obtain the estimator

$$\hat{\alpha}_{(1)} = \frac{2m_1' \, m_{(2)}}{m_{(3)} - m_1' \, m_{(2)}} \quad . \tag{2-14}$$

If, instead of using factorial moments in (2-13), we use factorial cumulants, we obtain the estimator

$$\hat{\alpha}_{(2)} = \frac{2m_1' m_{(2)} - 2m_1'^3}{m_{(3)} - 3m_1' m_{(2)} + 2m_1'^3} \quad . \tag{2-15}$$

It is, in fact, possible to develop an infinite number of moment estimators for α. A general method of constructing moment estimators is discussed in Chapter IV and by that method we are able to obtain modified moment estimators, which are the primary concern of this dissertation.

(c) Practical Applications for the NBD.

The negative binomial distribution is one of the most widely used discrete distributions. It has been used in work with biological data, accident statistics, industrial sampling, and psychological data. Other two-parameter distributions, such as the Neyman Type A, are sometimes used, but the negative binomial is usually attempted when a two-parameter discrete distribution is hypothesized. To illustrate the utility of the NBD, some examples of its practical use will be given.

Many successful applications of the NBD have arisen from the following model. Suppose there is an occurrence of events or individuals, which may be counted, in each of many units of time or space. If each unit has an equal chance of an occurrence, it is well known that the Poisson model holds. However, if there is an unequal chance of occurrence from unit to unit, the negative binomial law holds. This hypothesis is due to Greenwood and Yule [9].

Sichel [13] demonstrated the use of the model in a study of absence proneness. Through an intensive consideration of absenteeism among industrial workers, it was established that liability of absence from work differs from person to person. Absence data for 318 steel workers over a period of six months was fitted successfully by use of the negative distribution. The method of estimation, however, for the parameter α, was found to have a substantial influence on the results. The sample mean (in this case, $m_1' = .670$) was used to estimate λ and both the maximum likelihood estimator ($\alpha^* = .396$) and the simple moment estimator ($\hat{\alpha} = .522$) were used to estimate α. A satisfactory fit of the data was obtained by use of α^* and a questionable fit by use of $\hat{\alpha}$. Notice, from the values of the estimators, that here we are in an "apparent" (α,λ) region where the asymptotic efficiency is low for $\hat{\alpha}$. Also, it is a region in which one can safely rely on asymptotic results for $\hat{\alpha}$.

Sichel also used the negative binomial to fit Two-Hand Co-ordination Test data. The Two-Hand Co-ordination Test is used primarily as a measure of speed of performance. The liability to making errors in the test varies among individuals, but remains remarkably constant for a particular individual from trial to trial. Results of the Two-Hand Co-ordination Test for 504 subjects in ten trials were successfully fitted using both α^* and $\hat{\alpha}$. In this case, $\hat{\lambda} = 7.284$, $\alpha^* = .775$, and $\hat{\alpha} = .631$. Again, a better fit was obtained with α^* as the estimator for α.

An interesting application was provided by Wise [23] in the solution of an industrial sampling problem. Batches of enamelled wire, consisting of 50 or more reels, were being tested for pinholes in the enamel. Previously, the procedure was to take a sample of one third of the reels in the batch and test 50-yard lengths from each reel in the sample for pinholes. If any length had more than R pinholes, the whole batch was tested (R being predetermined according to the gauge of the wire). It was desired to devise a method for which a smaller sample would be needed to accept the batch. Since there was an inevitable change in manufacturing conditions over a period of time, it was naturally expected that the quality of the wire would differ from reel to reel. It was therefore assumed that the 50-yard lengths were being taken at random from a population in which the number of pinholes was distributed according to the negative binomial law.

The expected number of leaks, h, in a wire used in practice was determined as a function of the population parameters and the physical properties of the wire, i.e.,

$$h = K\,\lambda^2\left(1 + \frac{1}{\alpha}\right) \quad ,$$

where K is a constant depending on the dimensions of the wire and the areas of the pinholes. Thus the mean λ was more important than α in determining the reliability of the wire. It was decided to sample sequentially from a batch until the sample mean m_1' was less than a certain λ_b, at which time the batch would be conditionally accepted. The value of λ_b was chosen so that about 6 per cent of a batch with mean λ_b had more than R pinholes in 50 yards and also so that the probability of accepting a bad batch would be no more than .05. The decision to accept or continue sampling was to be made at sample sizes of 10, 15, 20, 30, and 60.

If the batch was accepted on the basis of its sample mean at a sample size N, a quantity we shall call $\hat{R}$ was computed by use of estimates of the parameters λ and α. The quantity $\hat{R}$ is such that:

$$P_r(\text{not more than } \hat{R} \text{ pinholes in } N) = .90 \quad .$$

The batch was rejected if the observed maximum number of pinholes in the sample exceeded $\hat{R}$.

Williamson and Bretherton [22] found an application of the NBD to an inventory control problem. It is often reasonable to assume that demands for items at a store follow a Poisson process. The number of items requested per demand may also be regarded as a random variable. Feller [7] has shown that if $G_1(t)$ is the probability generating function of the distribution of demands per unit time, and $G_2(t)$ is the probability generating function of the distribution of items requested per demand, then the distribution of items requested per unit time has the compound generating function $G_1[G_2(t)]$.

It was shown that the number of items requested per demand could be reasonably well fitted by the logarithmic distribution. If $G_1(t)$ is the Poisson generating function and $G_2(t)$ is the logarithmic generating function, the resulting compound generating function is that of the negative binomial distribution. Thus a reasonable model, representing withdrawals of items from a store, is the negative binomial probability law.

CHAPTER III

ORTHOGONAL STATISTICS

Orthogonal statistics have been used to a large extent in work with the negative binomial distribution to facilitate exploration of the momental properties of moment estimators. They are of importance, both in the development of modified moment estimators and in the investigation of their properties. At this point we shall review the concept of orthogonal statistics and, more basically, the concept of orthogonal polynomials with respect to a statistical distribution.

(a) Orthogonal Polynomials.

Let $F(x)$ be a distribution function with finite moments of all orders. Then there exists a set of orthogonal polynomials $\{q_r(x)\}$ such that

$$\begin{aligned} \int_{-\infty}^{\infty} q_r(x)\, q_s(x)\, dF(x) &= \varphi_r \qquad \text{for} \quad r = s \quad , \\ &= 0 \qquad \text{for} \quad r \neq s \quad , \\ & \qquad\qquad (r,s = 0,1,2,\ldots) \end{aligned}$$

where $\varphi_r > 0$, $\varphi_o = 1$, and the coefficient of x^r in $q_r(x)$

is unity (Cramer [4]). In the case of a discrete distribution for which

$$P_x = P_r[X=x] \qquad (x = 0,1,2,\ldots) \quad ,$$

the set $\{q_r(x)\}$ would have the properties:

$$\begin{aligned} \sum_x q_r(x) q_s(x) P_x &= \varphi_r \qquad \text{for} \quad r = s \quad , \\ &= 0 \qquad \text{for} \quad r \neq s \quad , \end{aligned}$$

where $\varphi_r > 0$. One point, that must be emphasized, is that the orthogonal polynomials are not only functions of the random variables but also contain the population parameters. The orthogonal polynomials for many of the classical distributions are well known. A general treatment of orthogonal polynomials and the method of obtaining them is given by Szegö [20].

For the negative binomial distribution, Shenton [19] has developed the following expression:

$$q_r(x) = (1 - \frac{\lambda}{\alpha}\,\Delta)^{\alpha+r-1}\; x^{(r)} \quad ,$$

where $x^{(r)} = x(x-1)\ \ldots\ (x-r+1)$, $x^{(o)} = 1$, and Δ is a symbolic difference operator, i.e., such that

$$\Delta\, x^{(r)} = r\, x^{(r-1)} \quad .$$

Hence,

$$q_1(x) = (1 - \frac{\lambda}{\alpha}\Delta)^{\alpha} x^{(1)} \quad ,$$

$$= (1 - \lambda\Delta + \ldots) x \quad .$$

Notice that there is no need to expand beyond the term involving Δ since

$$\Delta^{r+1} x^{(r)} = 0 \quad .$$

Then

$$q_1(x) = x - \lambda \quad .$$

Likewise we find

$$q_2(x) = x^{(2)} - \frac{2\lambda}{\alpha}(\alpha+1)x + \frac{(\alpha+1)}{\alpha}\lambda^2 \quad ,$$

$$q_3(x) = x^{(3)} - 3\frac{(\alpha+2)}{\alpha}\lambda x^{(2)} + 3\frac{(\alpha+2)}{\alpha^2}(\alpha+1)\lambda^2 x$$

$$- \frac{\lambda^3}{\alpha^2}(\alpha+2)(\alpha+1) \quad .$$

Shenton was also able to show that

$$\varphi_r = \frac{\lambda^r(\lambda+\alpha)^r(\alpha+r-1)^{(r)}\ r!}{\alpha^{2r}} \quad ,$$

where

$$(\alpha+r-1)^{(r)} = (\alpha+r-1)(\alpha+r-2)\ldots(\alpha+1)\alpha \quad .$$

Then

$$\varphi_1 = \lambda(\lambda+\alpha)/\alpha \quad ,$$

$$\varphi_2 = 2\lambda^2(\lambda+\alpha)^2(\alpha+1)/\alpha^3 \quad ,$$

$$\varphi_3 = 6\lambda^3(\lambda+\alpha)^3(\alpha+1)(\alpha+2)/\alpha^5 \quad .$$

(b) <u>Definition of Orthogonal Statistics</u>.

Suppose we take a random sample $(x_1,x_2,\ldots,x_n)$ from a population with distribution function $F(x)$. The orthogonal statistic Q_r is then defined as

$$Q_r = \sum_{i=1}^{n} \frac{q_r(x_i)}{n} \quad . \tag{3-1}$$

Note that Q_r is a pseudo-statistic since it will contain parameters of the distribution. It is easily seen that orthogonal statistics have the important property

$$\begin{aligned} E(Q_rQ_s) &= \varphi_{r/n} && \text{for } r = s \\ &= 0 && \text{for } r \neq s \quad , \end{aligned}$$

where $(r,s = 0,1,2,\ldots)$. Also obvious, from the definition of the orthogonal statistic Q_r, is that Q_r is a linear function of sample moments. The highest order of these sample moments will be r.

(c) Application of Orthogonal Statistics.

Since orthogonal statistics are linear functions of sample moments, it follows that sample moments can be expressed in terms of orthogonal statistics. With this knowledge, we are led to an important application of orthogonal statistics.

Suppose we are interested in finding the sampling moments of a moment statistic $t(m_1', m_2', \ldots, m_k')$. Suppose further that t can be expanded in the form

$$\begin{aligned} t = {} & a_{0,0,0,\ldots,0} + a_{1,0,0,\ldots,0}\, Q_1 + a_{0,1,0,\ldots,0}\, Q_2 \\ & + a_{0,0,1,0,\ldots,0}\, Q_3 + \ldots + a_{1,1,0,\ldots,0}\, Q_1 Q_2 \\ & + a_{1,0,1,0,\ldots,0}\, Q_1 Q_3 + \ldots + a_{2,0,0,\ldots,0}\, Q_1^2 + \ldots \\ & + a_{u,v,\ldots,w}\, Q_1^u Q_2^v \ldots Q_k^w + \ldots \quad . \end{aligned} \tag{3-2}$$

In some cases, this expansion may be non-terminable. Theoretically, however, the first moment of t is found by obtaining the expectation of each term in (3-2) as a function of the population parameters and the sample size n. Furthermore, powers of t can be expanded in a series such as (3-2). Hence, if the expected values of the various Q-products were known, it would be possible to exactly obtain any moment of t or to approximate that moment with an asymptotic expansion. The method described here is an alternative to the use of k-statistics (David [5]). We shall now discuss this application of orthogonal statistics with particular reference to the negative binomial distribution.

1. Relationships between Orthogonal Statistics and Sample Moments.

Shenton and Wallington [16] give the following relationship for the case of the NBD:

$$x^{(r)} = (1 + \frac{\lambda}{\alpha} \Delta)^{\alpha+r-1} q_r(x) \quad , \tag{3-3}$$

where $x^{(r)} = x(x-1)\ldots(x-r+1)$ and Δ is a symbolic difference operator, i.e.,

$$\Delta\, q_r(x) = r\, q_{r-1}(x) \quad .$$

With the relationship given by (3-3), we can develop other useful results. Suppose, for example, we start with

$$x^{(2)} = (1 + \frac{\lambda}{\alpha} \Delta)^{\alpha+1} q_2(x) \quad .$$

Notice that, in order to simplify the right hand side, it would be unnecessary to carry a binomial expansion further than the term which involves Δ^2. Thus,

$$x^{(2)} = [1 + \frac{(\alpha+1)}{\alpha}\lambda\Delta + \frac{(\alpha+1)}{2\alpha}\lambda^2\Delta^2 + \dots]q_2(x)$$

$$= q_2(x) + 2(1 + \frac{1}{\alpha})\lambda\ q_1(x) + (1 + \frac{1}{\alpha})\lambda^2 \quad .$$

Summing both sides over the sample $(x_1, x_2, \dots, x_n)$, and dividing each side by n,

$$\sum_{i=1}^{n} \frac{x_i(x_i-1)}{n} = \sum_{i=1}^{n} \frac{[q_2(x_i) + 2(1+\frac{1}{\alpha})\lambda q_1(x_i)]}{n} + (1+\frac{1}{\alpha})\lambda^2 \quad ,$$

or

$$m_{(2)} = Q_2 + 2(1+\frac{1}{\alpha})\lambda\ Q_1 + (1+\frac{1}{\alpha})\lambda^2 \quad .$$

Using the same procedure we can find the general relationships between sample factorial moments and orthogonal statistics. Starting with (3-3) and expanding the right hand side, we have

$$x^{(r)} = [1 + \binom{\alpha+r-1}{1}\frac{\lambda}{\alpha}\Delta + \binom{\alpha+r-1}{2}(\frac{\lambda}{\alpha})^2\Delta^2 + \dots$$

$$+ \binom{\alpha+r-1}{r-1}(\frac{\lambda}{\alpha})^{r-1}\Delta^{r-1} + \binom{\alpha+r-1}{r}(\frac{\lambda}{\alpha})^r\Delta^r + \dots]q_r(x)$$

$$= q_r(x) + \binom{r}{1}\lambda(1+\frac{r-1}{\alpha})q_{r-1}(x)$$

$$+ \binom{r}{2}\lambda^2(1+\frac{r-1}{\alpha})(1+\frac{r-2}{\alpha})q_{r-2}(x) + \dots$$

$$+ \binom{r}{r-1}\lambda^{r-1}(1+\frac{r-1}{\alpha})\ \dots\ (1+\frac{1}{\alpha})q_1(x)$$

$$+ \lambda^r(1+\frac{r-1}{\alpha})\ \dots\ (1+\frac{1}{\alpha}) \quad .$$

Hence the result:

$$m_{(r)} = Q_r + \sum_{i=1}^{r} \binom{r}{i} \lambda^i \left(1+\frac{r-1}{\alpha}\right)\left(1+\frac{r-2}{\alpha}\right)\cdots\left(1+\frac{r-i}{\alpha}\right) Q_{r-i} \quad . \tag{3-4}$$

We are now in a position to demonstrate an expansion such as that given by (3-2). Consider the simple moment estimator for the case of the NBD,

$$\hat{\alpha} = \frac{m_1'^2}{m_2 - m_1'} \quad .$$

Expressed in terms of sample factorial moments,

$$\hat{\alpha} = \frac{m_{(1)}^2}{m_{(2)} - m_{(1)}^2} \quad .$$

Using the relationship given by (3-4), we obtain

$$\hat{\alpha} = \frac{\alpha + \frac{2\alpha}{\lambda} Q_1 + \frac{\alpha}{\lambda^2} Q_1^2}{1 + \frac{2}{\lambda} Q_1 + \frac{\alpha}{\lambda^2} Q_2 - \frac{\alpha}{\lambda^2} Q_1^2} \quad .$$

We can then expand $\hat{\alpha}$ to obtain the form of (3-2), i.e.,

$$\hat{\alpha} = \alpha - \frac{\alpha^2}{\lambda^2} Q_2 + \frac{\alpha(\alpha+1)}{\lambda^2} Q_1^2 + \frac{\alpha^3}{\lambda^4} Q_2^2 + \frac{2\alpha^2}{\lambda^3} Q_1 Q_2 - \frac{2\alpha(\alpha+1)}{\lambda^3} Q_1^3$$
$$- \frac{\alpha^4}{\lambda^6} Q_2^3 - \frac{4\alpha^3}{\lambda^5} Q_1 Q_2^2 - \frac{\alpha^2(2\alpha+5)}{\lambda^4} Q_1^2 Q_2 + \cdots \quad . \tag{3-5}$$

Clearly, the expansion is non-terminating. A question arises as to the convergence of an expansion such as (3-5).

If $(\frac{2}{\lambda}Q_1 + \frac{\alpha}{\lambda^2}Q_2 - \frac{\alpha}{\lambda^2}Q_1^2)^2 < 1$, then (3-5) will converge to $\hat{\alpha}$. This will depend, not only on the values of α and λ, but also on the sample.

2. Expected Q-Products.

Shenton and Myers [15] have considered the problem of obtaining the expectation of Q-products for the negative binomial distribution. They introduced the following notation:

$$E(Q_r^\alpha Q_s^\beta Q_t^\gamma \ldots) = (r^\alpha s^\beta t^\gamma \ldots) \quad ,$$

$$E(q_r^\alpha q_s^\beta q_t^\gamma \ldots) = [r^\alpha s^\beta t^\gamma \ldots] \quad ,$$

and found that

$$(r^\alpha s^\beta t^\gamma \ldots) = \sum_{j=a}^{b} \frac{(r^\alpha s^\beta t^\gamma \ldots)_j}{n^j} \quad ,$$

where $a = -1+\alpha+\beta+\gamma\ldots$, and $b = \{\frac{1+\alpha+\beta+\gamma+\ldots}{2}\}$ ($\{z\}$ refers to the integral part of z). The coefficient $(r^\alpha s^\beta t^\gamma \ldots)_j$ is a function of expected orthogonal polynomial products. For example, we can write

$$(r^4 s) = \frac{(r^4 s)_3}{n^3} + \frac{(r^4 s)_4}{n^4} \quad ,$$

and from Myers [12], we find

$$(r^4s)_3 = 6[r^2][r^3]$$

$$(r^4s)_3 = [r^4s] - 6[r^2][r^3] \quad .$$

Now it becomes apparent that, even though an expansion of the moment statistic t, expressed in terms of Q-products, may not terminate, it is still possible to find approximations to the moments of t. To do this we take expectation term by term in the expansion of t (or powers of t) and consider the resulting asymptotic expansion. The accuracy of the approximation will depend on how far the asymptotic expansion is extended (assuming that it converges). Shenton and Myers [14] used the expansion (3-5) to investigate the properties of $\hat{\alpha}$. Some empirical sampling work by Myers [12] supported the validity of this technique in the case of $\hat{\alpha}$.

Myers [12] developed a "library" for values of $(1^{\beta}2^{\gamma})_j$ to the eighth order, i.e., $\beta+\gamma \leq 8$, and $j = 1,2,3,4$ for the NBD. Hsing [11] extended this library to include values of $(1^{\beta}2^{\gamma}3^{\delta})_j$ to the eighth order and $j = 1,2,3,4$. The use of the latter library facilitates the study of properties of moment estimators involving the third sample moment.

3. An Example.

As a simple illustration of the application of orthogonal statistics described above, we shall find the expectation of the unbiased estimator of the variance of the negative binomial distribution, i.e.,

$$s^2 = \frac{nm_2}{n-1} \quad .$$

In terms of orthogonal statistics,

$$s^2 = \frac{n}{n-1}\left[-Q_1^2 + Q_2 + \left(\frac{2\lambda+\alpha}{\alpha}\right)Q_1 + \frac{\lambda(\lambda+\alpha)}{\alpha}\right] \quad .$$

Then, taking advantage of the orthogonal property,

$$E(s^2) = \frac{n}{n-1}\left[-(1^2) + \frac{\lambda(\lambda+\alpha)}{\alpha}\right]$$

$$= \frac{n}{n-1}\left[-\frac{(1^2)_1}{n} + \frac{\lambda(\lambda+\alpha)}{\alpha}\right] \quad .$$

From the library given by Myers [12],

$$(1^2)_1 = \frac{\lambda(\lambda+\alpha)}{\alpha} \quad .$$

Therefore

$$E(s^2) = \frac{\lambda(\lambda+\alpha)}{\alpha} \quad .$$

CHAPTER IV

MODIFIED MOMENT ESTIMATORS

(a) General Method of Obtaining Moment Estimators.

Before we can discuss modified moment estimators, it is necessary first to introduce a more general method of deriving moment estimators. This method was developed by Shenton [19] and was later extended by Shenton and Wallington [17] to include modified moment estimators.

Suppose P_x is a discrete frequency function depending on h parameters $(\theta_1,\theta_2,\ldots,\theta_h)$. (The following derivation is essentially unaltered by the consideration of a continuous distribution.) Consider the Fourier expansion (see Shenton [19]):

$$\frac{1}{P_x}\frac{\partial P_x}{\partial \theta_a} = A_1^{(a)}q_1(x) + A_2^{(a)}q_2(x) + \ldots + A_p^{(a)}q_p(x) + \ldots \quad , \tag{4-1}$$

where the q's are orthogonal polynomials and the A's are coefficients which we shall determine later. Now suppose we sum both sides of (4-1) over the sample $(x_1,x_2,\ldots,x_n)$. The left hand side becomes

$$\sum_{i=1}^{n} \frac{1}{P_{x_i}}\frac{\partial P_{x_i}}{\partial \theta_a} \quad .$$

It is easily seen that this expression, when equated to zero, is a maximum likelihood equation. Therefore we obtain maximum likelihood estimators by solving for $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_h$ in the equations;

$$\begin{aligned}
\hat{A}_1^{(1)}\hat{Q}_1 + \hat{A}_2^{(1)}\hat{Q}_2 + \ldots + \hat{A}_p^{(1)}\hat{Q}_p + \ldots &= 0 \\
\hat{A}_1^{(2)}\hat{Q}_1 + \hat{A}_2^{(2)}\hat{Q}_2 + \ldots + \hat{A}_p^{(2)}\hat{Q}_p + \ldots &= 0 \qquad (4\text{-}2) \\
\hat{A}_1^{(h)}\hat{Q}_1 + \hat{A}_2^{(h)}\hat{Q}_2 + \ldots + \hat{A}_p^{(h)}\hat{Q}_p + \ldots &= 0 \quad ,
\end{aligned}$$

where $\hat{A}_s^{(a)} = A_s^{(a)}\Big|_{\theta_i=\hat{\theta}_i}$ and $\hat{Q}_s = Q_s\Big|_{\theta_i=\hat{\theta}_i}$, $i = 1,2,\ldots,h.$

Here the Q's are orthogonal statistics obtained by summing the right hand side of (4-1) over the sample and dividing by n.

If we terminate the expansions (4-2) at, say, the p^{th} term, then the solution of (4-2) will yield moment estimators for the parameters $(\theta_1, \theta_2, \ldots, \theta_h)$. This procedure, as a general method for approximating maximum likelihood solutions, relies on the validity of (4-1), i.e., the convergence of the right hand side to $\frac{1}{P_x}\frac{\partial P_x}{\partial \theta_a}$. The interest here, however, is strictly that of finding alternatives to the maximum likelihood estimators in the case of the negative binomial distribution. We shall see in the sequel that terminating equations (4-2), in the case of the NBD, does generate the usual moment estimators. In fact, for p=2, the solution for $\hat{\alpha}$ in (4-2) is the simple moment estimator discussed earlier. Even if we cannot obtain valid approximations to the likelihood solutions, we can still generate moment and, as we shall see later, modified moment estimators. Thus while the procedure for approximating the maximum likelihood solutions does rely on the convergence of (4-1), the same procedure as a method of generating moment estimators (and modified moment estimators) does not.

To obtain an expression for the coefficient $A_s^{(a)}$, we multiply both sides of (4-1) by $q_s(x)P_x$ and sum over all admissible values of x. Then, using the orthogonality property, we arrive at

$$\sum_x q_s(x)\frac{\partial P_x}{\partial \theta_a} = A_s^{(a)} \sum_x [q_s(x)]^2 P_x$$

$$= A_s^{(a)} \varphi_s \quad . \tag{4-3}$$

Since $E[q_s(x)] = 0$, we can write

$$\sum_x q_s(x) P_x = 0 \quad . \tag{4-4}$$

Differentiating both sides of (4-4) with respect to θ_a, we have

$$\sum_x q_s(x)\frac{\partial P_x}{\partial \theta_a} = -\sum_x \frac{\partial q_s(x)}{\partial \theta_a} P_x \quad .$$

Substituting this result into (4-3), we obtain

$$A_s^{(a)} = \frac{-E[\frac{\partial q_s(x)}{\partial \theta_a}]}{\varphi_s} \quad . \tag{4-5}$$

(b) <u>Construction of Modified Moment Estimators</u>.

Suppose that we have terminated the equations (4-2), each at the p^{th} term, and that θ_a can be expressed as a function of the population moments, i.e.,

$$\theta_a = f_a(\mu_{r_1}, \mu_{r_2}, \ldots, \mu_{r_w}) \quad , \tag{4-6}$$

where $r_1, r_2, \ldots, r_w$ are distinct positive integers. We can also write

$$A_s^{(a)} = g_s^{(a)}(\mu_{r_1}, \mu_{r_2}, \ldots, \mu_{r_w}) \quad . \tag{4-7}$$

Now let us replace some (partially modify) or all (completely modify) the μ's by their corresponding sample moments. Then $A_s^{(a)}$ becomes

$$\underline{A}_s^{(a)} = g_s^{(a)}(m_{r_1}, m_{r_2}, \ldots, m_{r_v}, \mu_{r_{v+1}}, \ldots, \mu_{r_w}). \quad \text{(4-7a)}$$

The estimating equations (4-2) now become

$$\sum_{s=1}^{p} \hat{\underline{A}}_s^{(a)} \hat{Q}_s = 0 \quad , \qquad (a = 1,2,\ldots,h) \quad , \quad \text{(4-8)}$$

where $\hat{\underline{A}}_s^{(a)} = \underline{A}_s^{(a)}\Big|_{\theta_i=\hat{\theta}_i}$, $i = 1,2,\ldots,h$. Notice that $\hat{Q}_s$ remains unaltered and is still defined as in (4-2). We now solve the equations (4-8) for $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_h$ to obtain modified moment estimators.

(c) Asymptotic Covariances of Moment Estimators and Modified Moment Estimators.

Consider the expansion

$$\hat{\theta}_a = \theta_a + \sum_{i=1}^{p} G_i^{(a)} Q_i + \sum_{j=1}^{p} H_i^{(a)} Q_j^2 + \ldots \quad , \quad \text{(4-10)}$$

where $G_i^{(a)}$ and $H_i^{(a)}$ are unknown coefficients. Note that in the expansion we do not include terms such as $Q_r Q_s$, with $s \neq r$, which will vanish in expectation and hence will not be needed. (Refer to the properties of orthogonal statistics in Chapter III.)

Since $\hat{A}_s^{(a)}$ and $\hat{Q}_s$ are purely functions of $\{\hat{\theta}_i\}_{i=1}^{h}$, we can write the following Taylor expansions:

$$\hat{A}_s^{(a)} = A_s^{(a)} + \sum_{i=1}^{h} \frac{\partial A_s^{(a)}}{\partial \theta_i}(\hat{\theta}_i - \theta_i) + \frac{1}{2!} \sum_{j=1}^{h} \sum_{k=1}^{h} \frac{\partial^2 A_s^{(a)}}{\partial \theta_j \partial \theta_k}(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k) + \cdots \quad ; \tag{4-11}$$

$$\hat{Q}_s = Q_s + \sum_{i=1}^{h} \frac{\partial Q_s}{\partial \theta_i}(\hat{\theta}_i - \theta_i) + \frac{1}{2!} \sum_{j=1}^{h} \sum_{k=1}^{h} \frac{\partial^2 Q_s}{\partial \theta_j \partial \theta_k}(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k) + \cdots \quad ; \tag{4-12}$$

assuming that the appropriate derivatives exist and $|\hat{\theta}_i - \theta_i|$ is finite. Later it will be apparent that we do not need the terms beyond those involving first derivatives. Hence by substituting (4-10) into (4-11) and (4-12), we have

$$\hat{A}_s^{(a)} = A_s^{(a)} + \sum_{i=1}^{h} \frac{\partial A_s^{(a)}}{\partial \theta_i}\left(\sum_{j=1}^{p} G_j^{(i)} Q_j + \sum_{k=1}^{p} H_k^{(i)} Q_k^2 + \cdots\right) + \cdots \quad , \tag{4-13}$$

and

$$\hat{Q}_s = Q_s + \sum_{i=1}^{h} \frac{\partial Q_s}{\partial \theta_i}\left(\sum_{j=1}^{p} G_j^{(i)} Q_j + \sum_{k=1}^{p} H_k^{(i)} Q_k^2 + \cdots\right) + \cdots \quad . \tag{4-14}$$

We can also write

$$\frac{\partial Q_s}{\partial \theta_i} = i_{s-1}^{(s)} Q_{s-1} + i_{s-2}^{(s)} Q_{s-2} + \ldots + i_o^{(s)} \quad , \qquad (4\text{-}15)$$

this expression being of degree s-1 (at most) since $q_s(x)$ has unity for its highest coefficient. Now (4-14) becomes

$$\hat{Q}_s = Q_s + \sum_{i=1}^{h} (i_{s-1}^{(s)} Q_{s-1} + i_{s-2}^{(s)} Q_{s-2} + \ldots + i_o^{(s)}) \qquad (4\text{-}16)$$
$$\left(\sum_{j=1}^{p} G_j^{(i)} Q_j + \sum_{k=1}^{p} H_k^{(i)} Q_k^2 + \ldots\right) + \ldots \quad .$$

If we substitute (4-13) and (4-16) into the estimating equations (4-2) and isolate the coefficients of linear terms in Q_s, we find

$$A_s^{(a)} + \sum_{i=1}^{h} \sum_{j=1}^{p} A_j^{(a)} \; i_o^{(j)} \; G_s^{(i)} = 0 \quad . \qquad (4\text{-}17)$$

From (4-15), we can write

$$E\left(\frac{\partial Q_s}{\partial \theta_i}\right) = i_o^{(s)} \quad ;$$

and from (4-5),

$$E\left(\frac{\partial Q_s}{\partial \theta_i}\right) = -A_s^{(i)} \; \varphi_s \quad ,$$

then

$$i_o^{(s)} = -A_s^{(i)} \varphi_s \quad . \qquad (4\text{-}18)$$

Upon substitution of (4-18) in (4-17), we now have

$$A_s^{(a)} = \sum_{i=1}^{h} \sum_{j=1}^{p} A_j^{(a)} A_j^{(i)} \varphi_j G_s^{(i)} \quad . \qquad (4\text{-}19)$$

Thus, we may write the entire scheme in matrix notation, i.e.,

$$\begin{bmatrix} A_1^{(1)} & A_2^{(1)} & \cdots & A_p^{(1)} \\ A_1^{(2)} & A_2^{(2)} & \cdots & A_p^{(2)} \\ \vdots & \vdots & & \vdots \\ A_1^{(h)} & A_2^{(h)} & \cdots & A_p^{(h)} \end{bmatrix} =$$

$$\begin{bmatrix} \sum_{j=1}^{p} A_j^{(1)} A_j^{(1)} \varphi_j & \sum_{j=1}^{p} A_j^{(1)} A_j^{(2)} \varphi_j & \cdots & \sum_{j=1}^{p} A_j^{(1)} A_j^{(h)} \varphi_j \\ \sum_{j=1}^{p} A_j^{(2)} A_j^{(1)} \varphi_j & \sum_{j=1}^{p} A_j^{(2)} A_j^{(2)} \varphi_j & \cdots & \sum_{j=1}^{p} A_j^{(1)} A_j^{(h)} \varphi_j \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{p} A_j^{(h)} A_j^{(1)} \varphi_j & \sum_{j=1}^{p} A_j^{(h)} A_j^{(2)} \varphi_j & \cdots & \sum_{j=1}^{p} A_j^{(h)} A_j^{(h)} \varphi_j \end{bmatrix} \begin{bmatrix} G_1^{(1)} & G_2^{(1)} & \cdots & G_p^{(1)} \\ G_1^{(2)} & G_2^{(2)} & \cdots & G_p^{(2)} \\ \vdots & \vdots & & \vdots \\ G_1^{(h)} & G_2^{(h)} & \cdots & G_p^{(h)} \end{bmatrix} ,$$

or more simply stated

$$[A_c^{(r)}] = [\sum_{j=1}^{p} A_j^{(r)} A_j^{(c)} \varphi_j][G_c^{(r)}] \quad . \qquad (4\text{-}20)$$

Let

$$[\Pi^{(r,c)}] = [\sum_{j=1}^{p} A_j^{(r)} A_j^{(c)} \varphi_j]^{-1} \quad , \qquad (4\text{-}21)$$

which we assume exists. Then we can write

$$G_s^{(a)} = \sum_{i=1}^{h} \Pi^{(a,i)} A_s^{(i)} \quad . \qquad (4\text{-}22)$$

Now we shall consider the asymptotic term of $\mathrm{Cov}(\hat{\theta}_a, \hat{\theta}_b)$, where $\hat{\theta}_a$ and $\hat{\theta}_b$ are moment estimators (unmodified) and $a,b = 1,2,\ldots,h$. From (4-10), we have

$$E(\hat{\theta}_a) = \theta_a + \sum_{j=1}^{p} \frac{H_j^{(a)} \varphi_j}{n} + \ldots \quad ;$$

$$E(\hat{\theta}_b) = \theta_b + \sum_{j=1}^{p} \frac{H_j^{(b)} \varphi_j}{n} + \ldots \quad ;$$

$$E(\hat{\theta}_a \hat{\theta}_b) = \theta_a \theta_b + \sum_{i=1}^{p} \frac{G_i^{(a)} G_i^{(b)} \varphi_i}{n}$$

$$+ \sum_{j=1}^{p} \frac{\theta_a H_j^{(b)} \varphi_j}{n} + \sum_{j=1}^{p} \frac{\theta_b H_j^{(a)} \varphi_j}{n} + \ldots \quad .$$

Then

$$\mathrm{Cov}(\hat{\theta}_a,\hat{\theta}_b) = \sum_{i=1}^{p} \frac{G_i^{(a)} G_i^{(b)} \varphi_i}{n} + \ldots \quad ,$$

and therefore from equation (4-22),

$$\mathrm{Cov}(\hat{\theta}_a,\hat{\theta}_b) \doteq \sum_{i=1}^{p} \frac{\left(\sum_{j=1}^{h} \Pi^{(a,j)} A_i^{(j)}\right)\left(\sum_{k=1}^{h} \Pi^{(b,k)} A_i^{(k)}\right)\varphi_i}{n} \quad . \tag{4-23}$$

If we let

$$[\Pi_{r,c}] = \left[\sum_{j=1}^{p} A_j^{(r)} A_j^{(c)} \varphi_j\right] \quad ,$$

then (4-23) becomes

$$\mathrm{Cov}(\hat{\theta}_a,\hat{\theta}_b) \doteq \sum_{j=1}^{h} \sum_{k=1}^{h} \frac{\Pi^{(a,j)} \Pi^{(b,k)} \Pi_{j,k}}{n} \quad . \tag{4-24}$$

However,

$$\begin{aligned} \sum_{k=1}^{h} \Pi^{(b,k)} \Pi_{j,k} &= 1 \qquad \text{when} \quad j = b \quad ; \\ &= 0 \qquad \text{when} \quad j \neq b \quad . \end{aligned}$$

Therefore (4-24) becomes

$$\mathrm{Cov}(\hat{\theta}_a,\hat{\theta}_b) \doteq \frac{\Pi^{(a,b)}}{n} \quad . \tag{4-25}$$

We shall now consider the covariances for modified moment estimators. Since $\hat{\underline{A}}_s^{(a)}$, as given by equation (4-8), is a function of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_h$, it is therefore a function of $Q_1, Q_2, \ldots, Q_p$. However, the sample moments will contribute Q's also. If m_{r_v} is the sample moment of highest order in $\hat{\underline{A}}_s^{(a)}$, then $\hat{\underline{A}}_s^{(a)}$ is a function of $Q_1, Q_2, \ldots, Q_w$, where $w = \max(r_v, p)$. We can write

$$\hat{\underline{A}}_s^{(a)} = A_s^{(a)} + \sum_{i=1}^{w} K_i^{(a,s)} Q_i + \ldots \quad , \tag{4-26}$$

where $K_i^{(a,s)}$ is a general coefficient. Now we substitute (4-26) and (4-16) into (4-8) and again isolate the coefficients of linear terms in Q_s to get

$$A_s^{(a)} + \sum_{i=1}^{h} \sum_{j=1}^{p} A_j^{(a)} i_o^{(j)} G_s^{(i)} = 0 \quad ,$$

which is identical to (4-17). We conclude, therefore, that the asymptotic covariances of the modified moment estimators are equal to those of corresponding moment estimators (for the same value of p). Moreover, the asymptotic covariances of all modified moment estimators (for the same p) will be the same.

(d) <u>Application to the Negative Binomial Distribution</u>.

We shall now use the methods described at the beginning of this chapter to develop moment and modified moment estimators for the parameters of the negative binomial distribution. Let $\theta_1 = \lambda$ and $\theta_2 = \alpha$. From the expressions for the negative binomial orthogonal polynomials given in Chapter III, we find

$$-E\left[\frac{\partial q_s(x)}{\partial \lambda}\right] = 0 \qquad \text{for } s > 1 \quad ;$$

and that

$$-E\left[\frac{\partial q_s(x)}{\partial \alpha}\right] = 0 \qquad \text{for } s = 1 \quad .$$

So that, from (4-2), our estimating equations are

$$\hat{A}_1^{(1)}\hat{Q}_1 = 0 \tag{4-27}$$

$$\hat{A}_2^{(2)}\hat{Q}_2 + \hat{A}_3^{(2)}\hat{Q}_3 + \ldots + \hat{A}_p^{(2)}\hat{Q}_p = 0 \quad . \tag{4-28}$$

Consider equation (4-27). From (4-5),

$$A_1^{(1)} = \frac{\alpha}{\lambda(\alpha+\lambda)} \quad .$$

Then (4-27) becomes

$$\frac{\hat{\alpha}}{\hat{\lambda}(\hat{\alpha}+\hat{\lambda})}(m_1' - \hat{\lambda}) = 0 \quad .$$

Therefore, $\hat{\lambda} = m_1'$ for all p.

Since it is impossible to obtain an estimator for α at $p = 1$, we first consider the case $p = 2$. Now from (4-5)

$$A_2^{(2)} = -\frac{\alpha}{2(\alpha+\lambda)^2(\alpha+1)} \quad , \tag{4-29}$$

and (4-28) is now

$$\hat{A}_2^{(2)}[m_{(2)} - \frac{m_1'^2(\hat{\alpha}+1)}{\hat{\alpha}}] = 0 \quad .$$

So that

$$\hat{\alpha} = \frac{m_1'^2}{m_{(2)}-m_1'^2} \quad ,$$

$$= \frac{m_1'^2}{m_2-m_1'} \quad .$$

Notice that this is the simple moment estimator which we have discussed previously. Note also that any exchange of sample moments for population moments in $A_1^{(1)}$ or $A_2^{(2)}$ would still not alter the results for $\hat{\lambda}$ or $\hat{\alpha}$. Thus we do not obtain modified moment estimators for λ or for α at $p = 2$.

We now consider the case $p = 3$ and construct the moment estimator, the completely modified moment estimator, and two partially modified moment estimators for α. At $p = 3$, (4-28) becomes

$$\hat{A}_2^{(2)}[m_{(2)} - \frac{m_1'^2(\hat{\alpha}+1)}{\hat{\alpha}}] + \hat{A}_3^{(2)}[m_{(3)} - 3m_1'm_2\frac{(\hat{\alpha}+2)}{\hat{\alpha}} + \frac{2m_1'^3(\hat{\alpha}+2)(\hat{\alpha}+1)}{\hat{\alpha}^2}] = 0 \quad . \qquad (4\text{-}30)$$

We can now express the factorial moments in terms of factorial cumulants, i.e.,

$$m_1' = k_1$$

$$m_{(2)} = k_{(2)} + k_1^2$$

$$m_{(3)} = k_{(3)} + 3k_{(2)}k_1 + k_1^3 \quad .$$

After transposing the first term of (4-30) and dividing both sides by $\hat{A}_3^{(2)}$, we have

$$k_{(3)} - \frac{6k_{(2)}k_1}{\hat{\alpha}} + \frac{4k_1^3}{\hat{\alpha}^2} = \hat{B}(k_{(2)} - \frac{k_1^3}{\hat{\alpha}}) \quad , \qquad (4\text{-}31)$$

where

$$\hat{B} = -\frac{\hat{A}_2^{(2)}}{\hat{A}_3^{(2)}} \quad .$$

From (4-5),

$$A_3^{(2)} = \frac{\alpha^2}{3(\alpha+\lambda)^3(\alpha+2)(\alpha+1)} \quad .$$

Hence

$$\hat{B} = \frac{3}{2}\,\frac{(\hat{\alpha}+\hat{\lambda})(\hat{\alpha}+2)}{\hat{\alpha}} \quad .$$

We find, after some algebraic manipulation of (4-31), that the moment estimator (unmodified) which we shall denote by $\hat{\alpha}_u$, is the solution to

$$C_3\hat{\alpha}_u^3 + C_2\hat{\alpha}_u^2 + C_1\hat{\alpha}_u + C_o = 0 \quad , \tag{4-32}$$

where:

$$C_3 = k_{(2)}$$

$$3C_2 = -2k_{(3)} + 3k_{(2)}(k_1+2) - 3k_1^2$$

$$C_1 = 6k_{(2)}k_1 - k_1^2(k_1+2)$$

$$3C_o = -14k_1^3 \quad .$$

Notice that, in the estimating equation (4-31), the A coefficients are only involved in B. Clearly then, any modification we wish to make must be done only through B. Thus, for the completely modified case, we replace λ by k_1 and α by $k_1^2/k_{(2)}$ (their simple moment estimators). Then

$$\underline{\hat{B}} = \frac{3}{2}\,\frac{(k_1^2/k_{(2)} + k_1)(k_1^2/k_{(2)} + 2)}{k_1^2/k_{(2)}} \quad .$$

Therefore the completely modified estimator $\hat{\alpha}_c$ is the solution to

$$D_2\hat{\alpha}_c^2 + D_1\hat{\alpha}_c + D_o = 0 \quad ; \tag{4-33}$$

where:

$$2D_2 = 2k_{(3)} - 3k_{(2)}(k_1+2) - 3k_1^2 - 6k_{(2)}^2/k_1$$

$$2D_1 = -6k_{(2)}k_1 + 3k_1^2(k_1+2) + 3k_1^4/k_{(2)}$$

$$D_o = 4k_1^3 \quad .$$

In order to obtain partially modified moment estimators, we shall consider B in the form S_1+S_2/α, where S_1 and S_2 are functions of sample moments only and

$$B = (S_1+S_2/\alpha)\Big|_{m_i=\mu_i} \quad . \tag{4-34}$$

Now the estimating equation (4-31) becomes

$$\frac{2t_1^3}{t_2t_3} - \frac{6t_1^3}{\hat{\alpha}t_2} + \frac{4t_1^3}{\hat{\alpha}^2} = (S_1+S_2/\hat{\alpha})\left(\frac{t_1^2}{t_2} - \frac{t_1^2}{\hat{\alpha}}\right) \quad , \tag{4-35}$$

where:

$$t_1 = k_1$$

$$t_2 = k_1^2/k_{(2)}$$

$$t_3 = 2k_1k_{(2)}/k_{(3)} \quad .$$

If we clear fractions, replace S_2 by $-4t$, and divide both sides by $t_1^2\hat{\alpha}$, equation (4-35) becomes linear in $\hat{\alpha}$, i.e.,

$$(2t_1-S_1t_3)\hat{\alpha} = -t_2\left(S_1t_3 - \frac{2t_1t_3}{t_2}\right)$$

or

$$\hat{\alpha} = t_2 + \frac{2t_1(t_2-t_3)}{S_1 t_3 - 2t_1} \quad . \tag{4-36}$$

To obtain S_1 we write

$$S_1 = \left(\frac{4\lambda}{\alpha} + B\right)\Big|_{\lambda=t_1,\alpha=t_2}$$

$$= \frac{4t_1}{t_2} + \frac{3}{2}\,\frac{(t_1+t_2)(t_2+2)}{t_2} \quad . \tag{4-37}$$

For convenience in substituting into (4-36), we can write (4-37) as

$$2(S_1t_3-2t_1) = 3t_1t_3 + 3t_2t_3 + 6t_3 + \frac{14t_1t_3}{t_2} - 4t_1 \quad .$$

Thus we have the partially modified moment estimator $\hat{\alpha}_1$ by solving (4-36), i.e.,

$$\hat{\alpha}_1 = t_2 + \frac{4t_1(t_2-t_3)}{3t_1t_3 + 3t_3(t_2+2) + 14t_1t_3/t_2 - 4t_1} \quad . \tag{4-38}$$

Likewise, we can obtain another partially modified moment estimator by writing

$$S_1 = \left(\frac{4\lambda}{\alpha} + B - \frac{7\lambda}{\alpha}\right)\Big|_{\lambda=t_1,\alpha=t_2} + \frac{7\lambda}{\alpha}\Big|_{\lambda=t_1,\alpha=t_3}$$

$$= \frac{3}{2}t_2 + \frac{3}{2}t_1 + 3 + \frac{7t_1}{t_3}$$

or

$$2(S_1t_3-2t_1) = 3t_2t_3 + 3t_1t_3 + 6t_3 + 10t_1 \quad .$$

Substituting the above into (4-36), we have

$$\hat{\alpha}_2 = t_2 - \frac{4(t_3-t_2)}{10 + 3t_3 + 3t_3(t_2+2)/t_1} \quad . \tag{4-39}$$

Note that, in constructing $\hat{\alpha}_2$, we modify by replacing α by the simple moment estimator t_2 and also by t_3. It might be mentioned that t_3 is the method of moments estimator given in (2-15).

We could construct many other modified and unmodified moment estimators for α. However, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ seem to be the simplest available at $p \leq 3$. It is these two modified moment estimators whose small-sample properties we shall attempt to investigate.

CHAPTER V

PROPERTIES OF TWO MODIFIED MOMENT ESTIMATORS

(a) Computational Technique.

The properties, such as bias, variance, etc., of the modified moment estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were determined with the aid of orthogonal statistics through the procedure described in Chapter III. Thus $\hat{\alpha}_1$ and $\hat{\alpha}_2$, as given in equations (4-38) and (4-39), were first expressed in terms of orthogonal statistics. The next step was to obtain expansions for the estimators similar to that of the form (3-2). This involved use of the expansion

$$(1+\Psi)^{-1} = 1 - \Psi + \Psi^2 - \Psi^3 + \ldots \quad , \qquad (5\text{-}1)$$

where, in this case, $(1+\Psi)$ is the denominator of the estimator expressed in terms of orthogonal statistics. For each estimator, however, Ψ proved to be a multi-termed polynomial of such complexity that expansion by hand would have been extremely difficult (Ψ contained 47 terms for $\hat{\alpha}_1$ and 35 terms for $\hat{\alpha}_2$). It became necessary, therefore, to develop a computational procedure in order that the problem could be solved by an electronic computer. The technique which was ultimately formulated seems general enough to be of use to other investigators. Thus, we shall discuss it in some detail.

Suppose that the moment estimator $\hat{\theta}$ can be expressed in the form

$$\hat{\theta} = \frac{v_o + \underline{V}_1'\underline{Q}}{1 + \underline{V}_2'\underline{Q}} \quad , \tag{5-2}$$

where:

$$\underline{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_1^2 \\ Q_2^2 \\ \vdots \\ Q_1 Q_2 \\ \vdots \\ Q_1^3 \\ \vdots \end{bmatrix}$$

and $\underline{V}_1$, $\underline{V}_2$ are vectors whose elements are functions of the population parameters. The elements of $\underline{V}_1$ and $\underline{V}_2$, and the constant v_o, are all known once $\hat{\theta}$ has been expressed in terms of orthogonal statistics. (For both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ the constant v_o is α.) Now

$$\Psi = \underline{V}_2'\underline{Q} = \underline{Q}'\underline{V}_2 \quad ,$$

and the expansion for $\hat{\theta}$ is given by

$$\hat{\theta} = (v_o+\underline{V}_1'\underline{Q})(1-\Psi+\Psi^2-\Psi^3+\ldots) \quad , \tag{5-3}$$

which converges in the parameter space where $\Psi^2 < 1$. The dimension of the vector $\underline{Q}$ and the highest power of Ψ in (5-3) will depend on how far one wishes to extend the expansion. In this investigation it was desired to extend the expansions to order eight in the powers and products of the Q's. Thus for our purposes, $\underline{Q}$ will contain all products of Q_1, Q_2, and Q_3 through order eight, or 164 elements.

Then

$$\Psi^2 = \underline{V}_2'\underline{Q}\underline{Q}'\underline{V}_2 = \underline{V}_3'\underline{Q}$$

$$\Psi^3 = \underline{V}_2'\underline{Q}\underline{Q}'\underline{V}_3 = \underline{V}_4'\underline{Q}$$

$$\vdots$$

$$\Psi^8 = \underline{V}_2'\underline{Q}\underline{Q}'\underline{V}_8 = \underline{V}_9'\underline{Q} \quad .$$

(These scalars will only contain Q-products through order 8. Q-products of higher order will not be needed.)

Let $$\underline{V}_{10} = \sum_{j=2}^{9} (-1)^{j+1}\, \underline{V}_j \quad .$$

Thus,

$$\begin{aligned}\hat{\theta} &= (v_o + \underline{V}_1'\underline{Q})(1 + \underline{V}_{10}'\underline{Q}) \\ &= v_o + v_o\underline{V}_{10}'\underline{Q} + \underline{V}_1'\underline{Q} + \underline{V}_1'\underline{Q}\underline{Q}'\underline{V}_{10} \\ &= v_o + v_o\underline{V}_{10}'\underline{Q} + \underline{V}_1'\underline{Q} + \underline{V}_{11}'\underline{Q} \\ &= v_o + \underline{V}_{12}'\underline{Q} \quad .\end{aligned}$$

Hence,

$$E(\hat{\theta}) = v_o + \underline{V}_{12}' \, E(\underline{Q}) \quad . \tag{5-4}$$

Furthermore,

$$\begin{aligned}\hat{\theta}^2 &= v_o^2 + 2v_o\underline{V}_{12}'\underline{Q} + \underline{V}_{12}'\underline{Q}\underline{Q}'\underline{V}_{12} \\ &= v_o^2 + 2v_o\underline{V}_{12}'\underline{Q} + \underline{V}_{13}'\underline{Q} \\ &= v_o^2 + \underline{V}_{14}'\underline{Q} \quad .\end{aligned}$$

We can continue the procedure to obtain higher powers of $\hat{\theta}$ and, in doing so, to find higher moments of $\hat{\theta}$.

In order to perform the above routine, we must first determine the vector $\underline{V}_{i+1}$, where

$$\underline{V}_{i+1}'\underline{Q} = \underline{V}_j'\underline{Q}\underline{Q}'\underline{V}_i \quad . \tag{5-5}$$

To do this, it is necessary to perform the matrix multiplication on both sides of (5-5) and then equate coefficients.

The left hand side of (5-5) becomes

$$v_{i+1,1}Q_1 + v_{i+1,2}Q_2 + v_{i+1,3}Q_3 + v_{i+1,4}Q_1^2$$
$$+ v_{i+1,5}Q_2^2 + v_{i+1,6}Q_3^2 + v_{i+1,7}Q_1Q_2 + \cdots$$

and the right hand side is

$$v_{i,1}v_{j,1}Q_1^2 + v_{i,2}v_{j,2}Q_2^2 + v_{i,3}v_{j,3}Q_3^2$$
$$+ v_{i,1}v_{j,2}Q_1Q_2 + v_{i,2}v_{j,1}Q_1Q_2 + \cdots \quad .$$

Hence,

$$v_{i+1,1} = 0$$

$$v_{i+1,2} = 0$$

$$v_{i+1,3} = 0$$

$$v_{i+1,4} = v_{i,1}v_{j,1}$$

$$v_{i+1,5} = v_{i,2}v_{j,2}$$

$$v_{i+1,6} = v_{i,3}v_{j,3}$$

$$v_{i+1,7} = v_{i,1}v_{j,2} + v_{i,2}v_{j,1}$$
$$\vdots \qquad \vdots \qquad \vdots$$

etc.

These relationships will depend somewhat on the definition of $\underline{Q}$, i.e., the arrangement of its elements.

Since $E(\underline{Q})$ is known for the NBD (for expected Q-products through order eight), it was possible to write a computer program, using the vector expansion technique, to find the first four terms of the bias, variance, covariance determinant, third moment and fourth moment for a moment estimator of α. This program, which includes a routine to generate the elements of $\underline{V}_{i+1}$, was executed on an IBM 7040 to obtain the properties of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

1. <u>An Example of the Vector Expansion Technique</u>.

To illustrate the vector expansion technique, we shall obtain the expansion, through Q-products to order three, for the simple moment estimator $\hat{\alpha}$. First, we express $\hat{\alpha}$ as a function of orthogonal statistics, i.e.,

$$\hat{\alpha} = \frac{\alpha + 2\frac{\alpha}{\lambda}Q_1 + \frac{\alpha}{\lambda^2}Q_1^2}{1 + \frac{2}{\lambda}Q_1 + \frac{\alpha}{\lambda^2}Q_2 - \frac{\alpha}{\lambda^2}Q_1^2} \quad .$$

It is now evident that $\hat{\alpha}$ can be put in the form (5-2). Since we are interested in obtaining an expansion through Q-products to order three, we define

$$\underline{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_1^2 \\ Q_2^2 \\ Q_1Q_2 \\ Q_1^3 \\ Q_2^3 \\ Q_1Q_2^2 \\ Q_1^2Q_2 \end{bmatrix} \quad . \tag{5-6}$$

Clearly, $v_o = \alpha$ and

$$\underline{V}_1 = \begin{bmatrix} 2\alpha/\lambda \\ 0 \\ \alpha/\lambda^2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} , \qquad \underline{V}_2 = \begin{bmatrix} 2/\lambda \\ \alpha/\lambda^2 \\ -\alpha/\lambda^2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} .$$

Hence, the expansion for $\hat{\alpha}$ is given by

$$\hat{\alpha} = (\alpha+\underline{V}_1'\underline{Q})(1-\underline{V}_2'\underline{Q}+\underline{V}_3'\underline{Q}-\underline{V}_4'\underline{Q})$$

$$= (\alpha+\underline{V}_1'\underline{Q})(1+\underline{V}_{10}'\underline{Q}) \quad .$$

With $\underline{Q}$ defined as it is in (5-6), we have the following relationships:

$$v_{i+1,1} = 0$$

$$v_{i+1,2} = 0$$

$$v_{i+1,3} = v_{i,1}v_{j,1}$$

$$v_{i+1,4} = v_{i,2}v_{j,2}$$

$$v_{i+1,5} = v_{i,1}v_{j,2} + v_{i,2}v_{j,1}$$

$$v_{i+1,6} = v_{i,1}v_{j,3} + v_{i,3}v_{j,1}$$

$$v_{i+1,7} = v_{i,2}v_{j,4} + v_{i,4}v_{j,2}$$

$$v_{i+1,8} = v_{i,1}v_{j,4} + v_{i,4}v_{j,1} + v_{i,2}v_{j,5} + v_{i,5}v_{j,2}$$

$$v_{i+1,9} = v_{i,1}v_{j,5} + v_{i,5}v_{j,1} + v_{i,2}v_{j,3} + v_{i,3}v_{j,2} .$$

Then since

$$\underline{V}_3'\underline{Q} = \underline{V}_2'\underline{Q}\underline{Q}'\underline{V}_2 \quad ,$$

we have (recall that we shall not need Q-products of order 4 or greater)

$$\underline{V}_3 = \begin{bmatrix} 0 \\ 0 \\ 4/\lambda^2 \\ \alpha^2/\lambda^4 \\ 4\alpha/\lambda^3 \\ -4\alpha/\lambda^3 \\ 0 \\ 0 \\ -2\alpha^2/\lambda^4 \end{bmatrix} .$$

Furthermore,

$$\underline{V}_4'\underline{Q} = \underline{V}_2'\underline{Q}\underline{Q}'\underline{V}_3 ,$$

and

$$\underline{V}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 8/\lambda^3 \\ \alpha^3/\lambda^6 \\ 6\alpha^2/\lambda^5 \\ 12\alpha/\lambda^4 \end{bmatrix} .$$

Therefore,

$$\underline{V}_{10} = \begin{bmatrix} -2/\lambda \\ -\alpha/\lambda^2 \\ (\alpha+4)/\lambda^2 \\ \alpha^2/\lambda^4 \\ 4\alpha/\lambda^3 \\ -4(\alpha+2)/\lambda^3 \\ -\alpha^3/\lambda^6 \\ -6\alpha^2/\lambda^5 \\ -2\alpha(\alpha+6)/\lambda^4 \end{bmatrix} .$$

Now we have

$$\hat{\alpha} = \alpha + \alpha\underline{V}_{10}'\underline{Q} + \underline{V}_1'\underline{Q} + \underline{V}_{11}'\underline{Q} \quad ,$$

where $\underline{V}_{11}$ is defined by

$$\underline{V}_{11}'\underline{Q} = \underline{V}_1'\underline{Q}\underline{Q}'\underline{V}_{10} \quad .$$

Then

$$\underline{V}_{11} = \begin{bmatrix} 0 \\ 0 \\ -4\alpha/\lambda^2 \\ 0 \\ -2\alpha^2/\lambda^3 \\ 2\alpha(\alpha+3)/\lambda^3 \\ 0 \\ 2\alpha^3/\lambda^5 \\ 7\alpha^2/\lambda^4 \end{bmatrix} .$$

Let

$$\underline{V}_{12} = \alpha\underline{V}_{10} + \underline{V}_{1} + \underline{V}_{11} \quad .$$

Hence

$$\hat{\alpha} = \alpha + \underline{V}_{12}'\underline{Q} \quad .$$

Since

$$\underline{V}_{12} = \begin{bmatrix} 0 \\ -\alpha^2/\lambda^2 \\ \alpha(\alpha+1)/\lambda^2 \\ \alpha^3/\lambda^4 \\ 2\alpha^2/\lambda^3 \\ -2\alpha(\alpha+1)/\lambda^3 \\ -\alpha^4/\lambda^6 \\ -4\alpha^2/\lambda^5 \\ -\alpha^2(2\alpha+5)/\lambda^4 \end{bmatrix} ,$$

we have the expansion

$$\hat{\alpha} = \alpha - \frac{\alpha^2}{\lambda^2}Q_2 + \frac{\alpha(\alpha+1)}{\lambda^2}Q_1^2 + \frac{\alpha^3}{\lambda^4}Q_2^2 + 2\frac{\alpha^2}{\lambda^3}Q_1Q_2$$
$$- \frac{2\alpha(\alpha+1)}{\lambda^3}Q_1^3 - \frac{\alpha^4}{\lambda^6}Q_2^3 - 4\frac{\alpha^2}{\lambda^5}Q_1Q_2^2 - \frac{\alpha^2(2\alpha+5)}{\lambda^4}Q_1^2Q_2 \quad .$$

From the example, one can see that this procedure can be readily adapted to more involved problems and facilitated by use of a computer. Of course, for computer solution, we must deal with specific values of α and λ. The properties of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were evaluated in the parameter subspace $(1 \leq \alpha \leq 100,\ 1 \leq \lambda \leq 100)$. We shall now discuss these results.

(b) <u>Biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$</u>.

The first four terms of

$$E(\hat{\alpha}_k - \alpha) = \frac{B_1^{(k)}}{n} + \frac{B_2^{(k)}}{n^2} + \frac{B_3^{(k)}}{n^3} + \frac{B_4^{(k)}}{n^4} + \ldots, \quad (k=1,2) \tag{5-7}$$

were determined numerically. The results for the bias of $\hat{\alpha}_1$ are presented in Table II and those for the bias of $\hat{\alpha}_2$ are presented in Table III for, in both cases, a sample size of 100.

As was the case with the maximum likelihood estimator and the simple moment estimator, the asymptotic bias (n^{-1} term) seems to be a poor approximation to the actual bias, especially in the $\alpha > \lambda$ region. Even with the use of all four terms, we can only hope to get a good assessment of the biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for certain values of α and λ. In the case of $\hat{\alpha}_1$, one such point would be $\alpha=2$, $\lambda=1$. Here we see the expansion has diminished considerably at the fourth term.

Table II.

Terms in Expansion of $E(\hat{\alpha}_1-\alpha)$

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.1252	.0620	.0456	.0382	.0341	.0265	.0242	.0224	.0210	.0204
	(II)	.0355	.0108	.0072	.0060	.0054	.0044	.0042	.0040	.0038	.0038
	(III)	.0177	.0044	.0030	.0025	.0023	.0018	.0017	.0016	.0016	.0015
	(IV)	.3511	.1175	.0760	.0601	.0518	.0380	.0340	.0311	.0313	.0878
2	(I)	.5157	.2180	.1458	.1149	.0980	.0680	.0592	.0525	.0477	.0454
	(II)	.2855	.0477	.0212	.0134	.0100	.0055	.0045	.0039	.0035	.0033
	(III)	.2529	.0164	.0050	.0026	.0018	.0009	.0007	.0006	.0004	.0005
	(IV)	.0214	.0516	.0298	.0213	.0171	.0105	.0088	.0076	.0067	.0052
3	(I)	1.291	.4980	.3142	.2375	.1963	.1251	.1048	.0897	.0790	.0740
	(II)	1.252	.1684	.0633	.0350	.0235	.0095	.0068	.0052	.0043	.0039
	(III)	1.971	.0897	.0196	.0078	.0042	.0011	.0007	.0005	.0004	.0004
	(IV)	3.946	.0189	.0130	.0114	.0094	.0054	.0044	.0036	.0031	.0031
4	(I)	2.570	.9305	.5624	.4120	.3325	.1979	.1604	.1329	.1139	.1049
	(II)	3.875	.4574	.1159	.0798	.0502	.0166	.0107	.0074	.0056	.0048
	(III)	9.564	.3596	.0675	.0236	.0114	.0020	.0010	.0006	.0004	.0003
	(IV)	32.23	.3440	.0226	.0010	.0041	.0034	.0027	.0022	.0018	.0017
5	(I)	4.470	1.544	.9031	.6452	.5106	.2869	.2258	.1819	.1519	.1378
	(II)	9.684	1.044	.3319	.1605	.0965	.0274	.0163	.0104	.0072	.0059
	(III)	34.52	1.140	.1933	.0621	.0278	.0037	.0016	.0008	.0004	.0003
	(IV)	170.3	1.666	.1381	.0239	.0048	.0020	.0018	.0015	.0012	.0011
10	(I)	27.48	8.372	4.447	2.944	2.188	1.0080	.7126	.5116	.3816	.3233
	(II)	199.5	17.15	4.542	1.887	.9969	.1816	.0833	.0392	.0201	.0138
	(III)	2406	57.97	7.596	1.963	.7298	.0501	.0142	.0041	.0013	.0006
	(IV)	4056.[1]	272.8	17.58	2.808	.7291	.0175	.0026	.0001	.0002	.0003

Table II (continued)

α		λ = 1	2	3	4	5	10	15	25	50	100
15	(I)	83.99	24.20	12.27	7.809	5.610	2.293	1.507	.9943	.6770	.5440
	(II)	1290.	101.0	24.66	9.541	4.730	.6808	.2674	.1037	.0424	.0249
	(III)	3297.^{1}	698.9	81.88	19.16	6.520	.3202	.0727	.0156	.0035	.0013
	(IV)	1179.^{3}	6758.	378.9	53.51	12.46	.2054	.0265	.0030	.0002	.0000
25	(I)	357.5	97.87	47.43	28.98	20.08	7.145	4.279	2.507	1.481	1.066
	(II)	1450.^{1}	1044.	236.5	85.38	39.72	4.436	1.451	.4391	.1303	.0596
	(III)	9796.^{2}	1854.^{1}	1959.	417.0	130.0	4.483	.7863	.1182	.0162	.0042
	(IV)	9264.^{4}	4606.^{2}	2269.^{1}	2845.	593.5	6.281	.5866	.0434	.0027	.0004
50	(I)	2678.	702.0	326.8	192.4	128.7	39.58	21.21	10.59	5.007	2.980
	(II)	4178.^{2}	2811.^{1}	5969.	2028.	890.3	77.42	20.72	4.629	.8690	.2596
	(III)	1087.^{5}	1875.^{3}	1815.^{2}	3556.^{1}	1024.^{1}	250.5	33.27	3.267	.2323	.0321
	(IV)	3956.^{7}	1751.^{5}	7728.^{3}	8730.^{2}	1649.^{2}	1132.	74.43	3.198	.0853	.0055
100	(I)	2070.^{1}	5304.	2415.	1391.	911.4	255.2	162.2	55.38	2.109	10.01
	(II)	1267.^{4}	8223.^{2}	1686.^{2}	5535.^{1}	2351.^{1}	1755.	410.6	73.15	9.174	1.729
	(III)	1293.^{7}	2125.^{5}	1961.^{4}	3669.^{3}	1010.^{3}	2007.^{1}	2217.	159.3	6.446	.4607
	(IV)	1846.^{10}	7686.^{7}	3195.^{6}	3405.^{5}	6078.^{4}	3211.^{2}	1674.^{1}	484.7	6.294	.1689

Key to Table II -- (a) (I), (II), (III), (IV) refer to the four terms in (5-7) when n=100.

(b) Underlined entries in this table are negative.

(c) Indices in this table are to be taken as the power of 10 which multiplies the entry; thus 4056.^{1} = 40560.

Table III.

Terms in Expansion of $E(\hat{\alpha}_2 - \alpha)$

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.0235	.0100	.0167	.0192	.0204	.0222	.0227	.0230	.0232	.0232
	(II)	.4054	.2367	.1919	.1714	.1596	.1372	.1301	.1246	.1205	.1185
	(III)	4.338	2.041	1.506	1.276	1.149	.9196	.8501	.7970	.7585	.7387
	(IV)	102.7	37.87	25.34	20.37	17.74	13.23	11.94	10.96	10.34	9.288
2	(I)	.3267	.0980	.0486	.0290	.0189	.0023	.0021	.0053	.0074	.0084
	(II)	.6305	.2349	.1633	.1346	.1192	.0919	.0837	.0775	.0730	.0707
	(III)	2.931	1.109	.7150	.5560	.4721	.3286	.2876	.2571	.2356	.2251
	(IV)	52.00	13.13	7.271	5.201	4.186	2.592	2.174	1.875	1.670	1.624
3	(I)	1.011	.3294	.1824	.1241	.0939	.0446	.0315	.0221	.0158	.0129
	(II)	1.550	.3348	.1871	.1384	.1151	.0784	.0685	.0612	.0561	.0537
	(III)	1.006	.7853	.4905	.3654	.2996	.1894	.1590	.1369	.1215	.1142
	(IV)	43.84	8.147	3.984	2.636	2.008	1.085	.8597	.7047	.6020	.5597
4	(I)	2.197	.7132	.3968	.2720	.2079	.1040	.0765	.0572	.0443	.0383
	(II)	4.081	.6020	.2659	.1716	.1310	.0757	.0628	.0539	.0479	.0452
	(III)	6.439	.4141	.3511	.2688	.2190	.1311	.1066	.0889	.0768	.0712
	(IV)	66.98	6.507	2.804	1.739	1.268	.6137	.4639	.3640	.2997	.2712
5	(I)	4.004	1.279	.7040	.4793	.3645	.1804	.1325	.0991	.0769	.0667
	(II)	9.740	1.161	.4284	.2426	.1692	.0800	.0622	.0508	.0436	.0403
	(III)	30.88	.4000	.1791	.1875	.1633	.0991	.0790	.0643	.0543	.0496
	(IV)	201.0	6.828	2.327	1.333	.9304	.4084	.2962	.2236	.1780	.1565
10	(I)	26.55	7.868	4.086	2.653	1.941	.8459	.5789	.4006	.2877	.2378
	(II)	197.6	16.97	4.531	1.912	1.032	.2169	.1146	.0662	.0435	.0353
	(III)	2385.	56.57	7.176	1.747	.5887	.0046	.0235	.0230	.0192	.0169
	(IV)	4037.[1]	273.8	18.69	3.468	1.166	.1648	.0950	.0603	.0415	.0340

Table III (continued)

α		λ = 1	2	3	4	5	10	15	25	50	100
15	(I)	82.59	23.46	11.75	7.398	5.265	2.080	1.338	.8603	.5692	.4454
	(II)	1283.	100.1	24.42	9.458	4.703	7.022	2.908	.1253	.0612	.0420
	(III)	$3285.^{1}$	693.6	80.86	18.79	6.326	.2732	.0444	$\underline{.0025}$	$\underline{.0091}$	$\underline{.0088}$
	(IV)	$1176.^{3}$	6728.	377.4	53.61	12.68	.3010	.0825	.0361	.0209	.0158
25	(I)	355.2	96.66	46.59	28.33	19.54	6.835	4.044	2.331	1.351	.9579
	(II)	$1447.^{1}$	1040.	235.2	84.82	39.44	4.415	1.458	.4535	.1449	.0729
	(III)	$9782.^{2}$	$1849.^{1}$	1951.	414.8	129.1	4.409	.7576	.1046	.0085	$\underline{.0014}$
	(IV)	$9254.^{4}$	$4597.^{2}$	$2262.^{1}$	2835.	591.1	6.301	.6172	.0620	.0130	.0075
50	(I)	2673.	699.6	325.2	191.1	127.7	39.03	20.81	10.32	4.826	2.846
	(II)	$4176.^{2}$	$2808.^{1}$	5959.	2023.	888.1	77.11	20.63	4.616	.8769	.2696
	(III)	$1086.^{5}$	$1874.^{3}$	$1814.^{2}$	$3552.^{1}$	$1023.^{1}$	249.8	33.12	3.242	.2262	.0288
	(IV)	$3955.^{7}$	$1751.^{5}$	$7722.^{3}$	$8721.^{2}$	$1647.^{2}$	1130.	7421	3.192	.0900	.0087
100	(I)	$2070.^{1}$	5299.	2412.	1389.	909.5	254.2	125.5	54.91	20.81	9.823
	(II)	$1267.^{4}$	$8221.^{2}$	$1685.^{2}$	$5532.^{1}$	$2350.^{1}$	1752.	409.9	72.98	9.156	1.734
	(III)	$1292.^{7}$	$2124.^{5}$	$1961.^{4}$	$3668.^{3}$	$1010.^{3}$	$2085.^{1}$	2214.	159.1	6.426	.4572
	(IV)	$1846.^{10}$	$7685.^{7}$	$3194.^{6}$	$3404.^{5}$	$6076.^{4}$	$3210.^{2}$	$1672.^{1}$	484.0	6.279	.1698

Key to Table III -- (a) (I), (II), (III), (IV) refer to the four terms in (5-7) when n=100.

(b) Underlined entries in this table are negative.

(c) Indices in this table are to be taken as the power of 10 which multiplies the entry; thus $4037.^{1} = 40370$.

It must be mentioned that we are assuming, in regions of the parameter space where the expansion diminishes, that the expansions (5-7) converge to the biases of $\hat{\alpha}_1$ and $\hat{\alpha}_2$. It is very likely that for the $\alpha > \lambda$ region the expansions do not converge. (This would be impossible to show in general since one cannot hope to obtain a general term in (5-7).) However, since we have no satisfactory approximations in this region, the possible divergence of the expansions is not a critical issue. In the regions of the parameter space for which the terms in the expansions do diminish, the matter of convergence is important.

One could certainly generate (for any α and λ) a sample for which the expansion of the estimator, i.e., (5-3), does not converge. However, this does not necessarily mean that the resulting moment expansion (i.e., after one takes expected values in (5-3)) does not converge. To support the validity of the moment expansions, a Monte Carlo simulation study was done at a point in the parameter space where the expansions were well behaved. The results, which will be discussed at the end of this chapter, strongly supported the expansion approximations.

(c) <u>Variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.</u>

Numerical values for the first four terms of the expansion

$$\mathrm{Var}(\hat{\alpha}_k) = \frac{V_1^{(k)}}{n} + \frac{V_2^{(k)}}{n^2} + \frac{V_3^{(k)}}{n^3} + \frac{V_4^{(k)}}{n^4} + \cdots \quad , \qquad (k = 1,2) \qquad (5\text{-}8)$$

were determined. The results, for a sample size of 100, are presented in Tables IV and V for $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively.

Again we see an "explosive" type of behavior in the expansions, especially when $\alpha \gg \lambda$. Also, in the region where α is small, for both $\hat{\alpha}_1$ and $\hat{\alpha}_2$, we find that the behavior of the variance expansions is erratic. In fact, if we were to use the sum of the first four terms as an approximation, we would obtain a negative value for the variance at some α and λ. Obviously, for the region where α is small, it would be necessary to extend the expansion further or to use a larger sample size to get a good approximation of the variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

It is interesting to note that the asymptotic term of $\mathrm{Var}(\hat{\alpha}_1)$ and the asymptotic term of $\mathrm{Var}(\hat{\alpha}_2)$ are equal. This fact is in accordance with the theory discussed in Chapter IV.

Table IV.

Terms in the Expansion of Var($\hat{\alpha}_1$)

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.1309	.0694	.0533	.0461	.0420	.0345	.0321	.0303	.0290	.0283
	(II)	.0803	.0153	.0063	.0033	.0019	.0001	.0006	.0010	.0012	.0013
	(III)	.0561	.0052	.0021	.0013	.0011	.0009	.0009	.0009	.0009	.0008
	(IV)	1.227	.3688	.2255	.1724	.1454	.1013	.0892	.0805	.0854	.3099
2	(I)	.9720	.4114	.2778	.2209	.1900	.1352	.1191	.1070	.0983	.0941
	(II)	1.506	.2391	.0985	.0572	.0393	.0156	.0106	.0073	.0053	.0044
	(III)	2.378	.1409	.0359	.0157	.0090	.0026	.0018	.0013	.0011	.0010
	(IV)	1.501	.3770	.2112	.1448	.1125	.0635	.0514	.0430	.0376	.0245
3	(I)	3.600	1.355	.8470	.6378	.5266	.3366	.2828	.2432	.2154	.2022
	(II)	9.952	1.316	.4849	.2616	.1709	.0604	.0392	.0265	.0190	.0158
	(III)	28.04	1.270	.2713	.1040	.0536	.0109	.0058	.0034	.0023	.0019
	(IV)	82.55	.6443	.1138	.1212	.1003	.0538	.0413	.0328	.0273	.0289
4	(I)	9.574	3.352	1.988	1.440	1.154	.6766	.5460	.4517	.3869	.3565
	(II)	41.08	4.792	1.617	.8182	.5089	.1571	.0959	.0611	.0416	.0337
	(III)	180.9	6.834	1.284	.4466	.2131	.0338	.0157	.0800	.0046	.0035
	(IV)	887.9	9.635	.7349	.0628	.0375	.0462	.0357	.0277	.0224	.0203
5	(I)	20.94	6.971	3.982	2.800	2.191	1.198	.9333	.7456	.6188	.5599
	(II)	128.4	13.69	4.322	2.076	1.239	.3384	.1944	.1165	.0750	.0589
	(III)	813.3	27.03	4.610	1.486	.6666	.0870	.0363	.0163	.0083	.0058
	(IV)	5845.	57.55	4.852	.8949	.2266	.0310	.0297	.0239	.0192	.0168
10	(I)	263.5	77.76	40.28	26.12	19.09	8.337	5.729	3.995	2.900	2.418
	(II)	5308.	452.7	119.2	49.31	25.95	4.663	2.115	.9770	.4844	.3227
	(III)	1123.[2]	2719.	358.2	93.05	34.77	2.438	.7037	.2055	.0661	.0340
	(IV)	2750.[3]	1867.[1]	1214.	195.3	51.04	1.275	.2039	.0222	.0048	.0075

Table IV (continued)

α	λ	1	2	3	4	5	10	15	25	50	100
15	(I)	1223.	343.6	170.6	106.5	75.32	29.09	18.48	11.71	7.650	5.943
	(II)	$5153.^{1}$	4011.	975.3	375.8	185.7	26.41	10.29	3.940	1.579	.9075
	(III)	$2299.^{3}$	$4892.^{1}$	5752.	1352.	461.6	23.10	5.329	1.173	.2694	.1083
	(IV)	$1192.^{5}$	$6884.^{2}$	$3888.^{1}$	5528.	1295.	21.95	2.894	.3406	.0356	.0051
25	(I)	8771.	2361.	1126.	678.7	464.2	157.0	90.76	50.75	28.32	19.54
	(II)	$9661.^{2}$	$6934.^{1}$	$1565.^{1}$	5636.	2616.	289.2	93.92	28.13	8.219	3.695
	(III)	$1134.^{5}$	$2152.^{3}$	$2279.^{2}$	$4864.^{1}$	$1520.^{1}$	531.0	94.25	14.47	2.061	.5620
	(IV)	$1551.^{7}$	$7755.^{4}$	$3839.^{3}$	$4838.^{2}$	$1014.^{2}$	1098.	104.6	7.985	.5287	.0833
50	(I)	$1326.^{2}$	$3444.^{1}$	$1589.^{1}$	9277.	6157.	1828.	952.0	455.1	201.4	112.8
	(II)	$5570.^{4}$	$3741.^{3}$	$7931.^{2}$	$2690.^{2}$	$1179.^{2}$	$1019.^{1}$	2712.	600.8	111.2	32.75
	(III)	$2508.^{7}$	$4335.^{5}$	$4202.^{4}$	$8244.^{3}$	$2378.^{3}$	$5855.^{1}$	7828.	778.8	56.92	8.193
	(IV)	$1318.^{10}$	$5854.^{7}$	$2590.^{6}$	$2934.^{5}$	$5559.^{4}$	$3870.^{2}$	$2576.^{1}$	1133.	31.65	2.150
100	(I)	$2060.^{3}$	$5253.^{2}$	$2380.^{2}$	$1365.^{2}$	$8903.^{1}$	$2441.^{1}$	$1185.^{1}$	5037.	1810.	802.8
	(II)	$3379.^{6}$	$2191.^{5}$	$4488.^{4}$	$1472.^{4}$	$6294.^{3}$	$4646.^{2}$	$1084.^{2}$	$1919.^{1}$	2380.	442.5
	(III)	$5958.^{9}$	$9800.^{7}$	$9053.^{6}$	$1695.^{6}$	$4670.^{5}$	$9310.^{3}$	$1032.^{3}$	$7472.^{1}$	3073.	225.8
	(IV)	$1227.^{13}$	$5117.^{10}$	$2130.^{9}$	$2274.^{8}$	$4065.^{7}$	$2164.^{5}$	$1136.^{4}$	$3334.^{2}$	4466.	126.0

Key to Table IV -- (a) (I), (II), (IV) refer to the four terms in (5-8) when n=100.

(b) Underlined entries are negative.

(c) Indices imply multiplication by the corresponding power of ten.

Table V.

Terms in the Expansion of Var($\hat{\alpha}_2$)

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.1309	.0694	.0533	.0461	.0420	.0345	.0321	.0303	.0290	.0283
	(II)	.0167	.0368	.0344	.0323	.0308	.0275	.0263	.0253	.0246	.0242
	(III)	2.461	1.146	.8443	.7147	.6433	.5142	.4750	.4451	.4234	.4109
	(IV)	96.03	35.64	23.87	19.18	16.70	12.45	11.22	10.28	9.786	8.200
2	(I)	.9720	.4114	.2778	.2209	.1900	.1352	.1191	.1070	.0983	.0941
	(II)	1.196	.1260	.0239	.0019	.0117	.0213	.0223	.0225	.0224	.0222
	(III)	5.310	1.275	.7751	.5933	.5007	.3460	.3023	.2700	.2472	.2359
	(IV)	81.56	22.27	12.50	8.986	7.252	4.508	3.784	3.266	2.909	2.940
3	(I)	3.600	1.355	.8470	.6378	.5266	.3366	.2828	.2432	.2155	.2022
	(II)	9.109	1.090	.3578	.1699	.0967	.0137	.0001	.0076	.0115	.0130
	(III)	30.26	2.376	.9694	.6292	.4870	.2881	.2392	.2046	.1811	.1700
	(IV)	1.598	17.10	9.076	6.145	4.733	2.601	2.071	1.703	1.457	1.372
4	(I)	9.574	3.352	1.988	1.440	1.154	.6766	.5460	.4517	.3869	.3565
	(II)	39.10	4.358	1.404	.6778	.4022	.0992	.0498	.0230	.0089	.0034
	(III)	178.3	7.702	1.914	.9260	.6057	.2729	.2113	.1718	.1466	.1350
	(IV)	780.0	6.222	6.588	4.617	3.511	1.788	1.367	1.082	.8964	.8146
5	(I)	20.94	6.971	3.982	2.800	2.191	1.198	.9333	.7456	.6188	.5599
	(II)	124.3	12.90	3.972	1.862	1.086	.2666	.1406	.0744	.0403	.0274
	(III)	795.8	27.16	5.083	1.897	1.015	.2992	.2068	.1560	.1268	.1142
	(IV)	5626.	41.49	1.589	3.005	2.578	1.336	.9948	.7637	.6153	.5346
10	(I)	263.5	77.76	40.28	26.12	19.09	8.337	5.729	3.995	2.900	2.418
	(II)	5258	445.2	116.6	47.97	25.13	4.442	1.991	.9040	.4379	.2861
	(III)	1114.[2]	2679.	351.7	91.30	34.19	2.521	.7986	.2874	.1329	.0928
	(IV)	2729.[3]	1839.[1]	1185.	188.2	48.14	.6968	.1566	.2224	.1830	.1578

Table V (continued)

α		λ = 1	2	3	4	5	10	15	25	50	100
15	(I)	1223.	343.6	170.6	106.5	75.32	29.09	18.48	11.71	7.650	5.943
	(II)	$5130.^{1}$	3979.	964.5	370.7	182.7	25.78	9.994	3.804	1.512	.8634
	(III)	$2290.^{3}$	$4855.^{1}$	5693.	1335.	455.2	22.77	5.308	1.216	.3149	.1490
	(IV)	$1188.^{5}$	$6836.^{2}$	$3849.^{1}$	5456.	1274.	21.06	2.609	.2073	$\underline{.0486}$	$\underline{.0629}$
25	(I)	8771.	2361.	1126.	678.7	464.2	157.0	90.76	50.75	28.32	19.54
	(II)	$9644.^{2}$	$6912.^{1}$	$1558.^{1}$	5604.	2598.	286.1	92.71	27.69	8.070	3.624
	(III)	$1132.^{5}$	$2146.^{3}$	$2270.^{2}$	$4838.^{1}$	$1510.^{1}$	525.8	93.21	14.33	2.070	.5846
	(IV)	$1549.^{7}$	$7735.^{4}$	$3825.^{3}$	$4814.^{2}$	$1008.^{2}$	1085.	102.7	7.726	.4822	.0601
50	(I)	$1326.^{2}$	$3444.^{1}$	$1589.^{1}$	9277.	6157.	1828.	952.0	455.1	201.4	112.8
	(II)	$5568.^{4}$	$3738.^{3}$	$7921.^{2}$	$2685.^{2}$	$1177.^{2}$	$1015.^{1}$	2700.	597.3	110.4	32.50
	(III)	$2507.^{7}$	$4332.^{5}$	$4197.^{4}$	$8231.^{3}$	$2373.^{3}$	$5836.^{1}$	7795.	774.5	56.57	8.159
	(IV)	$1318.^{10}$	$5850.^{7}$	$2588.^{6}$	$2931.^{5}$	$5550.^{4}$	$3858.^{2}$	$2564.^{1}$	1124.	31.11	2.080
100	(I)	$2060.^{3}$	$5253.^{2}$	$2380.^{2}$	$1365.^{2}$	$8903.^{1}$	$2441.^{1}$	$1185.^{1}$	5037.	1810.	802.8
	(II)	$3379.^{6}$	$2190.^{5}$	$4487.^{4}$	$1472.^{4}$	$6246.^{3}$	$4642.^{2}$	$1082.^{2}$	$1915.^{1}$	2374.	440.9
	(III)	$5957.^{9}$	$9798.^{7}$	$9050.^{6}$	$1694.^{6}$	$4668.^{5}$	$9302.^{3}$	$1031.^{3}$	$7457.^{1}$	3064.	225.1
	(IV)	$1227.^{13}$	$5116.^{10}$	$2130.^{9}$	$2273.^{8}$	$4063.^{7}$	$2162.^{5}$	$1134.^{4}$	$3328.^{2}$	4448.	124.8

Key to Table V -- (a) (I), (II), (III), (IV) refer to the four terms in (5-8) when n=100.

(b) Underlined entries are negative.

(c) Indices imply multiplication by the corresponding power of ten.

(d) <u>Covariance Determinants for $\hat{\alpha}_1$ and $\hat{\alpha}_2$.</u>

The expansions

$$\begin{vmatrix} Var(\hat{\alpha}_k) & Cov(\hat{\alpha}_k,\hat{\lambda}) \\ Cov(\hat{\alpha}_k,\hat{\lambda}) & Var(\lambda) \end{vmatrix} = \frac{C_2^{(k)}}{n^2} + \frac{C_3^{(k)}}{n^3} + \frac{C_4^{(k)}}{n^4} + \cdots \quad , \qquad (k = 1,2) \qquad (5\text{-}9)$$

were obtained numerically through the n^{-4} term. For this work, the estimator for λ was the sample mean $\bar{x}$ (or m_1'). Hereafter, we shall denote the covariance determinants more simply by $|Cov(\hat{\alpha}_k,\bar{x})|$. The results for the covariance determinants are given in Tables VI and VII for $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively, for a sample of 100.

As in the case of the bias and variance expansions, we note chaotic behavior when $\alpha \gg \lambda$. However, the expansion of $|Cov(\hat{\alpha}_1,\bar{x})|$ appears to be better behaved in the region of small α than does the expansion of $|Cov(\hat{\alpha}_2,\bar{x})|$. For example at $\alpha=1$, $\lambda=3$,

$$|Cov(\hat{\alpha}_1,\bar{x})| = \frac{64}{n^2} + \frac{800}{n^3} + \frac{20000}{n^4} + \cdots \quad ;$$

and

$$|Cov(\hat{\alpha}_2,\bar{x})| = \frac{64}{n^2} - \frac{4100}{n^3} + \frac{10130000}{n^4} + \cdots \quad .$$

Similar data for the third and fourth moments of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were obtained. These results are given in Appendices A and B.

Table VI.

Terms in the Expansion of $|Cov(\hat{\alpha}_1, \bar{x})|$

α \ λ		1	2	3	4	5	10	15	25	50	100
	(I)	.0026	.0042	.0064	.0092	.0126	.0379	.0771	.1970	.7391	2.862
1	(II)	.0016	.0009	.0008	.0006	.0006	.0002	.0015	.0063	.0302	.1295
	(III)	.0011	.0003	.0002	.0003	.0003	.0010	.0021	.0058	.0228	.0858
	(I)	.0146	.0164	.0208	.0265	.0032	.0811	.1518	.3610	1.278	4.799
2	(II)	.0226	.0096	.0074	.0069	.0069	.0094	.0134	.0246	.0686	.2238
	(III)	.0356	.0056	.0027	.0019	.0016	.0016	.0023	.0045	.0147	.0529
	(I)	.0480	.0452	.0508	.0595	.0702	.1458	.2545	.5674	1.903	6.944
3	(II)	.1327	.0439	.0291	.0244	.0228	.0262	.0353	.0618	.1676	.5423
	(III)	.3732	.0422	.0162	.0097	.0071	.0047	.0052	.0080	.0206	.0666
	(I)	.1196	.1006	.1044	.1152	.1298	.2368	.3890	.8187	2.612	9.268
4	(II)	.5135	.1438	.0849	.0654	.0572	.0550	.0683	.1107	.2808	.8763
	(III)	2.258	.2047	.0673	.0356	.0239	.0118	.0112	.0144	.0313	.0914
	(I)	.2512	.1952	.1911	.2016	.2191	.3594	.5600	1.118	3.404	11.76
5	(II)	1.541	.3834	.2074	.1495	.1239	.1015	.1167	.1747	.4128	1.236
	(III)	9.752	.7559	.2210	.1068	.0666	.0261	.0218	.0244	.0458	.1228
	(I)	2.899	1.866	1.571	1.463	1.432	1.667	2.148	3.496	8.698	26.60
10	(II)	58.38	10.86	4.650	2.761	1.946	.9327	.7933	.8549	1.453	3.549
	(III)	1235.	65.23	13.96	5.208	2.606	.4873	.2637	.1797	.1981	.3740
	(I)	13.04	7.789	6.140	5.398	5.022	4.848	5.543	7.809	16.58	45.56
15	(II)	549.6	90.92	35.11	19.04	12.38	4.401	3.086	2.627	3.421	6.958
	(III)	2452.[1]	1108.	207.0	68.46	30.76	3.848	1.598	.7820	.5835	.8305

Table VI (continued)

α	λ	1	2	3	4	5	10	15	25	50	100
	(I)	91.22	50.99	37.85	31.49	27.85	21.98	21.78	25.37	42.48	97.70
25	(II)	1005.^{1}	1498.	526.0	261.5	156.9	40.48	22.54	14.07	12.33	18.48
	(III)	1180.^{3}	4648.^{1}	7658.	2257.	912.0	74.32	22.61	7.234	3.091	2.810
	(I)	1352.	716.4	505.4	400.8	338.6	219.4	185.6	170.7	201.4	338.5
50	(II)	5682.^{2}	7782.^{1}	2522.^{1}	1162.^{1}	6486.	1222.	528.9	225.3	111.2	98.26
	(III)	2559.^{5}	9017.^{3}	1336.^{3}	3561.^{2}	1308.^{2}	7026.	1526.	292.0	56.92	24.58
	(I)	2081.^{1}	1072.^{1}	7355.	5678.	4674.	2685.	2044.	1574.	1358.	1606.
100	(II)	3413.^{4}	4470.^{3}	1387.^{3}	6124.^{2}	3208.^{2}	5111.^{1}	1869.^{1}	5997.	1785.	884.9
	(III)	6017.^{7}	1999.^{6}	2797.^{5}	7050.^{4}	2452.^{4}	1024.^{3}	1780.^{2}	2335.^{1}	2305.	451.6

Key to Table VI -- (a) (I), (II), (III) refer to the three terms in (5-9) when n=100.

(b) Underlined entries are negative.

(c) Indices imply multiplication by the corresponding power of ten.

Table VII.

Terms in the Expansion of $|\mathrm{Cov}(\hat{\alpha}_2, \bar{x})|$

α \ λ		1	2	3	4	5	10	15	25	50	100
	(I)	.0026	.0042	.0064	.0092	.0126	.0379	.0771	.1970	.7391	2.861
1	(II)	.0003	.0022	.0041	.0064	.0092	.0302	.0631	.1644	.6265	2.444
	(III)	.0492	.0688	.1013	.1429	.1930	.5656	1.140	2.893	10.80	41.50
	(I)	.0146	.0164	.0208	.0265	.0332	.0811	.1518	.3610	1.278	4.799
2	(II)	.0179	.0050	.0018	.0002	.0020	.0128	.0285	.0761	.2912	1.134
	(III)	.0796	.0510	.0581	.0712	.0876	.2076	.3855	.9113	3.214	12.03
	(I)	.0480	.0452	.0508	.0595	.0702	.1458	.2545	.5674	1.903	6.944
3	(II)	.1214	.0363	.0215	.0159	.0129	.0060	.0001	.0178	.1019	.4466
	(III)	.4029	.0791	.0581	.0587	.0649	.1248	.2152	.4775	1.600	5.837
	(I)	.1197	.1006	.1044	.1152	.1298	.2368	.3890	.8187	2.612	9.268
4	(II)	.4887	.1308	.0737	.0542	.0452	.0347	.0355	.0417	.0600	.0892
	(III)	2.226	.2308	.1004	.0740	.0681	.0955	.1506	.3115	.9892	3.512
	(I)	.2512	.1952	.1911	.2016	.2191	.3594	.5600	1.118	3.404	11.76
5	(II)	1.492	.3613	.1906	.1341	.1086	.0800	.0844	.1115	.2218	.5765
	(III)	9.542	.7599	.2437	.1365	.1014	.0898	.1240	.2340	.6972	2.399
	(I)	2.899	1.866	1.571	1.463	1.432	1.667	2.148	3.496	8.698	26.60
10	(II)	57.84	10.68	4.546	2.686	1.885	.8883	.7465	.7910	1.314	3.147
	(III)	1225.	64.28	13.71	5.110	2.562	.5039	.2993	.2514	.3986	1.020
	(I)	13.04	7.789	6.140	5.398	5.022	4.848	5.543	7.809	16.58	45.56
15	(II)	547.2	90.19	34.72	18.78	12.18	4.297	2.998	2.536	3.276	6.619
	(III)	2442.[1]	1100.	204.9	67.61	30.34	3.793	1.592	.8106	.6821	1.142

Table VII (continued)

α \ λ		1	2	3	4	5	10	15	25	50	100
	(I)	91.22	50.99	37.85	31.49	27.85	21.98	21.78	25.37	42.48	97.70
25	(II)	1003.^{1}	1493.	523.6	260.0	155.9	40.06	22.25	13.85	12.10	18.12
	(III)	1178.^{3}	4634.^{1}	7626.	2245.	906.2	73.60	22.36	7.163	3.105	2.923
	(I)	1352.	716.4	505.4	400.8	338.6	219.4	185.6	170.7	201.4	338.5
50	(II)	5679.^{2}	7775.^{1}	2519.^{1}	1160.^{1}	6474.	1218.	526.6	224.0	110.4	97.50
	(III)	2558.^{5}	9010.^{3}	1335.^{3}	3556.^{2}	1305.^{2}	7003.	1520.	290.4	56.56	24.47
	(I)	2081.^{1}	1072.^{1}	7355.	5678.	4674.	2685.	2045.	1574.	1358.	1606.
100	(II)	3413.^{4}	4469.^{3}	1386.^{3}	6122.^{2}	3279.^{2}	5106.^{1}	1867.^{1}	5985.	1780.	881.8
	(III)	6016.^{7}	1999.^{6}	2796.^{5}	7047.^{4}	2451.^{3}	1023.^{3}	1778.^{2}	2330.^{1}	2298.	450.2

Key to Table VII -- (a) (I), (II), (III) refer to the three terms in (5-9) when n=100.

(b) Underlined entries are negative.

(c) Indices imply multiplication by the corresponding power of ten.

(e) Monte Carlo Verification.

Due to the rather involved nature of the computations, it was desired to have a Monte Carlo simulation performed to serve as a "partial check" on the results given in this chapter. The Monte Carlo results for the means and variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were obtained, with the aid of an IBM 7040 computer, at the point $\alpha=10$, $\lambda=10$. This point was chosen for study because the expansions, for both estimators, are reasonably well behaved there. A sample size of 100 was used and 4000 samples were generated. The results of the Monte Carlo study are given in Table VIII. Note the good agreement between the Monte Carlo values and the values obtained from the expansions.

Having obtained approximations to the properties of $\hat{\alpha}_1$ and $\hat{\alpha}_2$, we are now in a position to make various comparisons between them and also see how the modified moment estimators compare with other estimators. In Chapter VI we shall discuss such comparisons.

Table VIII.

Comparison of Results for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ with Monte Carlo Results

Estimator	Moment	Monte Carlo Result	Result from Expansion	% Error
$\hat{\alpha}_1$	Mean	11.2323	11.2230	.22
	Variance	16.8664	16.7133	.91
$\hat{\alpha}_2$	Mean	11.0745	11.2572	1.34
	Variance	16.3587	15.9960	2.22

(f) Further Monte Carlo Studies

In order to provide further support for some of the results reported earlier in this chapter, additional Monte Carlo simulation work was done for $\hat{\alpha}_1$. Several combinations of α and λ were chosen where the asymptotic expansions for the moments of $\hat{\alpha}_1$ appeared to converge reasonably well at n=100. Monte Carlo results for the mean and variance of $\hat{\alpha}_1$ were obtained based on 2000 samples of 100 each drawn from simulated negative binomial distributions. The results are shown in Table VIIIa along with the approximations for the moments obtained from the asymptotic expansions. Note that the expansion approximations compare favorably with the Monte Carlo results. Where slight discrepancies occur, e.g. in the case of $\mathrm{Var}(\hat{\alpha}_1)$ for $\alpha=3$, $\lambda=4$, one will note from Table IV that the n^{-4} term in the expansion does contribute to some extent and thus one would not expect a "near perfect" comparison. However, in cases where the n^{-4} term is negligible, there is extremely close agreement. Thus, at least in the regions of the parameter space in which this Monte Carlo work was done, the expansions appear to give good approximations for a sample size that renders the n^{-4} term negligible.

Monte Carlo work was also done for the purpose of making comparisons of the properties of $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}$ in

Table VIIIa

Monte Carlo Results and Expansion Approximations for the Moments of $\hat{\alpha}_1$ for n=100

		Mean		Variance	
α	λ	Monte Carlo Result	Result from Expansion	Monte Carlo Result	Result from Expansion
3	4	3.287	3.269	1.147	.883
3	5	3.248	3.215	.769	.652
4	4	4.495	4.515	2.753	2.767
4	5	4.416	4.389	2.113	1.838
5	4	5.850	5.881	7.419	7.256
5	5	5.641	5.640	4.523	4.322

some regions of the (α, λ) space in which the asymptotic expansions do not provide reliable approximations for the moments of all three estimators at sample sizes of 100. These results are given in Chapter VI.

CHAPTER VI

COMPARISON OF MODIFIED MOMENT ESTIMATORS WITH OTHER ESTIMATORS

We are now able to compare the properties of the modified moment estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ with those of the maximum likelihood estimator α^* and the simple moment estimator $\hat{\alpha}$. In doing so, it must be pointed out that we are forced to make such comparisons on the basis of the first few terms of the expansions. We are not, therefore, always led to obvious conclusions concerning the comparative quality of an estimator. However, this kind of comparison certainly seems more desirable than the misleading "first term" comparisons often made.

The classical means of comparison with maximum likelihood estimation is given by the efficiency. We shall discuss the efficiencies of the modified moment estimators and compare them with the efficiency of the simple moment estimator. However, bias is often an important consideration too. Certain bias comparisons are given later in the chapter.

(a) Efficiency.

For the two parameter NBD, we are concerned with a joint estimation problem. Thus the efficiency of the estimators $(\hat{\alpha}_k, \hat{\lambda})$ is given by

$$E = \frac{|\mathrm{Cov}(\alpha^*, \lambda^*)|}{|\mathrm{Cov}(\hat{\alpha}_k, \hat{\lambda})|} \quad , \tag{6-1}$$

where, of course, α^* and λ^* are the maximum likelihood estimators. Since $\lambda^* = \hat{\lambda} = \bar{x}$ for all methods of estimation, the asymptotic efficiency can be expressed as

$$\bar{E} = \frac{C_2^*}{C_2^{(k)}} \quad , \tag{6-2}$$

where the C's are coefficients given by (2-7) and (5-9).

The contours of constant asymptotic efficiency for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ (note that the same contours will apply to either estimator) are given in Figure 3. If we compare these contours with those shown in Figure 1, we find that the asymptotic efficiencies of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are greater than those of $\hat{\alpha}$ over the entire (α, λ) subspace. We may, however, be skeptical about this improvement when we remember that the asymptotic terms in $|\mathrm{Cov}(\hat{\alpha}_k, \bar{x})|$ sometimes contribute little to the total expansions for all the estimators in question.

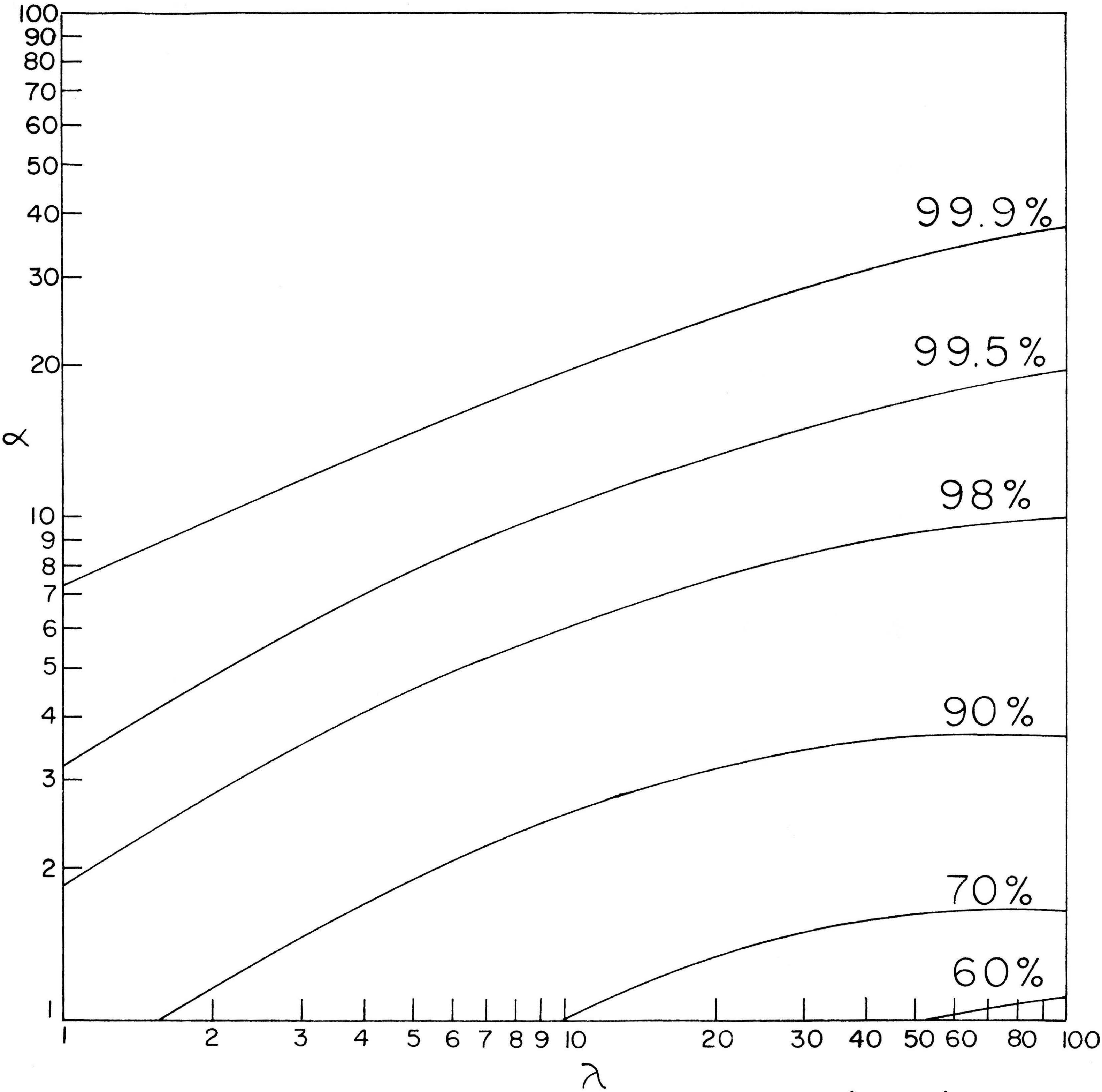

Figure 3. Contours of constant asymptotic efficiencies for $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

It seems logical, since two terms are available to us in the expansion for $|\mathrm{Cov}(\alpha^*,\lambda^*)|$, that we can obtain a more meaningful measure of the efficiency; namely we can use

$$E' = \frac{C_2^* + C_3^*/n}{C_2^{(k)} + C_3^{(k)}/n} \quad . \tag{6-3}$$

The significance of E' depends on the contribution of $C_2^{(k)}/n^2$ and $C_3^{(k)}/n^3$ to the expansion of $|\mathrm{Cov}(\hat{\alpha}_k,\bar{x})|$. Obviously, if $C_2^{(k)}/n^2 + C_3^{(k)}/n^3$ was small in comparison to the total expansion of $|\mathrm{Cov}(\hat{\alpha}_k,\bar{x})|$, E' could be as misleading as $\bar{E}$. Therefore, in addition to E', we shall be concerned with a quantity we refer to as the "relative contribution". The definition of relative contribution is given by the expression

$$\frac{C_2^{(k)} + C_3^{(k)}/n}{C_2^{(k)} + C_3^{(k)}/n + C_4^{(k)}/n^2} \quad .$$

Obviously, if the relative contribution is close to unity, the n^{-4} term in the covariance determinant expansion will be small and, although the possibility exists that terms beyond n^{-4} make a significant contribution, hopefully E' will be a "good" measure of the efficiency. Thus, we shall merely use the relative contribution as an aid in our evaluation of E' as an approximation of the true efficiency. Because of the obvious limitations of the relative contribution, it cannot be a <u>strict</u> guide for the use of E'.

Contours of constant E' for $(\hat{\alpha}_1,\hat{\lambda})$ are shown in Figure 4 at sample sizes of 100 and 1000. Relative contributions for these estimators are given in Table IX (expressed as percents). Notice that in the case of $(\hat{\alpha}_1,\hat{\lambda})$, E' does not change to a great extent with sample size even though the relative contribution varies considerably. For example at $\alpha=5$, $\lambda=2$, when n=1000, E'=99.4% while at n=100, E'=99.1%. On the other hand, the relative contribution increases from 43.37%, at n=100, to 96.85% at n=1000. We also see, from Figure 3, that an asymptotic sample will give us an efficiency of 99.6% for $(\hat{\alpha}_1,\hat{\lambda})$ at $\alpha=5$, $\lambda=2$. In fact, $(\hat{\alpha}_1,\hat{\lambda})$ appears highly efficient over a large region of the parameter space, regardless of sample size.

In the case of $(\hat{\alpha}_2,\hat{\lambda})$ there is an entirely different situation. The contours of constant E' for $(\hat{\alpha}_2,\hat{\lambda})$ at n=1000 are shown in Figure 5 and those at n=100 in Figure 6. The relative contribution values are presented in Table X. We find from Figures 5 and 6 that the efficiency picture

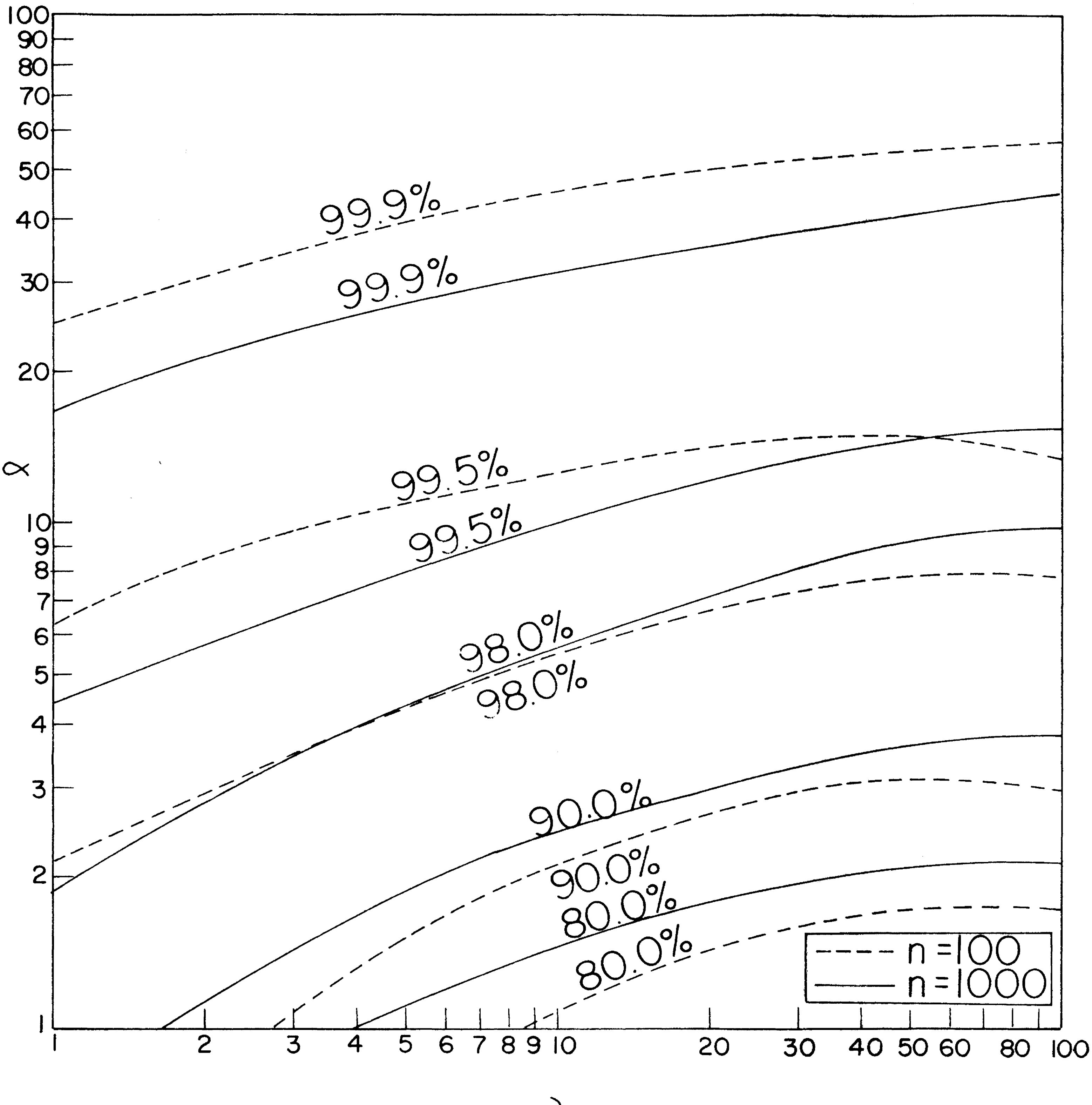

Figure 4. Contours of constant E' for $(\hat{\alpha}_1, \hat{\lambda})$ at sample sizes of 100 and 1000.

Table IX.

Relative Contribution in Percent of the First Two Terms in the Expansion of $|\mathrm{Cov}(\hat{\alpha}_1, \bar{x})|$

α	λ	n	% Relative Contribution	α	λ	n	% Relative Contribution
1	1	100	79.13	1	10	100	97.44
		1000	99.60			1000	99.97
5	1	100	15.52	10	10	100	84.22
		1000	80.61			1000	99.72
100	1	100	.06	100	10	100	4.99
		1000	.57			1000	43.22
1	2	100	94.22	1	25	100	97.04
		1000	99.93			1000	100.00
5	2	100	43.37	15	25	100	93.05
		1000	96.85			1000	99.90
25	2	100	3.22	100	25	100	24.48
		1000	30.17			1000	90.32
1	5	100	97.56	1	100	100	96.95
		1000	100.00			1000	99.96
5	5	100	83.74	25	100	100	97.65
		1000	99.70			1000	99.97
50	5	100	4.96	100	100	100	84.64
		1000	43.02			1000	99.76

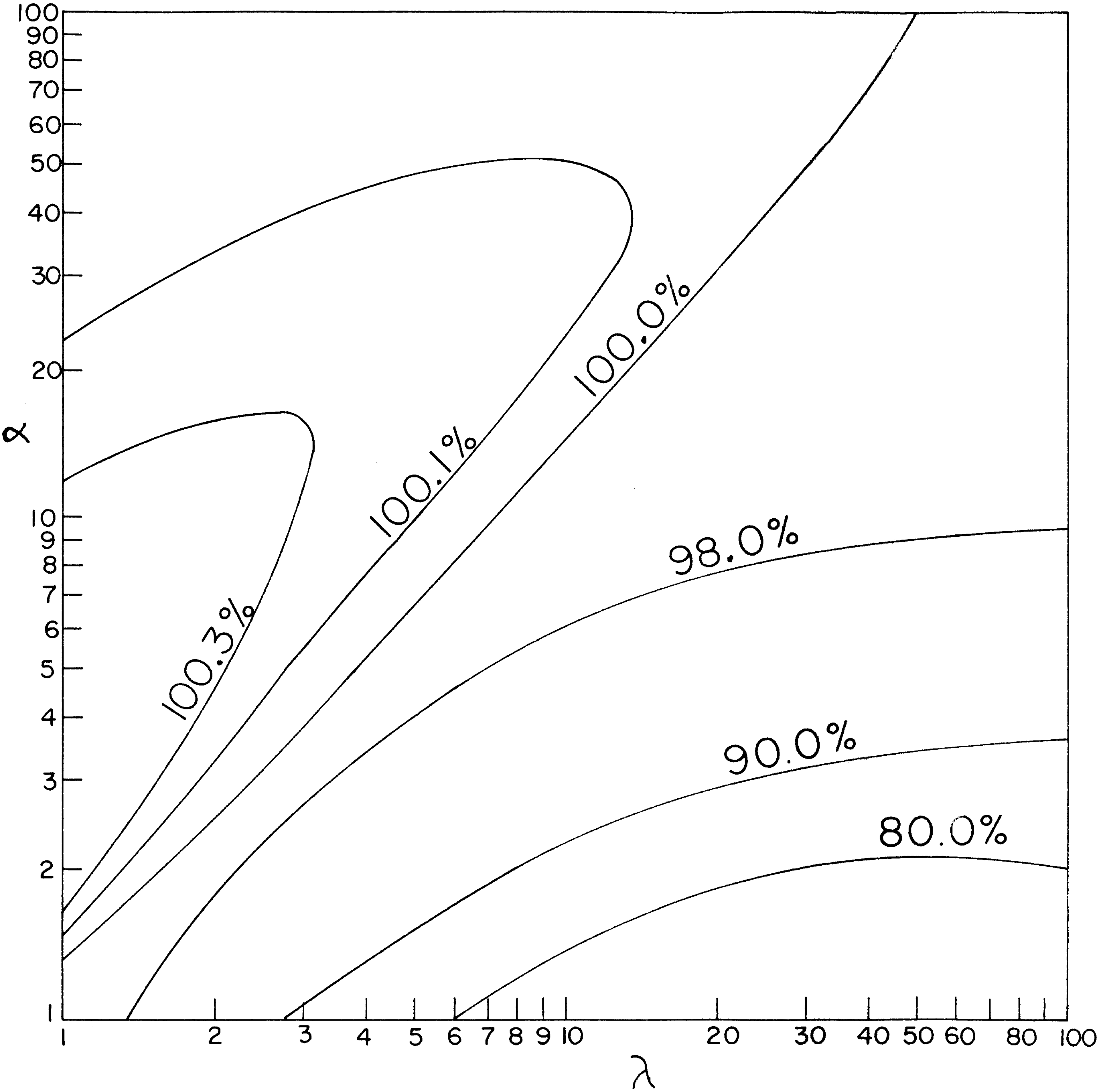

Figure 5. Contours of constant E' for $(\hat{\alpha}_2, \hat{\lambda})$ at n=1000.

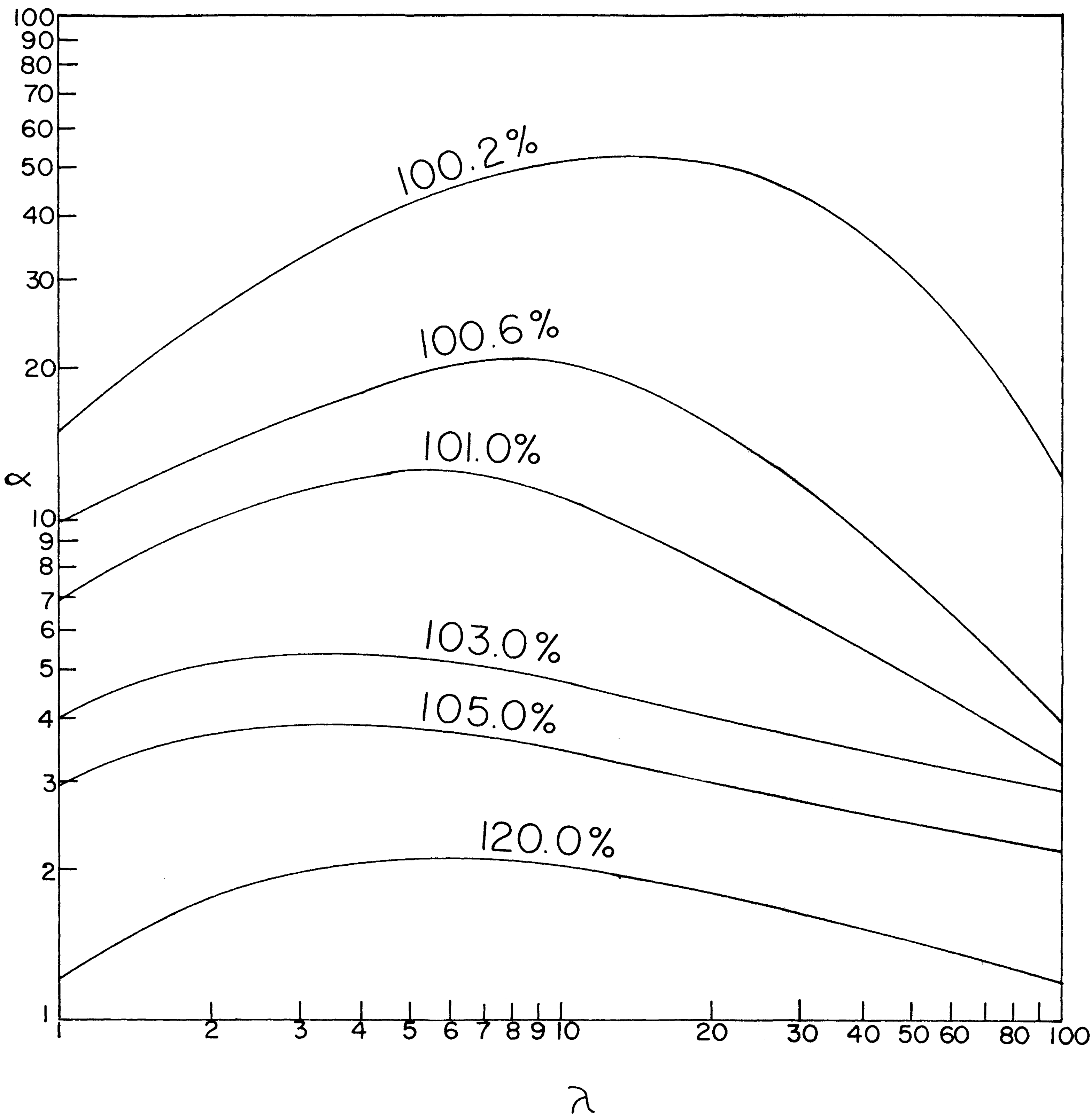

Figure 6. Contours of constant E' for $(\hat{\alpha}_2, \hat{\lambda})$ at n=1000.

Table X.

Relative Contribution in Percent of the First Two Terms in the Expansion of $|\mathrm{Cov}(\hat{\alpha}_2,\bar{x})|$

α	λ	n	% Relative Contribution	α	λ	n	% Relative Contribution
1	1	100	4.44	1	10	100	1.34
		1000	84.01			1000	86.04
5	1	100	15.45	10	10	100	83.52
		1000	80.75			1000	99.72
100	1	100	.06	100	10	100	4.99
		1000	.57			1000	43.24
1	2	100	2.77	1	25	100	1.11
		1000	85.15			1000	86.20
5	2	100	42.29	15	25	100	92.74
		1000	96.81			1000	99.90
25	2	100	3.22	100	25	100	24.49
		1000	30.18			1000	90.31
1	5	100	1.72	1	100	100	1.00
		1000	85.83			1000	86.32
5	5	100	76.37	25	100	100	97.56
		1000	99.57			1000	99.97
50	5	100	4.96	100	100	100	84.68
		1000	43.04			1000	99.76

changes considerably with sample size. In fact, at n=100 we have achieved efficiencies greater than 100% over the entire parameter subspace. Upon examination of Table X, however, we find that the relative contribution, at n=100, is low for $(\hat{\alpha}_2,\hat{\lambda})$, except perhaps where α and λ are both large. At n=1000, the relative contribution has improved somewhat but the efficiency has also decreased and the contours are quite different.

In the case of $(\hat{\alpha}_2,\lambda)$, E' seems quite sensitive to sample size. Its meaning in some portions of the (α,λ) subspace is then open to the same doubt that was cast upon asymptotic efficiency. However, while it is difficult to make general comparisons, with some caution we might base our decision of when to use $\hat{\alpha}_1$ or $\hat{\alpha}_2$ on a consideration of E' and the relative contribution.

For purposes of comparison, contours of constant E' and the relative contribution are given in Figure 7 and Table XI, respectively, for $(\hat{\alpha},\hat{\lambda})$. (The first two terms in the expansion of $|\mathrm{Cov}(\hat{\alpha},\vec{x})|$ were taken from Shenton and Myers [14]. The n^{-4} term was computed by the method described in Chapter V and is tabulated in Appendix C.) Note that the

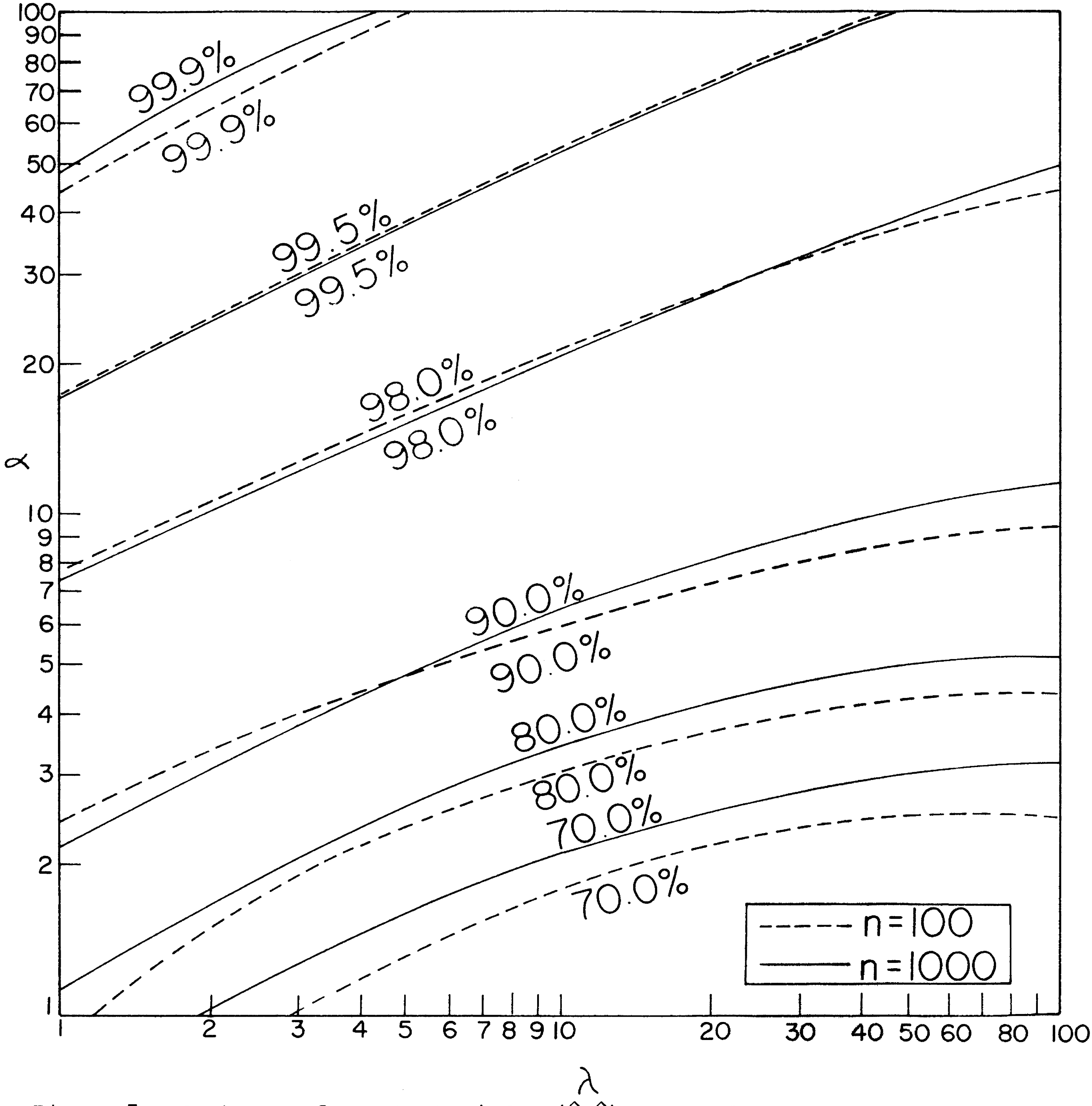

Figure 7. Contours of constant E' for $(\hat{\alpha},\hat{\lambda})$ at sample sizes of 100 and 1000.

Table XI.

Relative Contribution in Percent of the First two Terms in the Expansion of $|Cov(\hat{\alpha},\bar{x})|$

α	λ	n	% Relative Contribution	α	λ	n	% Relative Contribution
1	1	100	69.90	1	10	100	87.08
		1000	99.38			1000	99.85
5	1	100	15.52	10	10	100	84.26
		1000	80.60			1000	99.73
100	1	100	.06	100	10	100	4.99
		1000	.57			1000	43.23
1	2	100	81.95	5	25	100	97.74
		1000	99.74			1000	100.00
5	2	100	43.30	15	25	100	93.07
		1000	96.85			1000	99.90
50	2	100	.86	50	25	100	57.58
		1000	8.61			1000	98.53
1	5	100	86.13	15	100	100	98.46
		1000	99.83			1000	99.98
25	5	100	16.84	100	100	100	84.68
		1000	82.69			1000	99.76

contour pattern in Figure 7 is somewhat similar to that of $(\hat{\alpha}_1,\hat{\lambda})$. Notice too that E' for $(\hat{\alpha},\hat{\lambda})$ displays the same sort of insensitivity to sample size as did E' for $(\hat{\alpha}_1,\hat{\lambda})$.

When we compare Figure 7 with Figures 4, 5, and 6 we see again that the modified moment estimators provide an improvement in efficiency over simple moment estimation. Even more interesting is the evidence that $(\hat{\alpha}_2,\hat{\lambda})$ appears to be more efficient than (α^*,λ^*) in certain parts of the parameter space. Although we may be willing to discount some of the high values of E' for $(\hat{\alpha}_2,\hat{\lambda})$ on the basis of low relative contribution, certain of these values would seem to be good approximations to the true efficiencies. For example at $n=1000$, when $\alpha=5$, $\lambda=2$, the value of E' for $(\hat{\alpha}_2,\hat{\lambda})$ is 100.3% while the relative contribution is 96.8%.

Another interesting observation is the striking similarity of the covariance determinant expansions for all the estimators under consideration in the region where $\alpha \gg \lambda$. For example, at $\alpha=100$, $\lambda=1$,

$$|\mathrm{Cov}(\hat{\alpha}_1,\bar{x})| = \frac{2.081\times10^{8}}{n^2} + \frac{3.413\times10^{13}}{n^3} + \frac{6.017\times10^{18}}{n^4} + \dots ;$$

$$|\mathrm{Cov}(\hat{\alpha}_2,\bar{x})| = \frac{2.081\times10^{8}}{n^2} + \frac{3.413\times10^{13}}{n^3} + \frac{6.016\times10^{18}}{n^4} + \dots ;$$

$$|\mathrm{Cov}(\hat{\alpha},\hat{\lambda})| = \frac{2.081\mathrm{x}10^{8}}{n^{2}} + \frac{3.414\mathrm{x}10^{13}}{n^{3}} + \frac{6.018\mathrm{x}10^{18}}{n^{4}} + \dots ;$$

$$|\mathrm{Cov}(\alpha^{*},\lambda^{*})| = \frac{2.081\mathrm{x}10^{8}}{n^{2}} + \frac{3.413\mathrm{x}10^{13}}{n^{3}} + \dots \quad .$$

Notice the close agreement between the individual terms of the expansions. Apparently, all of these methods of estimation will be almost equally efficient when $\alpha \gg \lambda$, regardless of sample size. However, in this region all of these estimators are of doubtful value.

(b) Bias Comparisons.

Again we only have the first two terms of the expansion for $E(\alpha^*-\alpha)$ available for comparison. Thus the sums of the first two terms in the bias expansions for the modified moment estimators were compared with that for the bias of the maximum likelihood estimator. These comparisons are graphically represented in Figures 8 and 9 for $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively, for a sample size of 100. Note that in the region $\alpha < \lambda$, $\hat{\alpha}_1$ appears to have less bias than α^* and that in the region $\alpha \geq 3$, $\hat{\alpha}_2$ seems superior to α^*. One should keep in mind, however, that the bias expansions through n^{-2} are often poor approximations (see Tables II and III).

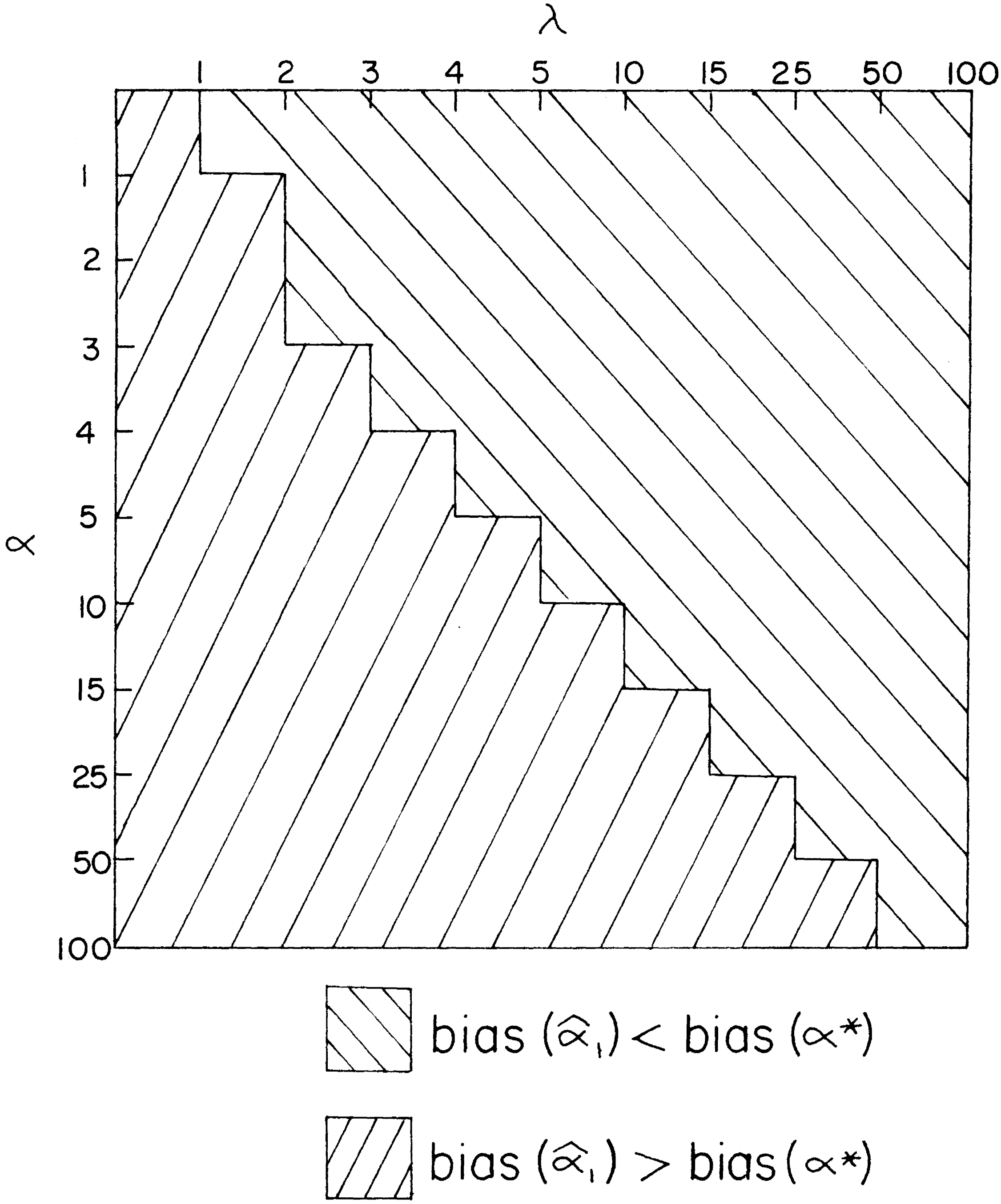

Figure 8. Comparison of the sum of the first two bias terms of $\hat{\alpha}_1$ with that of α* at n=100.

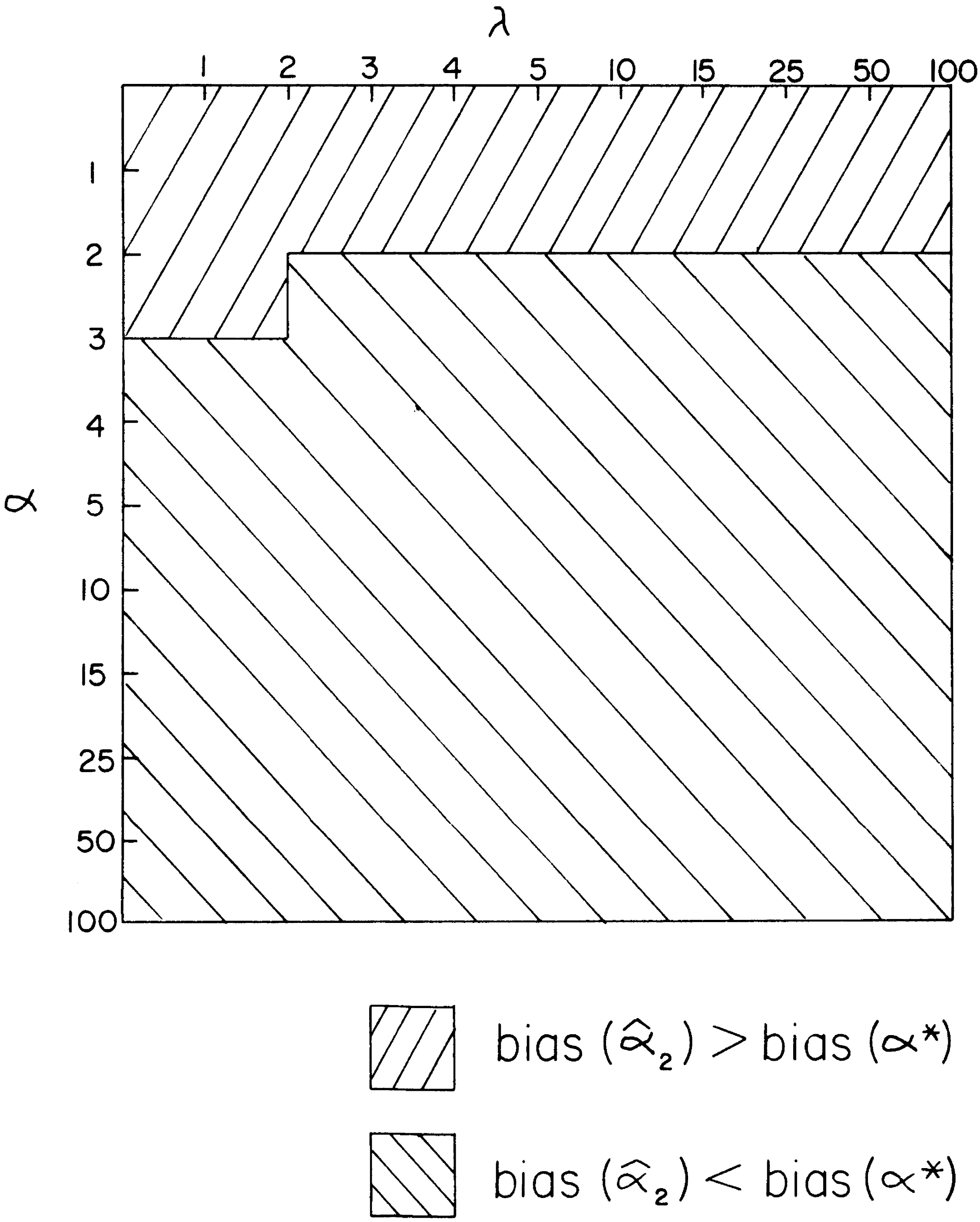

Figure 9. Comparison of the sum of the first two bias terms of $\hat{\alpha}_2$ with that of α^* at n=100.

In making bias comparisons between $\hat{\alpha}_1$ and $\hat{\alpha}_2$, and also with $\hat{\alpha}$, we are able to use the sum of four terms in the expansions. For such comparisons, we thus are able to make more reliable evaluations. At n=100, we find from Tables II and III that $\hat{\alpha}_2$ has less bias than either $\hat{\alpha}_1$ or $\hat{\alpha}$ when $\alpha \geq 10$. However, while $E(\hat{\alpha}-\alpha)$ is less than $E(\hat{\alpha}_2-\alpha)$ when $\alpha < 10$, $\underline{\hat{\alpha}_1}$ <u>always has less bias than</u> $\underline{\hat{\alpha}}$. It seems, then, that we have more evidence for choosing the modified moment estimators.

It is interesting to note, once again, an amazing agreement of the biases of all the estimators when $\alpha \gg \lambda$. At $\alpha=100$, $\lambda=1$,

$$E(\hat{\alpha}_1-\alpha) = \frac{2.070\times10^{6}}{n} + \frac{1.267\times10^{11}}{n^2} + \frac{1.293\times10^{16}}{n^3} + \frac{1.846\times10^{21}}{n^4} + \dots ;$$

$$E(\hat{\alpha}_2-\alpha) = \frac{2.070\times10^{6}}{n} + \frac{1.267\times10^{11}}{n^2} + \frac{1.292\times10^{16}}{n^3} + \frac{1.846\times10^{21}}{n^4} + \dots ;$$

$$E(\hat{\alpha}-\alpha) = \frac{2.071\times10^{6}}{n} + \frac{1.267\times10^{11}}{n^2} + \frac{1.293\times10^{16}}{n^3} + \frac{1.846\times10^{21}}{n^4} + \dots ;$$

$$E(\alpha^*-\alpha) = \frac{2.070\times10^{6}}{n} + \frac{1.267\times10^{11}}{n^2} + \dots \quad .$$

(c) Numerical Comparisons.

As another means of comparing the modified moment estimators with the maximum likelihood estimator and simple moment estimator, two sets of "live" data were fitted using all four methods of estimation. The first example involves the absence proneness data, given by Sichel [13], which was previously mentioned in Chapter II. The data and the fit obtained by each estimator are shown in Table XII. The numerical values of the estimators are also listed in this table.

The fit obtained for the absence proneness data appears to be consistent with the results given in this chapter on the efficiency of the various estimators. In fact, on the basis of the χ^2 Goodness of Fit Test, we would conclude that the fit obtained with $\hat{\alpha}$ is unsatisfactory, i.e., it results in a significant χ^2 value. From Figure 7 we see that the efficiency (E') of $\hat{\alpha}$ in this "apparent" (α,λ) region is only about 80%. On the other hand, we obtain non-significant values of χ^2 using either of the modified

Table XII.

Comparison of the Modified Moment Estimators with Other Estimators in Fitting the NBD to Absence Proneness Data

Number of Absences	Observed Frequency	Expected Frequency			
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	α^*	$\hat{\alpha}$
0	217	210.4	213.4	214.9	206.7
1	44	57.4	54.7	53.4	60.6
2	29	24.8	23.9	23.4	25.9
3	11	12.1	11.9	11.8	12.3
4	11	6.2	6.2	6.3	6.1
5	2	3.3	3.4	3.5	3.1
6	4	1.8	1.9	2.0	1.6
7 & over	0	2.0	2.6	2.7	1.7
	Value of Estimator	.460	.415	.396	.522
	χ^2	5.805	4.128	3.500	7.188

$n = 318$ d.f. = 2*

$\hat{\lambda} = .670$ $\chi^2_{.95} = 5.99$

* Two degrees of freedom result from a grouping of the last four frequencies.

moment estimators. The efficiencies (E') of these estimators are greater than 90%, that of $\hat{\alpha}_2$ being the highest, in this (α, λ) region.

The second set of data to be considered was taken from Fisher [8] and involves a sample of sheep classified according to the number of ticks found on each. This data, along with the fit obtained by each estimator, is given in Table XIII. Notice, from the values of the estimators given in the table, that we are now in an (α, λ) region where the efficiency (E') of each estimation method is greater than 90%. It is not surprising, then, that the estimators give similar values for χ^2.

Table XIII.

Comparison of Estimators in Fitting to the NBD to Sheep Tick Data

Ticks per sheep	Observed Frequency	Expected Frequency			
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	α^*	$\hat{\alpha}$
0	7	5.68	5.76	5.79	5.43
1	9	10.01	10.06	10.07	9.91
2	8	11.12	11.11	11.10	11.19
3	13	9.93	9.89	9.86	10.07
4	8	7.78	7.74	7.73	7.92
5	5	5.60	5.57	5.56	5.68
6	4	3.78	3.76	3.76	3.80
7	3	2.43	2.43	2.43	2.43
8 & over	3	3.66	3.68	3.70	3.55
	Value of Estimator	3.85	3.78	3.75	4.11
	χ^2	2.647	2.647	2.651	2.663

$n = 60$ d.f. $= 6$

$\hat{\lambda} = 3.25$ $\chi^2_{.95} = 12.59$

(d) Monte Carlo Comparison

Monte Carlo values for the means and variances of $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}$ were obtained in certain regions of the parameter space with n=100. The results were based on 2000 samples in each case and are presented in Table XIV. The parameter regions chosen are those in which the asymptotic expansions do not provide adequate approximations for the moments of all the estimators in question for a sample size of 100. Notice that in each case the modified moment estimators are superior to the simple moment estimator and that $\hat{\alpha}_2$ always has the smallest bias and variance. Note also that, in most cases, the differences between the means and variances of the three estimators are statistically significant at the 5% significance level.

One might have supposed $\hat{\alpha}_2$ to be the worst of these estimators since its moment expansions converge more slowly than those of $\hat{\alpha}_1$ and $\hat{\alpha}$ in most regions of the parameter space. However, the results from this Monte Carlo study seem to support the comparisons based on efficiency made earlier in this chapter. Thus the best moment estimator for α appears to be the modified moment estimator $\hat{\alpha}_2$.

Table XIV

Monte Carlo Comparisons of the Moments of $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}$ for n = 100

		Mean			Variance			
α	λ	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}$	Statistical Significance
1	1	1.178	1.102	1.244	1.167	.786	1.516	***
1	2	1.076	1.035	1.125	.092	.079	.118	***
1	3	1.054	1.020	1.098	.062	.055	.082	***
3	4	3.269	3.194	3.363	1.147	1.065	1.300	***
5	4	5.850	5.686	5.955	7.419	6.857	8.160	**
5	5	5.641	5.506	5.733	4.523	4.009	5.477	**
100	1	9.9×10^4	9.1×10^4	1.1×10^5	7.1×10^{12}	6.0×10^{12}	8.5×10^{12}	*
100	5	5.9×10^4	5.9×10^4	6.1×10^4	7.1×10^{12}	7.0×10^{12}	7.5×10^{12}	*

Key: *** The means are significantly different (compared pairwise) and the variances are significantly different at the 5% level.

** No significant difference between $E(\hat{\alpha}_1)$ and $E(\hat{\alpha})$. Otherwise, the means and variances are significantly different at the 5% level.

* No significant difference between the means, but the variances are significantly different at the 5% level.

CHAPTER VII

SUMMARY

The problem of estimating the two parameters, α and λ, of the negative binomial distribution has received considerable attention. The parameter λ, being the mean of the distribution, is easily and efficiently estimated by the sample mean. The task of finding a "good" estimator for α, however, has been a perplexing one. The maximum likelihood estimator α^* is difficult to obtain numerically. Furthermore it is possible that the maximum likelihood equation for α does not always have an admissible solution. Thus, many other estimators have been proposed by various workers. Among these are two modified moment estimators, $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

In this dissertation, terms through n^{-4}, in the asymptotic expansions of the bias, variance, covariance determinant, and higher moments, are given numerically for both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in the parameter subspace ($1 \leq \alpha \leq 100$, $1 \leq \lambda \leq 100$). The results show that the behavior patterns of these expansions are somewhat similar to those for the simple moment estimator $\hat{\alpha}$ and the maximum likelihood estimator α^*. For both $\hat{\alpha}_1$ and $\hat{\alpha}_2$, the n^{-4} term contributes

heavily in all the expansions when $\alpha > \lambda$. Thus, as with the other estimators, a first term approximation would not suffice for the properties of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

In terms of a comparison among $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}$, and α^*, the efficiency of the modified moment estimators suggests that their use may be more advantageous than that of either the simple moment estimator or the maximum likelihood estimator in certain instances. Here, a more general expression than the usual asymptotic efficiency was considered. This more general efficiency involves the ratio of the first two terms in the expansion of $|\mathrm{Cov}(\alpha^*,\lambda^*)|$ to those in the expansion of $|\mathrm{Cov}(\hat{\alpha}_k,\bar{x})|$, k=1,2. However, one must still be cautious since the first two terms in these expansions do not always "tell the complete story".

Comparison of the bias asymptotic expansions for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ with those of α^* and $\hat{\alpha}$ show again a possible advantage of the use of modified moment estimators in certain regions of the parameter space. Here again, the results are misleading for some α and λ because of the "inadequacy" of the asymptotic expansions. For this reason it is not always possible to decide which estimator is best. This is especially true, of course, when the n^{-4} term contributes heavily.

Some experimental data (assumed to be from negative binomial populations) was fitted to the negative binomial distribution by use of each of the estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}$, and α^*. Comparison of the results indicated that the "goodness of fit" was closely related to the efficiencies of the various estimators. In the examples given, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and α^* always provided a satisfactory fit of the data. The fit obtained by use of the modified moment estimators was always better than that with the simple moment estimator and, in one example, better than that with the maximum likelihood estimator. In one instance, the fit resulting from the use of $\hat{\alpha}$ was poor. This occurred in an "apparent" region of the parameter space in which the efficiency of $\hat{\alpha}$ was low.

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. Raymond H. Myers for suggesting this research and for his guidance during the course of the investigation.* Gratitude is also extended to Dr. Boyd Harshbarger, Head of the Statistics Department, for his assistance and encouragement.

The author acknowledges with gratitude the service of Mr. Murray Aitkin as co-reader of the manuscript. Thanks are also extended to Mr. Valiant Mah and Mr. Philip Bergstresser for their assistance in the computer work.

Finally, the author is indebted to Mrs. Lela Barnhisel for her co-operation and work in typing the manuscript.

*This investigation was carried out through fellowship funds available from the Department of Health, Education, and Welfare, to which the author is very grateful.

REFERENCES

[1] Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. Biometrika, 37, 358-382.

[2] Bowman, K. and Shenton, L. R. (1965). Biases and covariance of maximum likelihood estimators. Research and Development Report. Nuclear Division, Union Carbide Corporation.

[3] Cassidy, E. C. (1964). The estimation of the parameters of the negative binomial distribution. Thesis for M.S. degree. Virginia Polytechnic Institute.

[4] Cramer, H. (1946). Mathematical Methods of Statistics. Princeton: Princeton University Press.

[5] David, F. N. (1949). Note on the application of Fisher's k-statistics. Biometrika, 36, 383-393.

[6] Evans, D. A. (1953). Experimental evidence concerning contagious distributions in ecology. Biometrika, 40, 186-211.

[7] Feller, W. (1957). An Introduction to Probability Theory and Its Applications. Wiley, New York.

[8] Fisher, R. A. (1941). The negative binomial distribution. Annals of Eugenics, 11, 182-187.

[9] Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions of multiple happenings. Jour. Roy. Stat. Soc., 83, 255-279.

[10] Haldane, J. B. S. and Smith, S. M. (1956). The sampling distribution of a maximum-likelihood estimate. Biometrika, 43, 96-103.

[11] Hsing, P. S. (1964). Orthogonal statistics involving the third and fourth sample moments for negative binomial distribution. Thesis for M.S. degree. Virginia Polytechnic Institute.

[12] Myers, R. H. (1963). Orthogonal statistics and some sampling properties of moment estimators for the negative binomial distribution. Thesis for Ph.D. degree. Virginia Polytechnic Institute.

[13] Sichel, H. S. (1951). The estimation of the parameters of a negative binomial distribution with special reference to psychological data. Psychometrika, 16, 107-127.

[14] Shenton, L. R. and Myers, R. H. (1965). Comments on estimators for the negative binomial distribution. Proceedings of the Montreal Symposium on the Classical and Contagious Discrete Distributions.

[15] Shenton, L. R. and Myers, R. H. (1965). Orthogonal statistics. Proceedings of the Montreal Symposium on the Classical and Contagious Distributions.

[16] Shenton, L. R. and Wallington, P. A. (1962). The bias of moment estimators with an application to the negative binomial distribution. Biometrika, 49, 193-204.

[17] Shenton, L. R. and Wallington, P. A. (1961). Moment estimators and modified moment estimators with special reference to the negative binomial distribution. Unpublished.

[18] Shenton, L. R. (1964). Conjectures concerning the likelihood equation for a parameter in the negative binomial distribution. Unpublished.

[19] Shenton, L. R. (1958). Moment estimators and maximum likelihood. Biometrika, 45, 411-420.

[20] Szegö, G. (1939). Orthogonal Polynomials. American Mathematical Society: New York.

[21] Todhunter, I. (1865). A history of the mathematical theory of probability from the time of Pascal to that of Laplace. Cambridge and London, MacMillan and Co.

[22] Williamson, E. and Bretherton, M. H. (1963). Tables of the logarithmic series distribution. _Ann_. _Math_. _Stat_., 35, 284-297.

[23] Wise, M. E. (1946). The use of the negative binomial in an industrial sampling problem. _Jour_. _Roy_. _Stat_. _Soc_., _B_, _8_, 202-211.

The vita has been removed from the scanned document

A P P E N D I C E S

APPENDIX A

Third Central Moments of $\hat{\alpha}_1$ and $\hat{\alpha}_2$

The following tables give numerical values for terms in the expansions

$$\mu_3(\hat{\alpha}_k) = \frac{T_2^{(k)}}{n^2} + \frac{T_3^{(k)}}{n^3} + \frac{T_4^{(k)}}{n^4} + \ldots \quad , \quad k=1,2 \quad .$$

Table A.1.

Terms in the Expansion of $\mu_3(\hat{\alpha}_1)$

α \ λ		1		5		10		25		100	
1	$T_2^{(1)}$	.7303	3	.6596	2	.4378	2	.3370	2	.2943	2
	$T_3^{(1)}$	.1119	6	-.8710	3	-.1780	4	-.2027	4	-.2187	4
	$T_4^{(1)}$	-.1604	9	-.1660	8	-.1115	8	-.8650	7	-.4385	8
5	$T_2^{(1)}$	.4810	7	.4598	5	.1276	5	.4572	4	.2434	4
	$T_3^{(1)}$	.6993	10	.5930	7	.8000	6	.1515	6	.5087	5
	$T_4^{(1)}$	.8144	13	.3880	9	-.1217	8	-.1534	8	-.1024	8
10	$T_2^{(1)}$	.3976	9	.1919	7	.3409	6	.7152	5	.2399	5
	$T_3^{(1)}$	.1913	13	.6016	9	.4306	8	.3779	7	.6520	6
	$T_4^{(1)}$	.7556	16	.1449	12	.3749	10	.9084	8	-.5248	7
25	$T_2^{(1)}$	.1811	12	.4851	9	.5318	8	.5118	7	.6693	6
	$T_3^{(1)}$	.4796	16	.6473	12	.2282	11	.6378	9	.2640	8
	$T_4^{(1)}$	.1053	21	.7010	15	.7744	13	.5891	11	.6663	9
100	$T_2^{(1)}$	.2534	16	.4672	13	.3462	12	.1418	11	.3218	9
	$T_3^{(1)}$	.1003	22	.7881	17	.1575	16	.1274	14	.3981	11
	$T_4^{(1)}$	.3309	27	.1102	22	.5902	19	.9258	16	.3743	13

Key to Table A.1 -- The number to the immediate right of an entry implies multiplication by the corresponding power of ten.

Table A.2.

Terms in the Expansion of $\mu_3(\hat{\alpha}_2)$

α \ λ		1		5		10		25		100	
1	$T_2^{(2)}$	.7303	3	.6596	2	.4378	2	.3370	2	.2943	2
	$T_3^{(2)}$	-.2534	6	-.8893	5	-.7123	5	-.6160	5	-.5556	5
	$T_4^{(2)}$	.2815	10	.4744	9	.3517	9	.2885	9	.1732	9
5	$T_2^{(2)}$	.4810	7	.4598	5	.1276	5	.4572	4	.2434	4
	$T_3^{(2)}$	.6855	10	.5410	7	.6338	6	.6843	5	-.7773	4
	$T_4^{(2)}$	.7975	13	.8462	9	.1886	9	.8990	8	.5605	8
10	$T_2^{(2)}$	.3976	9	.1909	7	.3409	6	.7152	5	.2399	5
	$T_3^{(2)}$	.1902	13	.5900	9	.4185	8	.3602	7	.5908	6
	$T_4^{(2)}$	.7502	16	.1414	12	.3786	10	.1668	9	.3995	8
25	$T_2^{(2)}$	.1811	12	.4851	9	.5318	8	.5118	7	.6693	6
	$T_3^{(2)}$	.4790	16	.6445	12	.2267	11	.6317	9	.2613	8
	$T_4^{(2)}$	.1052	21	.6971	15	.7660	13	.5728	11	.6201	9
100	$T_2^{(2)}$	.2534	16	.4672	13	.3462	12	.1418	11	.3218	9
	$T_3^{(2)}$	.1003	22	.7878	17	.1574	16	.1272	14	.3972	11
	$T_4^{(2)}$	.3309	27	.1101	22	.5898	19	.9242	16	.3711	13

Key to Table A.2 -- The number to the immediate right of an entry implies multiplication by the corresponding power of ten.

APPENDIX B

Fourth Central Moments of $\hat{\alpha}_1$ and $\hat{\alpha}_2$

The following tables give numerical values for terms in the expansions

$$\mu_4(\hat{\alpha}_k) = \frac{F_2^{(k)}}{n^2} + \frac{F_3^{(k)}}{n^3} + \frac{F_4^{(k)}}{n^4} + \ldots \quad , \quad k=1,2 \quad .$$

Table B.1.

Terms in the Expansion of $\mu_4(\hat{\alpha}_1)$

α \ λ		1		5		10		25		100	
1	$F_2^{(1)}$	.5141	3	.5300	2	.3565	2	.2756	2	.2409	2
	$F_3^{(1)}$	.1542	6	.5634	4	.3604	4	.2782	4	.2355	4
	$F_4^{(1)}$	-.7606	8	-.7926	7	-.5267	7	-.4118	7	-.2906	8
5	$F_2^{(1)}$	.1315	8	.1440	6	.4305	5	.1668	5	.9406	4
	$F_3^{(1)}$	.3828	11	.3553	8	.5119	7	.1072	7	.4057	6
	$F_4^{(1)}$	.7028	14	.4468	10	.1645	9	-.2657	8	-.2543	8
10	$F_2^{(1)}$	.2084	10	.1094	8	.2085	7	.4788	6	.1754	6
	$F_3^{(1)}$	.2040	14	.6788	10	.5103	9	.4852	8	.9248	7
	$F_4^{(1)}$	.1270	18	.2594	13	.7243	11	.2375	10	.1337	9
25	$F_2^{(1)}$	.2308	13	.6464	10	.7398	9	.7726	8	.1145	8
	$F_3^{(1)}$	.1256	18	.1742	14	.6321	12	.1880	11	.8774	9
	$F_4^{(1)}$	.4350	22	.2975	17	.3389	15	.2769	13	.3663	11
100	$F_2^{(1)}$	.1273	18	.2378	15	.1788	14	.7611	12	.1993	11
	$F_3^{(1)}$	.1041	24	.8242	19	.1662	18	.1378	16	.4686	13
	$F_4^{(1)}$	.5416	29	.1817	24	.9822	21	.1580	19	.6994	15

Key to Table B.1 -- The number to the immediate right of an entry implies multiplication by the corresponding power of ten.

Table B.2.

Terms in the Expansion of $\mu_4(\hat{\alpha}_2)$

α \ λ		1		5		10		25		100	
1	$F_2^{(2)}$	.5141	3	.5300	2	.3565	2	.2756	2	.2409	2
	$F_3^{(2)}$	.7805	5	-.2612	4	-.2047	4	-.1644	4	-.2052	4
	$F_4^{(2)}$	-.1516	9	-.4170	8	-.3201	8	-.2609	8	.1096	8
5	$F_2^{(2)}$	.1315	8	.1440	6	.4305	5	.1668	5	.9406	4
	$F_3^{(2)}$	.3776	11	.3351	8	.4603	7	.8832	6	.3002	6
	$F_4^{(2)}$	.6890	14	.4706	10	.2943	9	.2435	8	.1263	8
10	$F_2^{(2)}$	.2084	10	.1094	8	.2085	7	.4788	6	.1754	6
	$F_3^{(2)}$	.2032	14	.6694	10	.4992	9	.4677	8	.8718	7
	$F_4^{(2)}$	.1262	18	.2537	13	.7047	11	.2453	10	.2106	9
25	$F_2^{(2)}$	.2308	13	.6464	10	.7398	9	.7726	8	.1145	8
	$F_3^{(2)}$	.1255	18	.1737	14	.6292	12	.1867	11	.8690	9
	$F_4^{(2)}$	.4346	22	.2962	17	.3361	15	.2719	13	.3482	11
100	$F_2^{(2)}$	.1273	18	.2378	15	.1788	14	.7611	12	.1993	11
	$F_3^{(2)}$	.1041	24	.8240	19	.1662	18	.1377	16	.4678	13
	$F_4^{(2)}$	.5416	29	.1816	24	.9815	21	.1577	19	.6959	15

Key to Table B.2 -- The number to the immediate right of an entry implies multiplication by the corresponding power of ten.

APPENDIX C

Coefficients of the n^{-4} Term in the Expansion of $|Cov(\hat{\alpha},\hat{\lambda})|$

α \ λ	1		2		3		4		5		10		15	
1	.2123	6	.1379	6	.1656	6	.2148	6	.2775	6	.7555	6	.1492	7
2	.4041	7	.7493	6	.4330	6	.3628	6	.3545	6	.5448	6	.8929	6
3	.3985	8	.4744	7	.1932	7	.1226	7	.9623	6	.8199	6	.1058	7
4	.2358	9	.2197	8	.7409	7	.4029	7	.2778	7	.1562	7	.1632	7
5	.1006	10	.7951	8	.2360	8	.1157	8	.7310	7	.3068	7	.2717	7
10	.1248	12	.6642	10	.1430	10	.5356	9	.2690	9	.5087	8	.2775	8
15	.2465	13	.1119	12	.2097	11	.6952	10	.3131	10	.3943	9	.1643	9
25	.1182	15	.4666	13	.7699	12	.2272	12	.9193	11	.7522	10	.2294	10
50	.2560	17	.9027	15	.1338	15	.3568	14	.1311	14	.7055	12	.1534	12
100	.6018	19	.2000	18	.2798	17	.7054	16	.2454	16	.1025	15	.1783	14

Key -- The number to the immediate right of an entry implies multiplication by the corresponding power of ten.

ABSTRACT

This dissertation deals with the properties of two modified moment estimators for parameters of the negative binomial distribution (NBD).

Several parametric forms have been suggested for the NBD. The estimation problems vary according to the form which is used. In particular, the form proposed by Anscombe [Biometrika, 37 (1950), pp. 358-382], with parameters λ and α, has received wide attention and was selected for study in this investigation. In Anscombe's parametric form, the mean of the NBD is λ and the variance is $\lambda + \lambda^2/\alpha$.

While the parameter λ is universally estimated by the sample mean, many different methods of estimation for α have been attempted. Among these, the maximum likelihood estimator α^* and the simple moment estimator $\hat{\alpha}$ are most often used. However, α^* is quite difficult to obtain numerically and often this computation requires the use of an electronic computer. In addition, $\hat{\alpha}$, while not difficult to compute, is often inefficient. For these reasons, it was felt that a study of the two modified moment estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$, suggested by Shenton and Wallington [Moment

Estimators and Modified Moment Estimators with Special Reference to the Negative Binomial Distribution (unpublished)], was needed.

In the text, the method of obtaining modified moment estimators in general is given in detail. The application of this method to the NBD is discussed and, in particular, the derivations of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are presented. Since orthogonal statistics play an important part in this work, their definition and applications are reviewed.

In order to evaluate the small sample properties of $\hat{\alpha}_1$ and $\hat{\alpha}_2$, asymptotic expansions, in powers of $1/n$, of their biases, variances, covariance determinants, and higher moments were determined numerically in the parameter space $(1 \leq \alpha \leq 100,\ 1 \leq \lambda \leq 100)$, through terms to n^{-4}. The computational method for this work is described in detail. Tables and charts which display the nature of the expansions are given in the text.

The results show that the behavior patterns of the moment expansions for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are somewhat similar to those for $\hat{\alpha}$ and α^*. For both $\hat{\alpha}_1$ and $\hat{\alpha}_2$, the n^{-4} term contributes heavily in all the expansions when $\alpha > \lambda$.

Thus, as with the other estimators, a first term approximation would not suffice for the properties of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

Further, the results give evidence that $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are highly efficient for most α and λ, and, in some regions of the parameter space, have less bias than α^* and $\hat{\alpha}$. Some experimental data was fitted to the NBD using the estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}$, and α^*. In all of the examples given, the modified moment estimators provided a better fit of the data than did the simple moment estimator and, in one instance, a better fit than was obtained by the maximum likelihood estimator.