ORTHOGONAL STATISTICS AND SOME SAMPLING PROPERTIES OF MOMENT ESTIMATORS FOR THE NEGATIVE BINOMIAL DISTRIBUTION

by

Raymond H. Myers, B.S., M.S.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

July 1963

Blacksburg, Virginia

TABLE OF CONTENTS

Chapter		Page
I	INTRODUCTION	6
	(a) Background	6
	(b) Purpose and Scope	7
II	PROPERTIES OF ORTHOGONAL POLYNOMIALS AND ORTHOGONAL STATISTICS	11
	(a) Method of Obtaining Orthogonal Polynomials	12
	1. General Case	12
	2. Certain Specific Cases	15
	(b) Determination and Use of Orthogonal Statistics	18
	(c) Simplified Method for Determining Expected Q-Products	22
	(d) Expected Value of q-products	24
	(e) Examples	30
III	MODELS, FORMS, AND ESTIMATORS FOR NEGATIVE BINOMIAL DISTRIBUTION	36
	(a) Physical Models	36
	(b) Forms	39
	(c) Estimators	41
	(d) Practical Work with the Negative Binomial .	44

TABLE OF CONTENTS (cont.)

Chapter		Page
IV	GENERAL USE OF ORTHOGONAL STATISTICS IN THE NEGATIVE BINOMIAL DISTRIBUTION	48
	(a) Relationships between Q's and Sample Moments	48
	(b) Some Properties of q's for Negative Binomial	50
	(c) Expected Q-Products for Negative Binomial .	51
V	MOMENT ESTIMATORS FOR THE NEGATIVE BINOMIAL . .	55
	(a) Moment Estimators for Anscombe's Parameters	56
	1. The Bias of $\hat{\alpha}$	56
	2. The Variance of $\hat{\alpha}$	64
	3. Covariance Determinant of $(\hat{\alpha},\hat{\lambda})$	72
	4. Discussion of the "Distribution" of $\hat{\alpha}$.	74
	(b) Evans' Estimator	82
	1. First Two Moments of $\hat{a}$	83
	2. Covariance Determinant	85
	3. Remarks on the Distribution of $\hat{a}$ . . .	88
	4. Examples — $\mu_3(\hat{a})$ and $\mu_4(\hat{a})$	89
	4.1 $\mu_3(\hat{a})$	90
	4.2 $\mu_4(\hat{a})$	95

TABLE OF CONTENTS (cont.)

Chapter Page

(c) Fisher's Form for the Negative Binomial Distribution 98

1. Covariance Determinant 99

(d) Verification of Expansions 102

VI COMPARISON WITH RESULTS OF MAXIMUM LIKELIHOOD ESTIMATORS 109

VII SUMMARY 117

(a) General Remarks on Orthogonal Statistics . 117

(b) General Remarks on Sampling Properties of Moment Estimators for the Negative Binomial Distribution 118

(c) Concluding Remarks on the Moment Estimators for the Negative Binomial Distribution 119

ACKNOWLEDGEMENTS 123

REFERENCES 124

VITA 127

TABLE OF CONTENTS (cont.)

	Page
APPENDICES	129
Appendix A	130
Appendix B	138
Appendix C (a)	142
Appendix C (b)	144
Appendix D (a)	146
Appendix D (b)	147

CHAPTER I

INTRODUCTION

(a) Background

We shall first consider certain interesting problems which arise in regard to estimation procedures for the parameters of the negative binomial distribution. Various forms have been proposed for the negative binomial distribution, and the estimation problems which arise depend on the form assumed. Several estimators have been suggested by various workers, among these estimators being those found by fitting moments. However, only the usual large sample properties, e.g., asymptotic bias, variance, efficiency, etc., have heretofore been discussed for these estimators.

The second general problem which we propose to investigate is that of developing a method for obtaining population characteristics, namely; mean, variance, and the third and fourth moment of statistics which are functions of the sample moments. The method proposed for development is the method of orthogonal statistics. This procedure is based on the set of polynomials, $q_0(x)$, $q_1(x)$, $q_2(x)$, ..., which exists for many statistical distributions and possesses the usual properties of orthogonal polynomials. These polynomials are established and well known for the classical distributions such as Normal, Poisson, Binomial, Chi-square, etc.

(b) Purpose and Scope

This dissertation will actually serve a dual purpose; (a) to develop and simplify the use of orthogonal statistics, and (b) to apply the latter to a study of characteristics of certain moment estimators of parameters of the negative binomial distribution. We shall consider first a few details involving the scope of development for the orthogonal statistics.

The theory behind the development and use of orthogonal statistics is discussed at length. Also, certain "users tables" were prepared, i.e., dictionaries which simplify the use of this technique. These tables will enable one to obtain, in ascending powers of $1/n$, expansions of sampling moments of estimators which are functions of the first few sample moments. The tables facilitate expansions, if desired, through terms in n^{-5}. Further tables for the particular case of the negative binomial distribution are given to ease the heavy algebra that might be involved in this technique. It might be said here that the technique of orthogonal statistics lends itself particularly well to discrete distributions.

Particular cases are given where expansions of moments of certain moment statistics are derived using the method of orthogonal statistics. Examples used to illustrate this technique are: (a) certain moment estimators from a negative binomial sample; (b) the square-root of the coefficient of skewness, $m_3/m_2^{3/2}$, from a normal sample; (c) the coefficient of variation from a normal sample, and others.

The orthogonal statistics are put to use in an investigation of the properties of moment estimators for the particular forms of the negative binomial distribution due to Anscombe [1], Evans [10], and Fisher [13]. The purpose was to determine the accuracy of the usual asymptotic formulae for various momental constants by assessing the contribution of higher order terms. It is worth noting that due to the wide use of the negative binomial distribution, and because of the complex nature of other proposed estimators, considerable interest has been generated toward this problem. The moment estimators for these cases are certainly attractive from the standpoint of ease in handling, i.e., in comparison to the maximum likelihood and other estimators. Because of this and as a result of the fact that moment estimators appear to possess <u>seemingly</u> high efficiency for certain combinations of the parameters, some practical workers have advocated their use (in particular, their use is recommended in the region of the parameter space for which the large sample efficiency is greater than 90%). However in all cases, only the usual asymptotic properties have been discussed. Preliminary investigations pointed toward the indication that for one particular estimator in the case of Anscombe's form, the sampling distribution might be quite different from that depending on the usual asymptotic moments. Thus it was felt that more thorough evaluations of these estimators were needed beyond the asymptotic properties.

Expressions for the expansions of the first four moments of these estimators are derived through terms in n^{-4}. Emphasis is put on the bias and variance of these estimators for a wide range in the parameter space and for various sample sizes. The object here is to shed some light on the small sample properties of the estimators and, at the same time, show how one might be mislead by considering, without further investigation, only the usual asymptotic properties.

Objective comparisons are made which show the difference between a "good" approximation and an asymptotic one with special reference to the bias, variance, and the determinant of the covariance matrix. Certain pictorial representations are given which illustrate this comparison. It should be noted that in many cases a good deal of formidable algebra preceeded the attainment and tabulation of the results of these properties. Very often an IBM 1620 was used to facilitate the lengthy computations.

Pearson curves are fitted using approximations for β_1 and β_2, the coefficients of skewness and kurtosis respectively. The purpose here is to show departures from normality of the estimators for various sample sizes. Certain checks are used in justifying the expansions given in the text. Among these is a brief Monte Carlo study which simulates the negative binomial and generates chance values of one of the interesting estimators. A small population of this estimator results and the moments are calculated and compared to those found from the expansions.

It is believed that this thesis might also lead to further thought and research, particularly in the field of estimation and asymptotic expansions of statistical properties of estimators.

CHAPTER II

PROPERTIES OF ORTHOGONAL POLYNOMIALS AND ORTHOGONAL STATISTICS

We shall introduce the concept of orthogonal polynomials with respect to a statistical distribution function (see Cramer [8]). 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

$$\int_{-\infty}^{\infty} q_r(x)q_s(x)dF(x) = \varphi_r \qquad \text{for} \quad r = s$$
$$= 0 \qquad \text{for} \quad r \neq s \quad ,$$

where $\varphi_r > 0$, $\varphi_0 = 1$, and the coefficient of x^r in $q_r(x)$ is unity. The polynomials will be infinite in number if the set of points of increase of F is infinite, and finite in number if F has only a finite number of points of increase (Cramer [8]).

For a discrete distribution for which

$$\Pr\{X = x\} = p_x \qquad (x = 0,1,\ldots.,.)$$

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

$$\Sigma\, q_r(x)q_s(x)p_x = \varphi_r \qquad \text{for} \quad r = s$$
$$= 0 \qquad \text{for } r \neq s$$

where $\varphi_r > 0$.

(a) Method of Obtaining Orthogonal Polynomials

1. General Case

In the same notation as before, let $f(x)$ be a frequency or probability function with all moments existing. Let $x-\mu = X$. The q's are obtained in the following manner;

$$q_1(x) = -\begin{vmatrix} 1 & X \\ 1 & \mu_1 \end{vmatrix} \quad , \tag{2-1}$$

$$q_2(x) = \begin{vmatrix} 1 & X & X^2 \\ 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \end{vmatrix} \div \mu_2 \quad , \tag{2-2}$$

where μ_j = j'th central moment for the distribution of x with $\mu_1 = 0$.

Thus

$$q_2(x) = X^2 - \frac{\mu_3}{\mu_2} X - \mu_2 \quad , \tag{2-3}$$

$$q_3(x) = -\begin{vmatrix} 1 & X & X^2 & X^3 \\ 1 & \mu_1 & \mu_2 & \mu_3 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ \mu_2 & \mu_3 & \mu_4 & \mu_5 \end{vmatrix} \div \begin{vmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix} \quad , \tag{2-4}$$

and

$$q_4(x) = \begin{vmatrix} 1 & x & x^2 & x^3 & x^4 \\ 1 & \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 \\ \mu_2 & \mu_3 & \mu_4 & \mu_5 & \mu_6 \\ \mu_3 & \mu_4 & \mu_5 & \mu_6 & \mu_7 \end{vmatrix} \div \begin{vmatrix} 1 & \mu_1 & \mu_2 & \mu_3 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ \mu_2 & \mu_3 & \mu_4 & \mu_5 \\ \mu_3 & \mu_4 & \mu_5 & \mu_6 \end{vmatrix} .$$

In general, the r'th orthogonal polynomial $q_r(x)$ can be written in determinantal form as follows:

$$q_r(x) = \delta \begin{vmatrix} 1 & x & x^2 & \cdots & x^r \\ 1 & \mu_1 & \mu_2 & \cdots & \mu_r \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{r+1} \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_{r+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \mu_{r-1} & \mu_r & \mu_{r+1} & \cdots & \mu_{2r-1} \end{vmatrix} \div \begin{vmatrix} 1 & \mu_1 & \mu_2 & \cdots & \mu_{r-1} \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_r \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{r-1} & \mu_r & \mu_{r+1} & & \mu_{2r-2} \end{vmatrix} ,$$

where $\delta = +1$ if r is even and -1 if r is odd.

Certain recurrence relationships exist for any set of orthogonal polynomials. Actually there is a fundamental recurrence of order 2 from which others of higher order can be derived. These can be extremely helpful and can be used not only in obtaining the q's themselves, but also in obtaining expected values of certain products. This is of particular use when the polynomials are eventually used to generate sampling moments. This, of course, will become more apparent later on in the text.

Szegö [28] gives the fundamental recurrence relationship in the following form:

$$q_r(x) = (x-a_r)q_{r-1}(x) - b_r q_{r-2}(x) \quad , \tag{2-5}$$

where $r \geq 2$. In this form, $q_0 = 1$ and $q_1 = x-a_0$. To obtain b_r we multiply both sides by q_{r-2} and take expectations. We can then write

$$b_r \varphi_{r-2} = E\{x \cdot q_{r-2}(x) \cdot q_{r-1}(x)\} \quad . \tag{2-6}$$

Using equation (2-5) once again, we can write

$$q_{r-1}(x) = (x-a_{r-1})q_{r-2}(x) - b_{r-1}q_{r-3}(x) \quad .$$

Multiplying both sides by $q_{r-1}(x)$ and taking expectations, we have:

$$\varphi_{r-1} = E\{x \cdot q_{r-2}(x) \cdot q_{r-1}(x)\} \quad . \tag{2-7}$$

Thus (2-6) can be written

$$b_r \varphi_{r-2} = \varphi_{r-1} \quad ,$$

and finally,

$$b_r = \frac{\varphi_{r-1}}{\varphi_{r-2}} \quad , \qquad r = 2,3,\ldots \quad . \tag{2-8}$$

To obtain a_r we first write

$$q_r(x) = x^r + c_{r-1}^{(r)} x^{r-1} + c_{r-2}^{(r)} x^{r-2} + \ldots + c_1^{(r)} x + c_0^{(r)} ,$$

where $c_\ell^{(r)}$ = coefficient of x^ℓ in $q_r(x)$. Then equating coefficients of x^{r-1} on both sides of (2-5) yields:

$$c_{r-1}^{(r)} = c_{r-2}^{(r-1)} - a_r ,$$

so that finally:

$$a_r = c_{r-2}^{(r-1)} - c_{r-1}^{(r)} , \quad (r = 2,3,\ldots.) . \qquad (2\text{-}9)$$

2. <u>Certain Specific Cases</u>

Consider the case of the normal distribution with the usual probability density

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} .$$

The first three orthogonal polynomials are as follows:

$$\begin{aligned} q_0 &= 1 , \\ q_1 &= x - \mu , \\ q_2 &= (x-\mu)^2 - \sigma^2 , \\ q_3 &= (x-\mu)^3 - 3(x-\mu)\sigma^2 . \end{aligned} \qquad (2\text{-}10)$$

Certain special properties are characteristic of these q's for the normal distribution. If we consider the normal distribution with $\mu = 0$ and $\sigma = 1$, the following property holds:

$$H_r(x) = e^{-\frac{1}{2}D_x^2}(x^r) \quad , \tag{2-11}$$

where $e^{-\frac{1}{2}D_x^2}$ is expanded in ascending powers of D_x^2. This property can be used to obtain the following polynomials:

$$\begin{aligned} H_2(x) &= x^2 - 1 \quad , \\ H_3(x) &= x^3 - 3x \quad , \\ H_4(x) &= x^4 - 6x^2 + 3 \quad , \\ H_5(x) &= x^5 - 10x^3 + 15x \quad , \\ &\text{etc..} \end{aligned} \tag{2-12}$$

These are called the Hermite Polynomials (see Kendall [20] and Szegö [28] for example). Another interesting formula for these Hermite Polynomials is

$$e^{-\frac{1}{2}x^2} H_r(x) = (-D_x)^r[e^{-\frac{1}{2}x^2}] \quad . \tag{2-13}$$

From observation we can see that the polynomials of (2-12) can be generated using this expression. From the property in (2-13) it can be shown that these "normalized" polynomials follow the recurrence relationship:

$$H_r(x) = xH_{r-1}(x) - (r-1)H_{r-2}(x) \quad .$$

There are other instances where recurrence relationships aid in the determination of these orthogonal polynomials. For example, in the case of a Poisson variate with discrete probability function of the form

$$f(x) = \frac{e^{-m} m^{x}}{x!} \quad , \quad x = 0,1,2,\ldots \quad ,$$

the following relationship holds for the generation of these polynomials: (We shall use Kendall's [20] notation here, i.e., use k's instead of q's.)

$$k_r(x) = e^{-m\Delta} x^{(r)} \tag{2-14}$$

where Δ is an advancing difference operator. We denote by $x^{(r)}$, the factorial term $x(x-1)(x-2)\ldots.(x-r+1)$. These polynomials are called Charlier's Polynomials [28]. The first few are:

$$k_0 = 1 \quad ,$$

$$k_1 = x - m \quad ,$$

$$k_2 = x^2 - x - 2xm + m^2 \quad , \text{ and}$$

$$k_3 = x(x-1)(x-2) - 3mx(x-1) + 3xm^2 - m^3 \quad .$$

Kendall [20] cites other classical distributions such as χ^2, binomial, etc. in which the polynomials are worked out.

(b) Determination and Use of Orthogonal Statistics

We shall now proceed to outline and illustrate details involving the actual use of these orthogonal polynomials in obtaining moments of the distributions of certain moment statistics. Consider a sample $(x_1, x_2, \ldots, x_n)$ from the distribution in question. Suppose we define

$$Q_r = \sum_{j=1}^{n} q_r(x_j)/n \qquad (2\text{-}15)$$

so that from the orthogonality property of the q's, we can write

$$\left.\begin{aligned} E(Q_r) &= 0 \ , && (r = 1,2,\ldots) \\ E(Q_r Q_s) &= \delta_{rs}\varphi_r/n \ , && (r,s = 1,2,\ldots) \end{aligned}\right\} , \qquad (2\text{-}16)$$

where δ_{rs} is the Kronecker delta symbol, i.e.,

$$\begin{aligned} \delta_{rs} &= 0 && \text{if } r \neq s \\ &= 1 && \text{if } r = s \ . \end{aligned}$$

We shall, from the outset of this discussion, refer to the Q's as orthogonal statistics. However, it must be said that they really are "pseudo statistics", i.e., they are functions of the random variable x but also involve the parameters of the distribution. It is easily seen here that these Q's are linear functions of the sample moments and are hence random variables. Consider for example the case of the polynomials associated with the normal distribution:

$$Q_0 = 1 \quad ,$$
$$Q_1 = m_1' - \mu \quad ,$$
$$Q_2 = m_2' - 2\mu m_1' + \mu^2 - \sigma^2 \quad ,$$
etc. ,

where $m_k' = \dfrac{\sum_{i=1}^{n} x_i^k}{n}$.

From (2-1), (2-2), and (2-3) and from the definition of Q_j, one can see that in general:

$$Q_0 = 1 \quad ,$$
$$Q_1 = m_1' - \mu \quad , \tag{2-17}$$
$$Q_2 = m_2' - m_1'(2\mu + \mu_3/\mu_2) - \mu_2' + 2\mu_1'^2 + \mu_1'\mu_3/\mu_2 \quad .$$

Here we denote μ_k' as the k'th moment about zero. Conversely, then we can write

$$m_1' = \mu_1' + Q_1 \quad ,$$
$$m_2' = \mu_2' + (2\mu + \mu_3/\mu_2)Q_1 + Q_2 \quad . \tag{2-18}$$

It would be proper at this time to introduce notation that will be used throughout the remainder of this text. We shall use the convention

$$E(Q_r^{\alpha}\, Q_s^{\beta}\, Q_t^{\gamma} \ldots.) = (r^{\alpha} s^{\beta} t^{\gamma} \ldots.) \quad ,$$
$$E(q_r^{\alpha}\, q_s^{\beta}\, q_t^{\gamma} \ldots.) = [r^{\alpha} s^{\beta} t^{\gamma} \ldots.] \quad . \tag{2-19}$$

Any moment statistic $t(m_1', m_2' \ldots . m_k')$ can be expanded in the following form (at least for a class of t functions);

$$\begin{aligned} t = {} & a_{0,0\ldots.0} + a_{1,0\ldots.0}Q_1 + a_{0,1,0\ldots.0}Q_2 \\ & + a_{0,0,1,0\ldots.0}Q_3 + \ldots. + a_{1,1,0\ldots.0}Q_1Q_2 \\ & + a_{1,0.1,0\ldots.0}Q_1Q_3 + \ldots. + a_{2,0,0\ldots.0}Q_1^2 + \ldots. \\ & + a_{\alpha,\beta,\ldots.\xi}Q_1^{\alpha}Q_2^{\beta} \ldots. Q_k^{\xi} + \ldots. \quad . \end{aligned} \tag{2-20}$$

This expansion may in some cases be non-terminable. Likewise, any power of t can be expanded in a similar fashion. Thus at least formally, the k'th moment of t is found by taking the expected value term by term of the expansion (2-20). This suggests a need for a system of obtaining the expected values of Q products, or in our notation values of $(r^{\alpha} s^{\beta} t^{\gamma} \ldots.)$. These can be obtained in ascending powers of $\frac{1}{n}$ with the coefficients being the square bracket terms defined by (2-19), i.e., the expected q-products.

To arrive at the expansion described in the previous paragraphs, we consider the joint moment generating function of the Q's,

$$E\{\exp[\alpha_r Q_r+\alpha_s Q_s+\alpha_t Q_t+\ldots.]\}$$

$$= \{E\ \exp[\frac{\alpha_r q_r+\alpha_s q_s+\alpha_t q_t+\ldots.}{n}]\}^n \quad , \qquad (2\text{-}21)$$

in which r, s, t, etc. are distinct positive integers. This expression follows from consideration of the independence of the observations and from the basic definition of the Q's given by (2-15). We can then write

$$E\{\exp[\alpha_r Q_r+\alpha_s Q_s+\alpha_t Q_t+\ldots.]\}$$

$$= \{E[1+(\frac{\alpha_r q_r+\alpha_s q_s+\alpha_t q_t+\ldots.}{n})+\frac{(\alpha_r q_r+\alpha_s q_s+\alpha_t q_t+\ldots.)^2}{n^2\ 2!}+\ldots.]\}^n .$$

This can be simplified further to the equation:

$$E\{1+(\alpha_r Q_r+\alpha_s Q_s+\alpha_t Q_t+\ldots.)+\frac{(\alpha_r Q_r+\alpha_s Q_s+\alpha_t Q_t+\ldots.)^2}{2!}+\ldots.\}$$

$$= 1+n\{\frac{\theta_2}{n^2 2!}+\frac{\theta_3}{n^3 3!}+\ldots.\}+\frac{n(n-1)}{2}\{\frac{\theta_2}{n^2 2!}+\frac{\theta_3}{n^3 3!}+\ldots.\}^2$$

$$+\frac{n(n-1)(n-2)}{6}\{\frac{\theta_2}{n^2 2!}+\frac{\theta_3}{n^3 3!}+\ldots.\}^3+\ldots.$$

$$+\frac{n(n-1)\ldots.(n-k+1)}{k!}\{\frac{\theta_2}{n^2 2!}+\frac{\theta_3}{n^3 3!}+\ldots.\}^k+\ldots. \quad ,(2\text{-}22)$$

where:

$$\theta_2 = \alpha_r^2\varphi_r + \alpha_s^2\varphi_s + \alpha_t^2\varphi_t + \ldots.$$

$$\theta_3 = \alpha_r^3[r^3] + \alpha_s^3[s^3] + \alpha_t^3[t^3] + 3\alpha_r^2\alpha_s[r^2s] + \ldots.$$

$$+ 6\alpha_r\alpha_s\alpha_t[rst]\ldots. \qquad \text{and so on.}$$

By actually performing the expectation on the left-hand side of (2-22), one can equate coefficients of appropriate products of powers of the α's and obtain the desired expected Q-products. (Of course to obtain the final values to be used in determining the moments of t, one must also evaluate the $[r^\alpha s^\beta t^\gamma \ldots.]$. This will be discussed subsequently.) Equation (2-22) leads to the general form as follows:

$$(r^\alpha s^\beta t^\gamma \ldots.) = \sum_{\lambda=a}^{b} (r^\alpha s^\beta t^\gamma \ldots.)_\lambda / n^\lambda \quad , \qquad (2\text{-}23)$$

where $a = -1+\alpha+\beta+\gamma+\ldots$ and $b = \{\frac{1+\alpha+\beta+\gamma+\ldots.}{2}\}$.

($\{z\}$ refers to the integral part of z.)

<u>Examples</u>

We can use (2-22) in conjunction with the notation of (2-23) to obtain the following expected Q-products:

(i) $(r^3) = (r^3)_2/n^2$

$= [r^3]/n^2$

(ii) $(r^4 s) = (r^4 s)_4/n^4 + (r^4 s)_3/n^3$

$= \{[r^4 s]-6[r^2][r^2 s]\}/n^4 + 6[r^2][r^2 s]/n^3$.

(c) <u>Simplified Method for Determining Expected Q-Products</u>

An interesting simplification for obtaining $(r^\alpha s^\beta t^\gamma \ldots.)$ can be noted here. Many of these can be obtained by applying

Kendall's [21] symbolic operator technique. We shall show an example of the technique and then proceed to generalize on it. Assume that by the previously described method one has obtained

$$(r^4) = 3[r^2]^2/n^2 + \{[r^4]-3[r^2]^2\}/n^3 \quad .$$

Then by applying to this the operator $s(\partial/\partial_r)$, we can write

$$4(r^3s) = 12[rs][r^2]/n^2 + \{4[r^3s]-12[r^2][rs]\}/n^3 \quad .$$

However, from orthogonality properties, $[rs] = 0$ and thus

$$(r^3s) = [r^3s]/n^3 \quad .$$

For another example, suppose one had arrived at the following:

$$(r^3s^2) = \{[r^3][s^2] + 3[r^2][rs^2]\}/n^3 + (r^3s^2)_4/n^4$$

for $r \neq s$.

By applying the operator $t(\partial/\partial_s)$ and simplifying, we have

$$(r^3st) = 3[r^2][rst]/n^3 + (r^3st)_4/n^4 \;, \quad (r \neq s \neq t) \;.$$

Thus if one wishes to obtain all expected Q-products of order, say k, one can begin with a known expression for (r^k) and by continuously applying the proper differential operator, derive all $(r^{k_1} s^{k_2} t^{k_3} \ldots .)$ for which $k_1+k_2+\ldots . = k$. One must be cautious, however, not to drop terms that are zero in intermediate stages in deriving expressions in later stages. For example, $[rs] = 0$ but will certainly make a contribution under the operator $s(\partial/\partial_r)$. A rule for one to follow here is to allow oneself to apply an operator of the type $s(\partial/\partial_r)$ to any existing final expression which is not conditioned by $r \neq s$. Actually the work involved here becomes quite lengthy for the higher powers of the Q-products. Thus a combination of the two methods described was used in obtaining the results in Appendix A. Appendix A gives a "library" of expected Q products in terms of expected q-products for any four Q's (denoted by Q_r, Q_s, Q_t, Q_u) through order 10. This will, of course, enable one to expand the moments of the distribution of any statistic involving any four sample moments to powers of n^{-5}.

(d) <u>Expected Value of q-products</u>

It should be noted at this point that one cannot truly evaluate the moments of a statistic via the "orthogonal statistic" method until one has a library of expected q-products or a library of expected Q's evaluated in terms of the population parameters.

All q-product expectations can be derived from a manipulation of the fundamental recurrence relationship given by (2-5). The preferred form of (2-5) is

$$q_1 q_r = q_{r+1} + c_{r+1} q_r + b_{r+1} q_{r-1} \quad , \qquad (r=1,2,\ldots.) \tag{2-24}$$

where $c_{r+1} = (a_{r+1} - a_0)$. This expression was obtained by replacing x by $q_1 + a_0$ in (2-5) and solving for $q_1 q_r$. (Strictly speaking, an alternative form of (2-5) was used; namely, one whereby r was replaced by r+1.) Repeated multiplication of both sides of (2-24) by appropriate q's gives formulae whereby one can obtain values of $[r^{\alpha} s^{\beta} t^{\gamma} \ldots.]$. Actually a generalized form of (2-24) can be written as follows:

$$q_r q_s = a_0^{(r,s)} q_{r+s} + a_1^{(r,s)} q_{r+s-1} + a_2^{(r,s)} q_{r+s-2} + \ldots. + a_{2s}^{(r,s)} q_{r-s} \qquad (r \geq s) \; , \tag{2-25}$$

where $a_0^{(r,s)} = 1$. (Further discussion on these expected q-products for the negative binomial distribution is given in Chapter IV.) The above equation is extremely useful. For instance from (2-25) and the orthogonality conditions, we have

$$[r\ s\ t] = \begin{cases} 0 \; , & \text{for } t > r+s, \; t < r-s \\ a_{r+s-t}^{(r,s)} \varphi_t \; , & \text{for } r-s \leq t \leq r+s \end{cases} \; .$$

Similarly, we could write

$$[r\ s\ t\ u] = E\{\sum_{\lambda=0}^{2s} a_\lambda^{(r,s)}\ q_{r+s-\lambda}\ \sum_{v=0}^{2u} a_v^{(t,u)}\ q_{t+u-v}\}\quad ,$$

where we assume $r \geq s$ and $t \geq u$.

Of course, orthogonality will remove certain terms on taking expectations. A continuation of this process will deal with expectations of high degree q-products. Of course, the magnitude of the degree of these q-products depends on how many sampling moments are involved in the statistic t and how many population moments of this statistic are needed. (In our work with the negative binomial distribution, we consider statistics which are functions of m_1' and m_2' with the first four moments of the statistics expanded to terms involving n^{-4}.)

As an illustration, consider a situation in which it is desired to find various values of the expectations of q-products for q_1 and q_2 up to degree six. One can obtain all of these by first deriving a few of the basic q-products as linear combinations of the individual q's. For example, from (2-24), by setting r equal to the appropriate values,

$$q_1^2 = b_2q_0 + c_2q_1 + q_2\quad ,$$

$$q_1q_2 = q_3 + c_3q_2 + b_3q_1\quad ,$$

where the coefficients are those defined by the original recurrence relationship of (2-5). We can also manipulate (2-24) to obtain

$$\begin{aligned} q_1^2 q_2 &= q_3 q_1 + c_3 q_2 q_1 + b_3 q_1^2 \\ &= q_4 + c_4 q_3 + b_4 q_2 + c_3(q_3 + c_3 q_2 + b_3 q_1) \\ &\quad + b_3(q_2 + c_2 q_1 + b_2 q_0) \\ &= q_4 + q_3(c_4 + c_3) + q_2(b_4 + c_3^2 + b_3) + q_1(c_3 b_3 + b_3 c_2) \\ &\quad + b_3 b_2 q_0 \quad . \end{aligned}$$

Similarly q_2^2, q_1^3, $q_1 q_2^2$, and q_2^3 are obtained and these seven q-products can be written as follows:

$$q_1^2 = \sum_{\lambda=0}^{2} A_\lambda q_\lambda \qquad q_1^2 q_2 = \sum_{\lambda=0}^{4} G_\lambda q_\lambda$$

$$q_1 q_2 = \sum_{\lambda=1}^{3} B_\lambda q_\lambda \qquad q_1 q_2^2 = \sum_{\lambda=0}^{5} H_\lambda q_\lambda$$

$$q_2^2 = \sum_{\lambda=0}^{4} C_\lambda q_\lambda \qquad q_2^3 = \sum_{\lambda=0}^{6} I_\lambda q_\lambda$$

$$q_1^3 = \sum_{\lambda=0}^{3} F_\lambda q_\lambda \quad ,$$

in which, in terms of the coefficients in (2-24),

$$A_2 = 1 \ , \quad A_1 = c_2 \ , \quad A_0 = b_2 \ ;$$

$$B_3 = 1 \ , \quad B_2 = c_3 \ , \quad B_1 = b_3 \ ;$$

$$C_4 = 1\ ,\quad C_3 = c_4 + c_3 - c_2\ ,\quad C_2 = b_4 + b_3 - b_2$$
$$+ c_3(c_3-c_2)\ ,\quad C_1 = b_3c_3\ ,\quad C_0 = b_3b_2\ ;$$

$$F_3 = 1\ ,\quad F_2 = C_2 + C_1\ ,\quad F_1 = b_3 + b_2 + c_2^2\ ,$$
$$F_0 = b_2c_2\ ;$$

$$G_4 = 1\ ,\quad G_3 = c_4 + c_3\ ,\quad G_2 = b_4 + b_3 + c_3^2\ ,$$
$$G_1 = b_3(c_3+c_2)\ ,\quad G_0 = b_3b_2\ ;$$

$$H_5 = 1\ ,\quad H_4 = c_5 + C_3\ ,\quad H_3 = b_5 + c_4C_3 + C_2\ ,$$
$$H_2 = b_4C_3 + c_3(C_2+b_3)\ ,\quad H_1 = b_3(b_4+b_3+c_3^2)\ ,$$
$$H_0 = b_3b_2c_3\ ;$$

$$I_6 = 1\ ,\quad I_5 = H_4 + c_6 - c_2\ ,\quad I_4 = b_6 + b_5 - b_2 + c_5(c_5-c_2)$$
$$+ C_3(c_5+c_4-c_2) + C_2\ ,$$
$$I_3 = b_5(c_5+c_4-c_2)+C_3(b_5+b_4-b_2+c_4^2-c_2c_4)+b_3c_3+C_3C_2\ ,$$
$$I_2 = b_5b_4 + b_3c_3^2 + b_3b_2 + C_2^2 + b_4C_3^2\ ,$$
$$I_1 = b_4b_3C_3 + b_3c_3C_2 + b_3^2c_3\ ,\quad I_0 = b_3b_2C_2\ .$$

We then use these seven basic linear combinations to derive the desired expected q-products as follows:

$$[1^3] = A_1\varphi_1 = F_0\ ,\quad [1^22] = A_2\varphi_2 = b_2\varphi_1 = G_0\ ,$$

$$[12^2] = B_2\varphi_2 = C_1\varphi_1 = H_0\ ,\quad [2^3] = C_2\varphi_2 = I_0\ ;$$

$$[1^4] = \sum_0^2 A_\lambda^2\varphi_\lambda = F_1\varphi_1\ ,\quad [1^32] = \sum_1^2 A_\lambda B_\lambda\varphi_\lambda = F_2\varphi_2 = G_1\varphi_1\ ,$$

$$[1^2 2^2] = \sum_0^2 A_\lambda C_\lambda \varphi_\lambda = \sum_1^3 B_\lambda \varphi_\lambda = G_2 \varphi_2 = H_1 \varphi_1 \ ,$$

$$[12^3] = \sum_1^3 B_\lambda C_\lambda \varphi_\lambda = H_2 \varphi_2 = I_1 \varphi_1 \ , \quad [2^4] = \sum_0^4 C_\lambda \varphi_\lambda = I_2 \varphi_2 \ ;$$

$$[1^5] = \sum_0^2 A_\lambda F_\lambda \varphi_\lambda \ , \quad [1^4 2] = \sum_1^3 B_\lambda F_\lambda \varphi_\lambda = \sum_0^2 A_\lambda G_\lambda \varphi_\lambda \ ,$$

$$[1^3 2^2] = \sum_0^2 A_\lambda H_\lambda \varphi_\lambda = \sum_1^3 B_\lambda G_\lambda \varphi_\lambda = \sum_0^3 C_\lambda F_\lambda \varphi_\lambda \ ,$$

$$[1^2 2^3] = \sum_0^2 A_\lambda I_\lambda \varphi_\lambda = \sum_1^3 B_\lambda H_\lambda \varphi_\lambda = \sum_0^4 C_\lambda G_\lambda \varphi_\lambda \ ,$$

$$[12^4] = \sum_1^3 B_\lambda I_\lambda \varphi_\lambda = \sum_0^4 C_\lambda H_\lambda \varphi_\lambda \ , \quad [2^5] = \sum_0^4 C_\lambda I_\lambda \varphi_\lambda \ ;$$

$$[1^6] = \sum_0^3 F_\lambda^2 \varphi_\lambda \ , \quad [1^5 2] = \sum_0^3 F_\lambda G_\lambda \varphi_\lambda \ , \quad [1^4 2^2] = \sum_0^3 F_\lambda H_\lambda \varphi_\lambda \ ,$$

$$[1^3 2^3] = \sum_0^3 F_\lambda I_\lambda \varphi_\lambda \ , \quad [1^2 2^4] = \sum_0^4 G_\lambda I_\lambda \varphi_\lambda \ ,$$

$$[12^5] = \sum_0^5 H_\lambda I_\lambda \varphi_\lambda \ , \quad [2^6] = \sum_0^6 I_\lambda^2 \varphi_\lambda \ .$$

It might be noted that these constants, i.e., the A's, B's, ..., H's, I's, and the φ's are in terms of the population parameters and could easily be introduced into a program for a high-speed computer. This would greatly simplify the work involved in obtaining desired moments of functions of Q_1 and Q_2. Note also that a number of checks can be applied in working out the above expected q-products and the same is true when further q's are involved.

(e) Examples

We shall consider two particular examples at length in which moments of certain statistics are found using the proposed method of orthogonal statistics. Consider first the origin and scale-free statistics $\sqrt{b_1} = m_3/m_2^{3/2}$, where m_3 is the third central sample moment for a sample of size n from a N(0,1) population, and, of course, m_2 is the second central sample moment from the same population.
E. S. Pearson [23] actually worked out the variance of the related statistic

$$y = [(n-1)/\sqrt{6(n-2)}\ ]\sqrt{b_1} \quad .$$

The expansion is as follows:

$$\mu_2(y) = 1 - \frac{6}{n} + \frac{22}{n^2} - \frac{70}{n^3} + \dots \quad . \tag{2-26}$$

Using the Q-statistics as defined in (2-15), we find that

$$m_1 = Q_1$$

$$m_2 = 1 + Q_2 - Q_1^2$$

$$m_3 = Q_3 - 3Q_1Q_2 + 2Q_1^3 \quad ,$$

so that

$$\mu_2(y) = Ey^2 = E\{\frac{n(1-1/n)^2(Q_3-3Q_1Q_2+2Q_1^3)^2}{6(1-2/n)(1+Q_2-Q_1^2)^3}\} ,$$

since $Ey = 0$.

If we expand this in terms of powers of the Q's, we find that

$$Ey^2 = f(n) \cdot E\{Q_3^2-Q_1Q_2Q_3-3Q_2Q_3^2+3Q_1^2Q_3^2+18Q_1Q_2^2Q_3+9Q_1^2Q_2^2 +4Q_1^3Q_3+6Q_2^2Q_3^2+\dots\} . \tag{2-27}$$

If we take expectations term by term and use the dictionary of Appendix A, we find that

$$Ey^2 = 1 - \frac{6}{n} + \frac{22}{n^2} + \dots .$$

If we added another 43 terms to (2-26), we would cover all Q-products that would generate terms in n^{-3} and would thus give the last term in (2-26).

Again consider the sample coefficient of variation

$$v = \frac{s}{m_1} , \quad \text{where} \quad s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} .$$

Here we consider $x_1,x_2,\dots,x_n$ a random sample from a normal population with mean $\mu_1 \neq 0$ and variance μ_2. F. N. David [9] found the first four moments of v/V (V is the population coefficient of variation) through terms in n^{-2}. The following paragraphs show an example of the determination of the mean and variance of this statistic using orthogonal statistics.

Expanding v/V in terms of Q-statistics, we have

$$v/V = \frac{\sqrt{n/n-1}\ \{1+[Q_2/\mu_2 + \mu_3 Q_1/\mu_2^2 - Q_1^2/\mu_2]\}^{\frac{1}{2}}}{(1 + Q_1/\mu_1)} \quad , \qquad (2\text{-}28)$$

where the Q's were obtained in terms of the m's from the following equations:

$$m_1 = Q_1 + \mu_1 \quad ,$$

$$m_2' = Q_2 + (\mu_3/\mu_2)Q_1 + \mu_2 + 2\mu_1 m_1 - \mu_1^2 \quad .$$

These were derived for the $N(\mu_1,\mu_2)$ case using (2-1) and (2-2) and the definitions of the Q's in terms of the q's.

$$\begin{aligned} v/V = & \{1 + 1/2n + \ldots.\} \cdot \{1 + 1/2[Q_2/\mu_2 + \mu_3 Q_1/\mu_2^2 \\ & - Q_1^2/\mu_2] - 1/8[Q_2/\mu_2 + \mu_3 Q_1/\mu_2^2 - Q_1^2/\mu_2]^2 + \ldots.\} \\ & \cdot \{1 - Q_1/\mu_1 + Q_1^2/\mu_1^2 + \ldots.\} \end{aligned}$$

$$\begin{aligned} = & \{1 + 2/n + \ldots.\} \cdot \{1 + L(Q) - 1/2(Q_1^2/\mu_2) \\ & - 1/8(Q_2^2/\mu_2^2) - 1/4(\mu_3 Q_1 Q_2/\mu_2^3) - 1/8(\mu_3^2 Q_1^2/\mu_2^4) \\ & - 1/2[\mu_3 Q_1^2/(\mu_2^2\mu_1)] + Q_1^2/\mu_1^2 + \ldots.\} \quad , \qquad (2\text{-}29) \end{aligned}$$

where $L(Q)$ represents linear terms in Q_1 and Q_2 which have zero expectation. If we take expectations of the right-hand side of (2-29) we obtain

$$E(v/V) = \{1 + 2/n + \ldots.\} \cdot \{1 - 1/n[1/2 + \mu_3/(2\mu_2\mu_1)$$

$$+ \mu_3^2/(8\mu_2^3) - \mu_2/\mu_1^2 - 1/8 + \mu_4/(8\mu_2^2)$$

$$- \mu_3^2/(8\mu_2^3)] + \ldots.\} \quad . \qquad (2\text{-}30)$$

This expression was found by using Appendix A. Here $q_1 = x - \mu_1$ and $q_2 = (x-\mu_1)^2 - (\mu_3/\mu_2)(x-\mu_1) - \mu_2$.

Finally, simplification of (2-30) yields:

$$E(v/V) = 1 - 1/4n + 1/n[-k_3/(2k_2k_1) + k_2/k_1^2 - k_4/(8k_2^2)]$$

$$+ s_2^{(1)}/n^2 + s_3^{(1)}/n^3 + \ldots. \quad , \qquad (2\text{-}31)$$

where $k_1 = \mu_1$,

$k_2 = \mu_2$,

$k_3 = \mu_3$,

$k_4 = \mu_4 - 3\mu_2^2$,

$s_2^{(1)}$ = term in n^{-2} for first moment of v/V , and

$s_3^{(1)}$ = term in n^{-3} for first moment of v/V .

The expression is given in terms of cumulants to facilitate comparison with David's results. This method was actually used to obtain the $1/n^2$ term which checked with that given by David.

In finding the variance of the statistic v/V, we have

$$(v/V)^2 = (n/n-1)\{1 + Q_2/\mu_2 + \mu_3 Q_1/\mu_2^2 - Q_1^2/\mu_2\}$$

$$\cdot\ \{1 - 2Q_1/\mu_2 + 3Q_1^2/\mu_1^2 - 4Q_1^3/\mu_1^3$$

$$+ 5Q_1^4/\mu_1^4 + \ldots\} \quad .$$

$$E(v/V)^2 = (n/n-1)E\{1 + \pi_1 + \pi_2 + \pi_3 + \ldots\} \quad , \qquad (2\text{-}32)$$

where: π_1 = that portion of the statistic which gives terms in n^{-1} ,

π_2 = that portion of the statistic which gives terms in n^{-2} ,

etc.

We shall proceed to obtain the expansion of the variance, in illustrating the use of these Q's, through the n^{-1} term. Thus, in (2-32) we need π_1.

$$\pi_1 = 3Q_1^2/\mu_1^2 - 2(Q_1/\mu_1)(Q_2/\mu_2 + \mu_3 Q_1/\mu_2^2) - Q_1^2/\mu_2 \quad .$$

Thus, taking expectations as indicated in (2-32), we have

$$E(v/V)^2 = \{n/n-1\}\cdot\{1 + 3(1^2)_1/\mu_1^2 - \frac{2(12)_1}{\mu_1\mu_2} - \frac{2\mu_3(1^2)_1}{\mu_1\mu_2^2}$$

$$- (1^2)_1/\mu_2 + \ldots\} \quad ,$$

$$= \{1 + 1/n + 1/n^2 + \ldots\}\cdot\ \{1 + [\frac{3\mu_2/\mu_1^2 - 2\mu_3/(\mu_1\mu_2) - 1}{n}]$$

$$+ \ldots\} \quad , \qquad (2\text{-}33)$$

$$= 1 + \frac{3\mu_2/\mu_1^2 - 2\mu_3/(\mu_1\mu_2)}{n} + \ldots \quad .$$

After subtracting the correction term $[E(v/V)]^2$ from this we have:

$$\mathrm{Var}(v/V) = 1/n[k_2/k_1^2 - k_3/(k_1k_2) + 1/2 + k_4/(4k_2^2)]$$

$$+ S_2^{(2)}/n^2 + S_3^{(2)}/n^3 + \ldots . \qquad (2\text{-}34)$$

Once again the n^{-2} term was also determined. The details are not shown, but the procedure would be the same as outlined, i.e., the expansion of v/V and $(v/V)^2$ would be carried further in powers of the Q's to include those terms of degree four. Expectations of this expansion along with the use of the table in Appendix A yields the following results:

$$S_2^{(2)} = 3/8 + 8k_2^2/k_1^4 - 10k_3/k_1^3 + k_2/k_1^2 + 5/2\ k_4/(k_1^2k_2)$$

$$+ 10\ k_2/k_1^2 + 7/32\ k_4^2/k_2^4 + 1/4\ k_4k_3/(k_1k_2^3)$$

$$- 1/8\ k_6/k_2^3 + 1/4\ k_3^2/(k_1^2k_2^2) - 1/4\ k_5/(k_1k_2^2)$$

$$- 5/8\ k_4/k_2^2 - 1/2\ k_3/(k_2k_1) - 1/2\ k_3^2/k_2^3 \quad .$$

CHAPTER III

MODELS, FORMS, AND ESTIMATORS FOR NEGATIVE BINOMIAL DISTRIBUTION

(a) Physical Models

The negative binomial distribution is one of the more widely used two-parameter distributions. It has application in many practical areas, including work with biological data, accident statistics, psychological data, reliability studies, etc. A literature review on this distribution was done by Bartko [5].

There is a natural similarity of the negative binomial to the positive binomial distribution. The basic positive binomial expansion with parameters p and n is, of course, $(q+p)^n$, while the corresponding expansion for the negative binomial with parameters p and k (notation of Fisher [13]) is $(q-p)^{-k}$, where $q = 1 + p$, $p > 0$, and $k > 0$. Gurland [16] suggests an interpretation of the probability term,

$$Pr(X=x) = \left(\frac{1}{q}\right)^k \frac{k(k+1)\dots(k+x-1)}{x!} \left(\frac{p}{q}\right)^x , \quad x=0,1,2\dots \quad .$$

If we imagine independent trials with $1/q$ as the probability of a "success" and P/q as the probability of a "failure", then the expression is the probability that x+k trials will be required to obtain k successes. According to Gurland, the negative binomial was actually formulated by Montmort in 1714.

One of the more significant properties of this distribution in that its variance ($kp + kp^2$ in the above form) exceeds its mean (kp). The Poisson distribution may be regarded as a limiting case of the negative binomial. If we let $k \to \infty$, $p \to 0$ while keeping the mean constant, equal to say λ, then the negative binomial probability function becomes that of the Poisson with parameter λ.

There are various physical models which give rise to the negative binomial distribution. The first is described by Feller [11] as "apparent contagion". The application here is that Poisson Distributions may be compounded, i.e., the parameter λ varies from case to case and is considered to have some statistical population. If the parameter λ has a Gamma frequency function given by

$$f(\lambda) = \frac{\alpha^k}{\Gamma(k)} \lambda^{k-1} e^{-\alpha\lambda} \quad , \qquad \lambda > 0, \ \alpha > 0, \ k > 0 \ ,$$

then the derived frequency generating function (parameter t) is

$$\int_0^\infty e^{\lambda(t-1)} f(\lambda)d\lambda = [1+(1-t)/\alpha]^{-k} \quad , \quad (\alpha + t > 1) \; ,$$

with probability function

$$Pr(X=x) = \frac{k(k+1)....(k+x-1)}{x!} \left(\frac{\alpha}{1+\alpha}\right)^k \left(\frac{1}{1+\alpha}\right)^x \quad .$$

If we set $p = 1/\alpha$ and $q = (1+\alpha)/\alpha$, then the above becomes identically equal to the negative binomial probability function. Feller also discussed "true" contagion, i.e., the situation in which the probability of a success depends on previous favorable events. For example, consider sampling from a binomial population containing N objects in which p is the probability of a success and q the probability of a failure. Suppose also after each observation is taken, we replace the object and likewise add $N\delta$ objects of the same type. Then if we denote by X the random variable representing the number of successes in n drawings, then

$Pr(X=x)$

$$=\binom{n}{x}\frac{[p][p+\delta][p+2\delta]....[p+(x-1)\delta][q][q+\delta][q+2\delta]....[q+(n-x+1)\delta]}{1 \cdot [1+\delta][1+2\delta]....[1+(n-1)\delta]} \; .$$

If we let $n \to \infty$, $p \to 0$, $\delta > 0$, and keep $np = \lambda$, $n\delta = \eta$ constant, then the above probability function becomes

$$\Pr(X=x) = \frac{k(k+1)\ldots(k+X-1)}{x!}\left(\frac{\eta}{1+\eta}\right)^{x}\left(\frac{1}{1+\eta}\right)^{\lambda/\eta} ,$$

which is a negative binomial upon setting $\lambda/\eta = k$.

Another model introduced by Kendall [19] which gives rise to the negative binomial is that of immigration and birth-death processes.

(b) Forms

There are a number of distinct forms of the negative binomial distribution. Thus, because of this and because of its varied use, a number of estimation problems arise. We are going to consider in particular three forms; those discussed by Fisher [13], Anscombe [1], and Evans [10]. Table 1 shows the probability function, mean, variance, and factorial moment generating function in the case of the three forms. It might be noted here that in Anscombe's form the special case of the Poisson distribution arises when α becomes infinite, i.e., the mean λ equals the variance. Note also that Table 1 contains the first two moment estimators for the three forms of the negative binomial. These are the estimators which are of primary interest here and will be discussed at length in Chapter V.

Table 1. Interesting Forms for Negative Binomial Distribution

Property	Anscombe	Fisher	Evans
*Prob. Function P_x	$\binom{\alpha+x-1}{x}\dfrac{\alpha^{\alpha}\lambda^{x}}{(\lambda+\alpha)^{\alpha+x}}$ Parameters λ,α	$\binom{k+x-1}{x}\dfrac{p^{x}}{q^{k+x}}$ Parameters k,p $q = p+1$	$\binom{m/a+x-1}{x}\dfrac{a^{x}}{(1+a)^{m/a+x}}$ Parameters m,a
F.M.G.F. parameter t	$(1 - \lambda t/\alpha)^{-\alpha}$	$(1 - pt)^{-k}$	$(1 - at)^{-m/a}$
Mean	λ	kp	m
Variance	$\lambda + \lambda^2/\alpha$	$pk(1+p)$	$m(1+a)$
First Two Moment Estimators	$\hat{\lambda} = m_1'$ $\hat{\alpha} = m_1'^2/(m_2-m_1')$	$\hat{p} = (m_2-m_1')/m_1'$ $\hat{k} = m_1'^2/(m_2-m_1')$	$\hat{m} = m_1'$ $\hat{a} = (m_2-m_1')/m_1'$

*The combinatorial part of P_x is taken to be unity when $x = 0$.

Other Notation: (to be used throughout the text)

$$m_1' = \sum_{i=1}^{n} x_i/n \quad , \qquad m_j' = \sum_{i=1}^{n} x_i^{\,j}/n \quad , \qquad m_j = \sum_{i=1}^{n} (x_i-m_1')^j/n$$

$$\mu = E(X) \quad , \qquad \mu_j' = EX^j \quad , \qquad \mu_j = E(X-\mu)^j$$

(c) Estimators

Fisher [13] actually discusses the problem of estimation and, in particular, the problem of fitting a negative binomial by maximum likelihood methods. He also gives some discussion on the asymptotic efficiency of moment estimators. Anscombe [1] discusses similar estimation problems with respect to the negative binomial and other contagious distributions; namely Neyman Types A, B, and C; Polya-Aeppli; and the discrete lognormal.

Anscombe actually suggests, other than that of maximum likelihood, four primary methods of estimation for the parameters of the negative binomial. Here $\hat{\lambda} = m_1'$ in all cases.

Second Moment

$$\begin{aligned} \hat{\lambda} &= m_1' \\ \hat{\alpha} &= m_1'^2/(m_2 - m_1') \end{aligned} \tag{3-1}$$

This is the first set of moment estimators we will discuss at length in Chapter V.

<u>Zeroth Frequency</u>

$$n_0/n = (1+m_1'/\hat{\alpha}_1)^{-\hat{\alpha}_1} \quad , \tag{3-2}$$

where: n_0 = number of zeros observed in the sample,

n = sample size.

<u>Inverse Moment</u>

$$1/n \sum_{x=0}^{\infty} \frac{n_x}{x+1} = \frac{(1-\hat{k}) - (1-\hat{k})\hat{\alpha}_2}{(\hat{\alpha}_2-1)\hat{k}} \quad , \tag{3-3}$$

where: $\hat{k} = m_1'/(m_1'+\hat{\alpha}_2)$,

n_x = number of times x occurs in the sample.

<u>Geometric Moment</u>

$$1/n \sum_{x=0}^{\infty} n_x C^x = \{1+m_1'(1-C)/\hat{\alpha}_3\}^{-\hat{\alpha}_3}$$

for a given value of C not equal to unity. For Anscombe's notation, the maximum likelihood estimator of α, say α^*, is given by a root of the equation

$$n \,\ell n(1+m_1'/\alpha^*) = \sum_{j=1}^{\infty} n_j [1/\alpha^* + 1/(\alpha^*+1) + \ldots + 1/(\alpha^*+j-1)] \;. \tag{3-4}$$

Haldane [17] gives an iterative technique for the solution of (3-4). We do not propose to actually investigate properties of α^*. However, some of them are given (see Chapter VI) for purposes of comparison with corresponding properties of the moment estimators.

Evans investigated a large amount of experimental data involving plant and insect quadrat counts. The evidence essentially indicated that for plant counts the Neyman Type A Distribution fits fairly well while for insect counts the negative binomial fits best. Evans, as indicated by Table 1, considers the parameters m and a for the negative binomial, where in reference to the form of Anscombe,

$$m = \lambda \quad ,$$

and $$a = \lambda/\alpha \quad .$$

Two general methods of estimation were considered for the negative binomial, Neyman Type A and Polya-aeppli distributions. These methods were; (a) the use of the mean and variance, and (b) the use of the mean and a proportion of zeros. Evans points out that neither method may be efficient in some cases and hence suggests the use of a third, namely an estimate which is a weighted function of the first two. These three methods yield the following estimators:

<u>Use of Mean and Variance</u>

$$\hat{m} = m_1' \quad ,$$

$$\hat{a} = (m_2 - m_1')/m_1' \quad .$$

<u>Mean and Proportion of Zeros</u>

$\hat{m} = m_1'$,

$\hat{a}_1$ is the solution of the equation ,

$$\hat{a}_1/\ell n(1+\hat{a}) = m_1'/\ell n(n/n_0) \quad . \qquad (3\text{-}5)$$

<u>Combined Estimate</u>

Evans calls this estimator a_w, defined by

$$\hat{a}_w = \hat{a} + w(\hat{a}_1 - \hat{a}) \quad ,$$

where: $\hat{a}$ = estimate found by using moments ,

$\hat{a}_1$ = estimate found by using proportion of zeros ,

$$w = \frac{\text{Var } \hat{a} - \text{Cov}(\hat{a}, \hat{a}_1)}{\text{Var } \hat{a} - 2 \text{ Cov}(\hat{a}, \hat{a}_1) + \text{Var } \hat{a}_1} \quad .$$

Contours of w are given by Evans.

There are many other estimators which will not be mentioned here.

(d) <u>Practical Work with the Negative Binomial</u>

The practical work in the paper by Bliss and Fisher [6] involved counts of a number of European red mites on apple leaves. The data revealed that a fairly close fit could be made with the negative binomial. The form used was actually that of Anscombe, i.e., with parameters λ and α. It is proposed that in estimating α, moment estimation, i.e., $\hat{\alpha} = m_1'^2/(m_2'-m_1'-m_1'^2)$ can safely be used for $\alpha > 13$ and $\alpha/\lambda > 6$ since this results in an efficiency (actually the author means asymptotic efficiency) of greater than 90%. Bliss and Fisher propose that under low efficiency conditions for moment estimation an alternative estimate such as that described by (3-2) might be used. It turns out that the efficiency of the latter estimator is greater than 90% if more than 1/3 of the units are empty, or in other words, 1/3 or more of the observations are zeros.

Bliss and Fisher actually proceed to fit a set of data using both kinds of estimators. The sample size used was $n = 150$. The variance (large sample variance) is quoted for these estimators and the given set of data. (See Chapter V for a discussion of the paradox in quoting the usual large sample variances of the moment estimator $\hat{\alpha}$.)

Perhaps one of the more intriguing pieces of work on use of the negative binomial was that of Arbous [3] on accident statistics and the concept of accident proneness. It is hypothesized that the distribution of the accidents incurred might be due to simple chance. It is theorized under the hypothesis that; (i) if the environmental circumstances are homogeneous for all individuals, and (ii) individuals are homogeneous with respect to certain qualities that might generate reasons for accidents to occur, then the theoretical distribution of accidents might follow a Poisson probability law. However an alternative and more realistic hypothesis is stated which essentially says that people do vary with respect to certain qualities and environment which brings about changes in accident proneness from person to person. Likewise, having had previously sustained a certain accident, a person's accident proneness is changed and thus the homogeneity conditions do not hold. This hypothesis is due to Greenwood and Yule [15]. This brings about the use of the negative binomial. Some accident data is given to which the negative binomial was successfully fitted.

The above examples represent only a small sample of uses of the negative binomial distribution. Table 2 gives a larger collection of practical examples. Note that the values of the parameters (in Anscombe's notational form) are given for each example. These, of course, are not the true parameters for the latter are not known. The values in the tables are in fact based on moment estimates of the parameters.

Table 2. Some Examples of Parameter Values
In Practical Cases

Author or (and) Experimenter	λ	α	n
Evans and Steiger [10]	39.00	1.70	40
" "	38.00	2.40	"
" "	16.00	1.00	"
" "	25.00	0.70	"
" "	16.00	1.00	"
" "	5.00	0.69	"
" "	111.00	8.46	"
" "	6.97	1.13	"
" "	1.73	3.84	"
" "	9.88	2.30	"
" "	1.82	1.00	"
Evans and Hanson [10]	1.40	2.12	325
" "	0.50	3.12	"
" "	22.80	1.58	384
" "	5.75	1.26	782
" "	4.03	1.31	120
" "	2.48	1.25	780
Fisher [13]	3.25	3.96	60
Fisher and Bliss [6]	1.15	1.17	150

CHAPTER IV

GENERAL USE OF ORTHOGONAL STATISTICS IN THE NEGATIVE BINOMIAL DISTRIBUTION

One of the primary purposes of this dissertation is to point out certain previously unknown characteristics of estimators of parameters of the negative binomial. Thus it is fitting to attempt to discuss here the application of certain of the concepts mentioned in Chapter II to the negative binomial distribution. Then in Chapter V, the results of the study of these estimators will be discussed.

(a) Relationships between Q's and Sample Moments

Consider the form of the negative binomial with parameters λ and α, (see Chapter III), i.e., Anscombe's form. Through use of the determinants in eq. (2-3) and (2-4) and from the definitions of the Q's, the following expressions can be derived:

$$Q_1 = m_1' - \lambda \quad ,$$

$$Q_2 = m_2' - m_1' - 2(1+1/\alpha)\lambda m_1' + \lambda^2(1+1/\alpha) \quad , \quad \text{and}$$

$$Q_3 = m_3' - 3m_2' + 2m_1' - 3(1+2/\alpha)\lambda(m_2'-m_1')$$
$$+ 3(1+2/\alpha)(1+1/\alpha)\lambda^2 m_1' - (1+2/\alpha)(1+1/\alpha)\lambda^3 \quad .$$

There is a form from which one can obtain the general Q_k, namely the difference operator expression:

$$q_k = (1-\lambda/\alpha\ \Delta)^{\alpha+k-1}\ X^{(k)} \quad , \tag{4-1}$$

which leads to the expression for the Q's,

$$Q_k = 1/n \sum_{j=1}^{n} (1-\lambda/\alpha\ \Delta)^{\alpha+k-1}\ X^{(k)} \Big|_{X=x_j} \quad , \tag{4-2}$$

where x_j represents the j'th observation from a sample $(x_1, x_2, \ldots, x_n)$. Here we denote $X^{(k)}$ as $X(X-1)\ldots(X-k+1)$. For example, in finding Q_2 we would expand as follows:

$$Q_2 = 1/n \sum_{j=1}^{n} (1-\lambda/\alpha\ \Delta)^{\alpha+1}\ x_j^{(2)} \quad ,$$

$$= 1/n \sum_{j=1}^{n} \{x_j^2 - x_j - \lambda/\alpha(\alpha+1)\ \Delta x_j^{(2)} + \frac{\alpha(\alpha+1)}{2}\frac{\lambda^2}{\alpha^2}\Delta^2 x_j^{(2)}\} \quad ,$$

$$= m_2' - m_1' - \frac{2\lambda(\alpha+1)}{\alpha} m_1' + \alpha(\alpha+1)\lambda^2/\alpha^2 \quad ,$$

$$= m_2' - m_1' - 2m_1'\lambda(1+1/\alpha) + \lambda^2(1+1/\alpha) \quad .$$

It is important to have a direct method of relating m_r' to some function of the Q's, i.e., $f(Q_1, Q_2, \ldots, Q_r)$. An inverse moment relationship was introduced by Shenton and Wallington [24] of the type;

$$X^{(r)} = (1+\lambda/\alpha\ \Delta X)^{\alpha+r-1}\ q_r(X) \quad , \tag{4-3}$$

where $\quad \Delta q_r(X) = rq_{r-1}(X)$.

One can then use (4-3) and the basic definition of the Q's to express the sample factorial moments in terms of the Q's. For example, the following relationships are derived:

$$m_1 = Q_1 + \lambda \quad ,$$

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

$$m_{(3)} = Q_3 + 3\lambda(1+2/\alpha)Q_2 + 3\lambda^2(1+2/\alpha)(1+1/\alpha)Q_1 + \lambda^3(1+2/\alpha)(1+1/\alpha) \quad .$$

Here $m_{(k)}$ refers to the k-th factorial sample moment, i.e., $\sum_{j=1}^{n} x_j(x_j-1)\ldots.(x_j-k+1)/n$.

(b) <u>Some Properties of q's for Negative Binomial</u>

There are other useful properties of the orthogonal polynomials associated with the negative binomial distribution, some of which are applicable here. For example, Shenton [25] pointed out the useful expression that gives a general evaluation of φ_r, where $r = 1,2,\ldots.$ as follows:

$$Eq_r^2(x) = \lambda^r(\lambda+\alpha)^r(\alpha+r-1)^{(r)}r!/\alpha^{2r} \qquad (4\text{-}4)$$
$$= \varphi_r \quad ,$$

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

Another property, although not directly connected with the orthogonal statistics but one which can be extremely useful in evaluating the $[r^{\alpha}s^{\beta}t^{\gamma}....]$ in case they are needed, is the expression which enables one to obtain the factorial moment

$$Ex^{(s)} = \lambda^{s} \prod_{i=1}^{s-1} (1+i/\alpha) = \mu_{(s)} \quad . \tag{4-5}$$

As examples we have the following:

$$\mu_{1} = \lambda \quad .$$

$$\mu_{(2)} = \lambda^{2}(1+1/\alpha) \quad ,$$

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

It might be noted here that the factorial moments are much easier to work with than the central moments. The factorial moment generating is found in Table 1 of Chapter III. Note from the above factorial moments that if we consider the case where α is large, the moments approach those of a Poisson.

(c) Expected Q-Products for Negative Binomial

Recall in section (a) of Chapter II that certain recurrence relationships were given which eventually led to relatively easy attainment of the expected q-products which were extremely useful in finding the actual moments in powers of $1/n$ of moment statistics. However, the material of

Chapter II was discussed only with respect to general recurrence relationships and no specific cases were discussed. It is proposed that here those recurrence relationships be discussed which were useful in obtaining the moments of certain estimators of negative binomial parameters.

Using the same (λ, α) form of the distribution as mentioned in (a) of this chapter, we have the first basic recurrence relationship for $r \geq 2$,

$$q_r(x) = \{x-\lambda-(r-1)(2\lambda+\alpha)/\alpha\}q_{r-1}(x) - \frac{\lambda(\lambda+\alpha)(r-1)(\alpha+r-2)}{\alpha^2}q_{r-2}(x) \ . \tag{4-6}$$

This relationship can easily be derived using the general form of the recurrence relationship (2-5) in Chapter II, i.e.,

$$q_r(x) = (x-a_r)q_{r-1}(x) - b_r q_{r-2}(x) \quad . \tag{4-7}$$

Here we note that by expanding (4-1) for $k = (r-1)$ and $k = (r-2)$, one finds that from the notation of section (a) of Chapter II,

$$a_r = \lambda + (r-1)(2\lambda+\alpha)/\alpha \quad ,$$

$$b_r = \varphi_{r-1}/\varphi_{r-2} \quad ,$$

$$= \frac{\lambda^{r-1}(\lambda+\alpha)^{r-1}(\alpha+r-2)^{(r-1)}(r-1)!\ \alpha^{2r-4}}{\alpha^{2r-2}\lambda^{r-2}(\lambda+\alpha)^{r-2}(\alpha+r-3)^{(r-2)}(r-2)!} \quad ,$$

$$= \lambda(\lambda+\alpha)(r-1)(\alpha+r-2)/\alpha^2 \quad .$$

Thus, after substitution of a_r and b_r into (4-7), (4-6) results.

From (4-6) one can derive certain q-products for the negative binomial just as they were derived in general in Chapter II. For example, since $x - \lambda = q_1$, one can solve for the product $q_1(x)\cdot q_{r-1}(x)$ and obtain

$$q_1 q_{r-1} = q_r + (r-1)(1+2\lambda/\alpha)q_{r-1} + \frac{\lambda(\lambda+\alpha)(r-1)(\alpha+r-2)}{\alpha^2} q_{r-2} \quad . \tag{4-8}$$

(Note that the argument x is omitted here.)

Similarly one can manipulate (4-7) and (4-8) to obtain the following:

$$\begin{aligned} q_2 q_{r-1} = q_{r+1} &+ 2(r-1)(1+2\lambda/\alpha)q_r + (r-1)\{\frac{2\lambda(\lambda+\alpha)(\alpha+r-1)}{\alpha^2} \\ &+ (r-2)(1+2\lambda/\alpha)^2\}q_{r-1} \\ &+ \frac{2\lambda(\lambda+\alpha)}{\alpha^2}(1+2\lambda/\alpha)(\alpha+r-2)(r-1)(r-2)q_{r-2} \\ &+ \frac{\lambda^2(\lambda+\alpha)^2}{\alpha^4}(r-1)(r-2)(\alpha+r-2)(\alpha+r-3)q_{r-3} \quad . \end{aligned} \tag{4-9}$$

From (4-8) and (4-9) we can derive any q-product of order 2 that contains q_1 and q_2. This will enable us to obtain many expected q's of order 3. For example, suppose we needed $[1^2 2]$. Setting r = 3 in (4-8) gives:

$$q_1 q_2 = q_3 + 2(1+2\lambda/\alpha)q_2 + \frac{2\lambda(\lambda+\alpha)(\alpha+1)}{\alpha^2} q_1 \quad .$$

Multiplying both sides of the above equation by q_1 and taking expectations yields:

$$[1^2 2] = \frac{2\lambda(\lambda+\alpha)(\alpha+1)}{\alpha^2}\,\varphi_1 \quad .$$

Since $\varphi_1 = \lambda(\lambda+\alpha)/\alpha$, the final result becomes,

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

It might be noted here that the orthogonal property causes many of the expected triple-products to become zero. For example, if we multiply both sides of (4-9) by q_s where $s > r+1$ or $< r-3$ and take expectations, the result will be zero.

The expected q-products found via these recurrence formulae can be used to obtain the expected Q-products for the negative binomial. The actual expected Q-products involving Q_1 and Q_2 not exceeding order 8 through terms in n^{-4} were worked out in terms of the parameters λ and α and are given in Appendix B. This work was accomplished by the use of the expected Q's in general which are found in Appendix A and by the manipulation of the recurrence formulae discussed above.

CHAPTER V

MOMENT ESTIMATORS FOR THE NEGATIVE BINOMIAL

It can be seen by eq. (3.4) of Chapter III that the likelihood equation is a rather formidable one and in practice is not easy to solve unless a digital machine is used. Thus the simplest and most attractive method of estimation seems to be that of moments. The estimators for three of the forms of the distribution are given in Table 1 in Chapter III. Various workers have used these estimators in practice. Table 2 in Chapter III shows a list of workers who, in practical cases, used moment estimators for the three forms of the negative binomial that we consider here. Note the particular values of the parameters (obtained, of course, through estimation procedures).

The moment estimators derived by equating the first two sample moments to the corresponding population moments will be considered in detail here. Bias, variance, and covariance properties for these estimators will be discussed. Particular emphasis is put on a discussion of the possible misinterpretation or misuse of asymptotic properties of these estimators.

Similar properties in the case of the maximum likelihood estimators for one of the forms of the distribution were worked out by Bowman [7]. A discussion of the comparison of the behavior of these estimators is given in Chapter VI.

(a) Moment Estimators for Anscombe's Parameters

We shall denote the moment estimators for Anscombe's form as $\hat{\alpha}$ and $\hat{\lambda}$, where of course, $\hat{\lambda} = m_1'$ and $\hat{\alpha} = m_1'^2/(m_2-m_1')$. Here $\hat{\alpha}$ is not defined for a null sample, i.e., when $x_j = 0$ for $j = 1,2,\ldots,n$. This does not present a problem since there is a small probability of a null sample occurring if n is large and for certain regions of the parameter space. When the maximum likelihood estimators are used in the discussion, they will be referred to as α^* and λ^*.

1. The Bias of $\hat{\alpha}$

An asymptotic series was developed for the bias of $\hat{\alpha}$ through terms in n^{-4}. The expression is as follows:

$$E(\hat{\alpha}-\alpha) = \Lambda_1^{(1)}/n + \Lambda_2^{(1)}/n_2 + \Lambda_3^{(1)}/n_3 + \Lambda_4^{(1)}/n_4 + \cdots, \quad (5-1)$$

where: $\Lambda_1^{(1)} = (\lambda+\alpha)(\alpha+1)(2\alpha+3\lambda)/\lambda^2$, (5-1-a)

$$\Lambda_2^{(1)} = (\lambda+\alpha)(\alpha+1)\{(9\alpha-28)\lambda^3 + (31\alpha-24)\lambda^2\alpha + (34\alpha+10)\lambda\alpha^2 + 4\alpha^3(3\alpha+2)\}/\alpha\lambda^4 \quad , \qquad (5\text{-}1\text{-}b)$$

$$\Lambda_3^{(1)} = (\lambda+\alpha)(\alpha+1) \sum_{i=0}^{5} A_i^{(1)}\lambda^i\alpha^{5-i}/\alpha^2\lambda^6 \quad . \qquad (5\text{-}1\text{-}c)$$

Here the $A_i^{(1)}$'s are given as follows:

$$A_5^{(1)} = 27\alpha^2 + 24\alpha + 1128 \quad ,$$

$$A_4^{(1)} = 218\alpha^2 - 268\alpha + 2352 \quad ,$$

$$A_3^{(1)} = 615\alpha^2 - 433\alpha + 1524 \quad ,$$

$$A_2^{(1)} = 804\alpha^2 + 154\alpha + 368 \quad ,$$

$$A_1^{(1)} = 500\alpha^2 + 456\alpha + 120 \quad ,$$

$$A_0^{(1)} = 120\alpha^2 + 160\alpha + 48 \quad .$$

$$\Lambda_4^{(1)} = (\lambda+\alpha)(\alpha+1) \sum_{i=0}^{7} B_i^{(1)}\lambda^i\alpha^{7-i}/\alpha^3\lambda^8 \quad , \qquad (5\text{-}1\text{-}d)$$

with $B_i^{(1)} = \sum_{j=0}^{3} b_j^{(1)}\alpha^j$ for all i .

The $b_j^{(1)}$'s are given in the body of the following table for each $B_i^{(1)}$.

$B_i^{(1)}$ \ α^j	α^3	α^2	α	Const.
$B_7^{(1)}$	81	-276	-22,400	-86,064
$B_6^{(1)}$	1,279	-1,388	-46,104	-268,128
$B_5^{(1)}$	6,811	-6,157	-9,078	-308,088
$B_4^{(1)}$	17,769	-7,997	36,556	-155,360
$B_3^{(1)}$	25,696	6,008	31,050	-29,232
$B_2^{(1)}$	21,100	20,040	14,804	1,184
$B_1^{(1)}$	9,240	14,440	7,760	1,456
$B_0^{(1)}$	1,680	3,360	2,080	384

The expression described above was derived using the orthogonal Q's by the technique described in Chapters II and IV.

The first four terms in the bias of $\hat{\alpha}$ given in (5-1) are tabulated for n = 100 in Table A.1. Note that in general the entries decrease as λ increases for a constant α and increase as α increases with a constant λ. The fact that the second, third, and fourth terms seem to be quite large

Table A.1. Bias of $\hat{\alpha} = m_1^2/(m_2-m_1)$

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.2000	.1200	.0978	.0875	.0816	.0704	.0668	.0641	.0620	.0610
	(II)	.0208	.0006	.0029	.0035	.0038	.0039	.0039	.0039	.0038	.0038
	(III)	.0317	.0090	.0059	.0047	.0041	.0031	.0029	.0026	.0025	.0024
	(IV)	.0287	.0106	.0067	.0052	.0044	.0032	.0028	.0025	.0023	.0023
2	(I)	.6300	.3000	.2167	.1800	.1596	.1224	.1111	.1024	.0961	.0930
	(II)	.2853	.0414	.0149	.0073	.0041	.0002	.0006	.0010	.0013	.0014
	(III)	.2833	.0223	.0079	.0046	.0034	.0018	.0015	.0012	.0011	.0010
	(IV)	.2799	.0002	.0024	.0020	.0016	.0010	.0008	.0007	.0006	.0005
3	(I)	1.440	.6000	.4000	.3150	.2688	.1872	.1632	.1452	.1323	.1261
	(II)	1.281	.1677	.0597	.0309	.0193	.0054	.0028	.0013	.0005	.0002
	(III)	2.058	.0994	.0234	.0101	.0059	.0019	.0013	.0010	.0008	.0007
	(IV)	4.386	.0600	.0051	.0003	.0005	.0005	.0004	.0003	.0003	.0003
4	(I)	2.750	1.050	.6611	.5000	.4140	.2660	.2238	.1926	.1706	.1602
	(II)	3.955	.4650	.1556	.0775	.0473	.0131	.0073	.0041	.0023	.0016
	(III)	9.808	.3779	.0732	.0265	.0133	.0027	.0016	.0010	.0007	.0006
	(IV)	33.56	.3944	.0380	.0078	.0022	.0002	.0002	.0002	.0002	.0002
5	(I)	4.680	1.680	1.013	.7425	.6000	.3600	.2933	.2448	.2112	.1953
	(II)	9.844	1.063	.3361	.1606	.0950	.0245	.0133	.0073	.0042	.0030
	(III)	35.13	1.176	.2025	.0662	.0303	.0045	.0021	.0012	.0007	.0006
	(IV)	174.6	1.763	.1563	.0320	.0100	.0002	.0001	.0001	.0001	.0001
10	(I)	27.83	8.580	4.608	3.080	2.310	1.100	.7944	.5852	.4488	.3872
	(II)	200.8	17.32	4.597	1.911	1.009	.1817	.0818	.0371	.0177	.0114
	(III)	2420	58.54	7.695	1.994	.7431	.0517	.0149	.0044	.0015	.0008
	(IV)	40830	276.5	17.94	2.882	.7524	.0189	.0031	.0004	.0000	.0002
15	(I)	84.48	24.48	12.48	7.986	5.760	2.400	1.600	1.075	.7488	.6072
	(II)	1294	101.5	24.84	9.620	4.773	.6861	.2680	.1023	.0404	.0228
	(III)	33063	702.4	82.44	19.32	6.582	.3248	.0741	.0161	.0037	.0015
	(IV)	1183^{3}	6801	382.5	54.19	12.65	.2108	.0273	.0031	.0003	.0001
25	(I)	358.3	98.28	47.72	29.22	20.28	7.280	4.391	2.600	1.560	1.138
	(II)	14518	1047	237.2	85.71	39.91	4.463	1.460	.4401	.1289	.0576
	(III)	9807^{2}	18580	1964	418.5	130.5	4.512	.7924	.1194	.0165	.0044
	(IV)	9274^{4}	4617^{2}	22767	2858	597.0	6.354	.5957	.0441	.0027	.0004
50	(I)	2679	702.8	327.4	192.8	129.0	39.78	21.36	10.71	5.100	3.060
	(II)	418^{3}	28130	5975	2030	891.6	77.60	20.79	4.644	.8703	.2583
	(III)	1087^{5}	1877^{3}	1817^{2}	35601	10257	251.1	33.36	3.279	.2334	.0324
	(IV)	3957^{7}	1753^{5}	7735^{3}	8740^{2}	1652^{2}	1136	74.78	3.221	.0859	.0054
100	(I)	20708	5305	2416	1392	912.0	255.5	126.5	55.55	21.21	10.10
	(II)	1267^{4}	8225^{2}	1686^{2}	5537^{1}	23523	1756	411.1	73.25	9.189	1.730
	(III)	1293^{7}	2125^{6}	1962^{4}	3670^{3}	1011^{3}	20081	2219	159.5	6.457	.4617
	(IV)	1846^{10}	7687^{7}	3195^{7}	3406^{5}	6080^{4}	3214^{2}	1676^{1}	485.7	6.317	.1696

Key to Table A.1. (a) (I), (II), (III), (IV) refer to the four terms in bias of $\hat{\alpha}$ when n = 100.

(b) Underlined figures in this table are negative.

(c) Indices in this table are to be taken as the power of 10 which multiplies the entry; thus $64^{1} = 640$

for many values of α and λ certainly generates much thought on the idea of the interpretation of what one might call "asymptotic bias". The following paragraphs seek to answer these questions.

An observation of Table A.1 will reveal that for the part of the parameter space in which $\lambda > \alpha$, it may be said that if one quotes only the first term in the bias one won't be in error by more than 15% for $n = 100$. Thus it certainly can be said that a "good" approximation of the bias is given by (5-1) for $n = 100$ and $\lambda > \alpha$. Consider, however, the case in which $\alpha > \lambda$. In some sections of this region of the parameter space not only is it "unsafe" to consider only the asymptotic bias, that is the first term of the expansion, but in some cases the next three terms completely "overwhelm" the first. For example, with a sample size of 1000 one might expect the higher order terms in the expansion to be insignificant. However, for a case in which $\lambda = 2$, $\alpha = 50$, the terms in the expansion become 702.8 + 28,130. + 1,877,000. + 175,300,000. Note the astronomical error that would be involved by using only asymptotic bias. It turns out that even in the case where $\lambda = 1$, $\alpha = 5$, and

$n = 1000$, the error in using the first term would be as high as 25%.

To further illustrate the chaotic nature of the bias of $\hat{\alpha}$ and also to point out possible pitfalls through not including higher order terms in asymptotic expansions in general, Figure 1 is given which shows the ratio, as a percentage, of the first term in the bias to the first four terms when $n = 100$. Note the relatively large contribution of the first term when $\lambda > \alpha$. Note also the small contribution when $\alpha > \lambda$. Figure 2 gives the minimum sample size that can be used in order to make the first four terms in (5-1) a "good" approximation (The criterion of "good" here was selecting an n for which $\Lambda_4^{(1)}/\Lambda_1^{(1)}n^3 < 1/20.$). Note the relatively large sample size needed in the regions of low λ and high α.

Two points need to be established at this stage as far as the bias of $\hat{\alpha}$ is concerned. First, it certainly can be said that in regions of low λ and high α, the estimator $\hat{\alpha}$ appears to have a relatively large bias (assuming of course that n is not extremely large). Secondly, it can be pointed out that to obtain an approximation of the bias one must use some caution in practical cases and be wary of using only

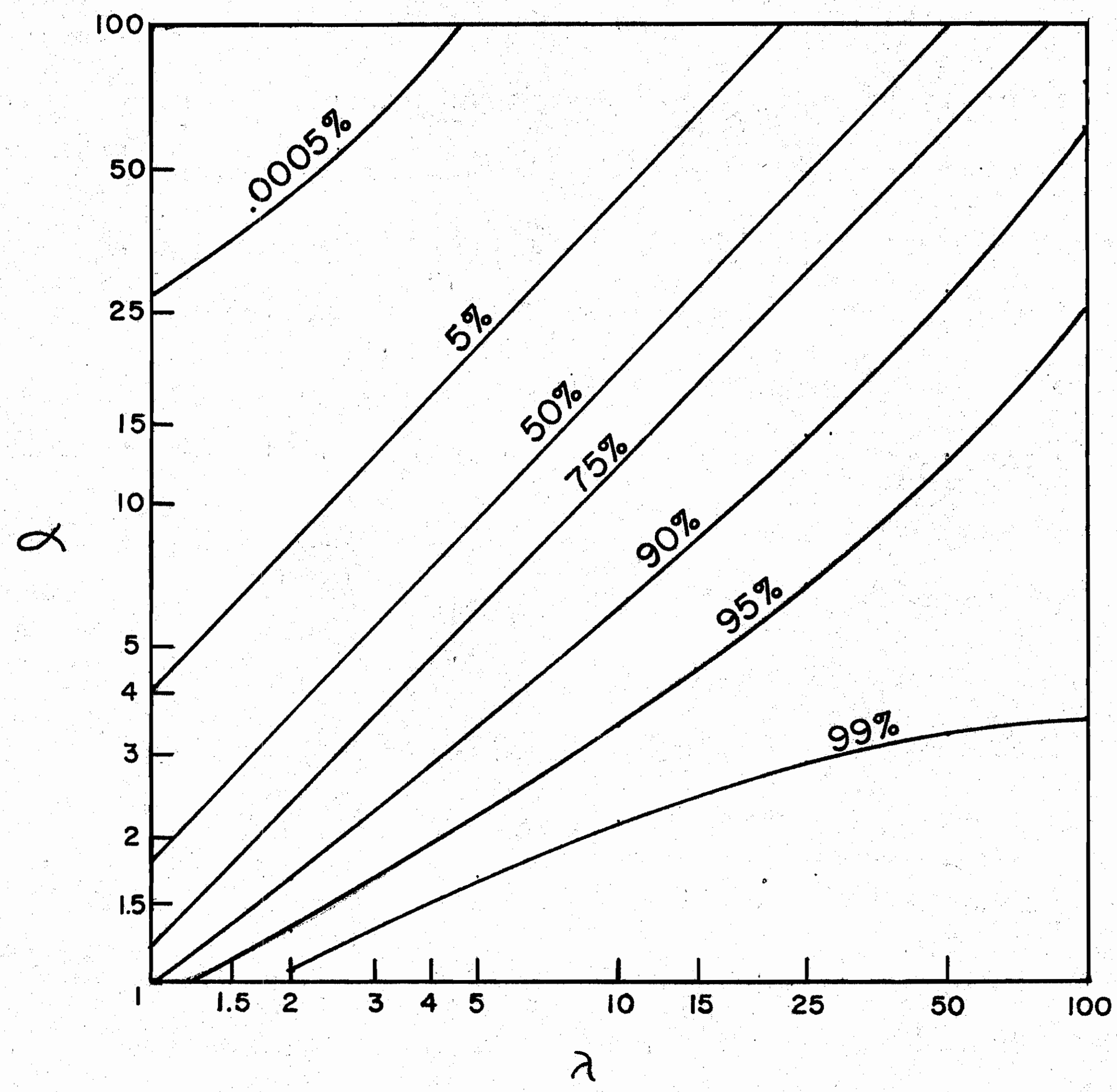

Figure 1. Bias of $\hat{\alpha} = m_1^2/(m_2-m_1)$.

Ratio of first to first four terms in asymptotic expansion of $E(\hat{\alpha}-\alpha)$ when sample size $n = 100$ (expressed as a percentage).

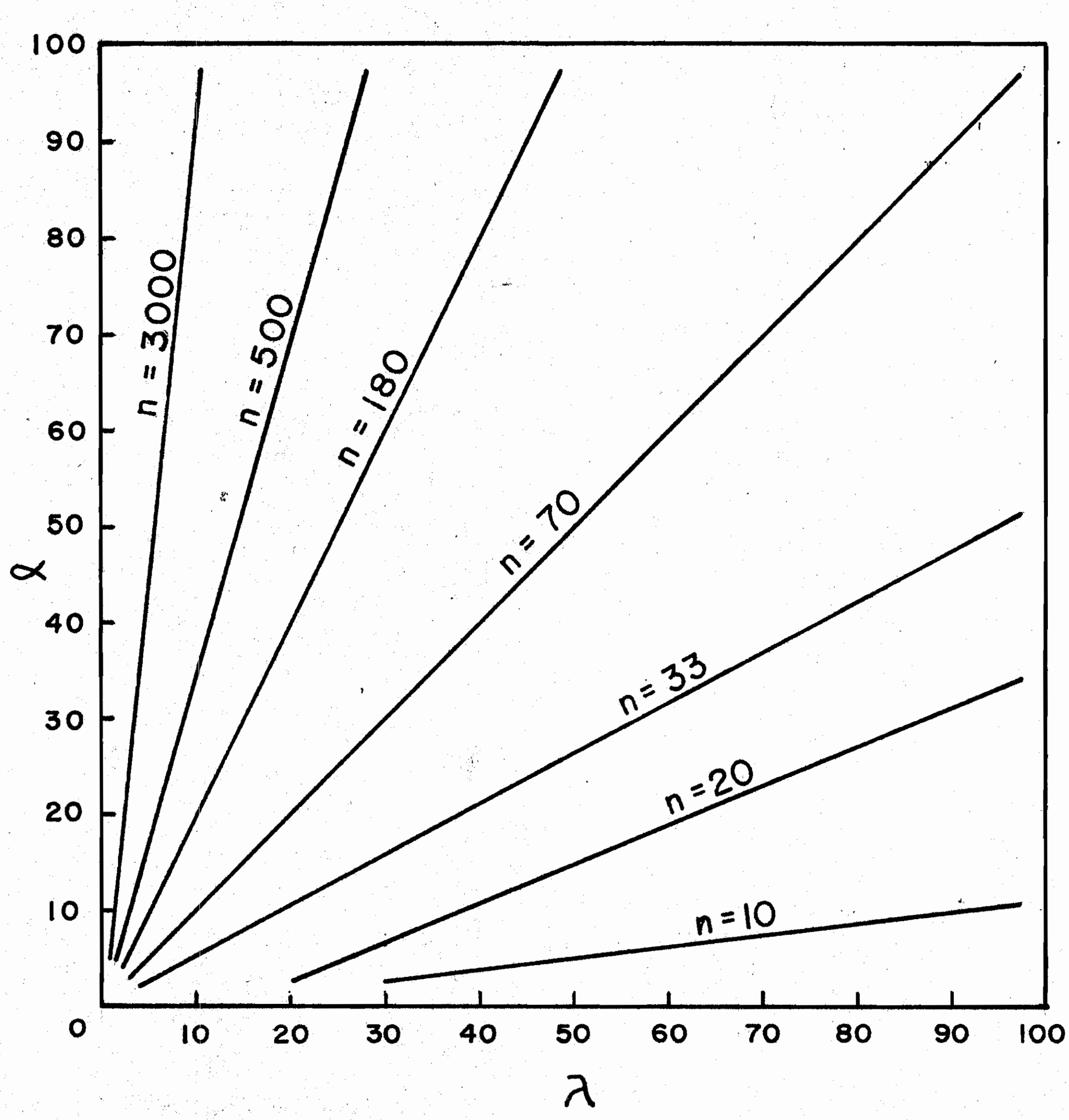

Figure 2. Bias of $\hat{\alpha} = m_1^2/(m_2-m_1)$.

Sample size n in asymptotic expansion of $E(\hat{\alpha}-\alpha)$ to make the ratio (Fourth term)/(First term) = 0.05 approx.

the first term in the expansion of the bias in cases where $\alpha > \lambda$. Certainly the above results point out that in many cases even four terms in the expansion of the bias will not be sufficient to obtain a good approximation.

2. The Variance of $\hat{\alpha}$

An asymptotic expansion was found of the type:

$$\mathrm{Var}(\hat{\alpha}) = \Lambda_1^{(2)}/n + \Lambda_2^{(2)}/n^2 + \Lambda_3^{(2)}/n^3 + \Lambda_4^{(2)}/n^4 + \ldots \quad , \tag{5-6}$$

using orthogonal statistics. Here the $\Lambda^{(2)}$'s are as follows:

$$\Lambda_1^{(2)} = 2\alpha(\alpha+1)(\lambda+\alpha)^2/\lambda^2 \quad , \tag{5-6-a}$$

$$\Lambda_2^{(2)} = (\alpha+1)(\lambda+\alpha)\{(22\alpha-38)\lambda^3 + (78\alpha-26)\lambda^2\alpha + (88\alpha+36)\lambda\alpha^2 + (32\alpha+24)\alpha^3\}/\lambda^4 \quad , \tag{5-6-b}$$

$$\Lambda_3^{(2)} = (\alpha+1)(\lambda+\alpha) \sum_{i=0}^{5} A_i^{(2)}\lambda^i\alpha^{5-i}/\alpha\lambda^6 \quad , \tag{5-6-c}$$

where:

$$A_5^{(2)} = 164\alpha^2 - 194\alpha + 2572 \quad ,$$

$$A_4^{(2)} = 1172\alpha^2 - 1302\alpha + 5332 \quad ,$$

$$A_3^{(2)} = 3092\alpha^2 - 1083\alpha + 3327 \quad ,$$

$$A_2^{(2)} = 3876\alpha^2 + 1696\alpha + 950 \quad ,$$

$$A_1^{(2)} = 2344\alpha^2 + 2472\alpha + 656 \quad ,$$

$$A_0^{(2)} = 552\alpha^2 + 800\alpha + 272 \quad .$$

$$\Lambda_4^{(2)} = (\alpha+1)(\lambda+\alpha) \sum_{i=0}^{7} B_i^{(2)} \lambda^i \alpha^{7-i} / \alpha^2 \lambda^8 \quad ,$$

where $B_i^{(2)} = \sum_{j=0}^{3} b_j^{(2)} \alpha^j$, the coefficients $b_j^{(2)}$ being given in the body of the table below for each $B_i^{(2)}$:

$B_i^{(2)}$ \ α^j	α^3	α^2	α	Const.
$B_7^{(2)}$	1,036	-1,516	-53,052	-279,160
$B_6^{(2)}$	12,684	-15,332	-69,428	-870,856
$B_5^{(2)}$	58,820	-48,224	95,268	-988,486
$B_4^{(2)}$	140,348	-29,360	211,992	-476,460
$B_3^{(2)}$	190,560	92,840	154,250	-72,314
$B_2^{(2)}$	149,288	172,208	99,416	15,840
$B_1^{(2)}$	63,040	108,160	61,536	11,840
$B_0^{(2)}$	11,136	23,744	16,032	3,360

Table A.2 shows a tabulation of the four terms in (5-6). As in the case of the bias, charts were prepared illustrating the behavior of this property. For instance, Figures 3 and 4 show the relative importance of the first term to the first four for samples of size 100 and 250 respectively. Figure 5 shows a chart of minimum sample size in order that $\Lambda_4^{(2)}/(\Lambda_1^{(2)}n^3) < 1/20$.

We can easily see from Table A.2 that not unlike the case of the bias described in A.1, the asymptotic behavior becomes more explosive as α exceeds λ more and more. Figures 3 and 4 illustrate perfectly the trap into which one can easily fall in judging an estimator on variance properties when only the asymptotic variance has been considered. We note that even when $\lambda = 1$, $\alpha = 2$, and $n = 100$ the first term is approximately 10% of the sum of the first four. When $\lambda = 1$, $\alpha = 3$, and $n = 100$, i.e., if we increase α by one unit, this reduces the contribution of the first term to 2%. Thus it can be seen that in practical cases (disregarding extremely large sample sizes) one's assessment of the variance through use of the asymptotic term can be a gross <u>underestimate</u> of the true variance.

It should be noted that in Figures 3 and 4, superimposed on the contours of % contribution of the first term to the first four in Var($\hat{\alpha}$) are contours of constant asymptotic efficiency. The purpose for this comparison was to better point out the trap involved in using asymptotic properties

Table A.2. Terms in Var $\hat{\alpha}$ n = 100

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.1600	.0900	.0711	.0625	.0576	.0484	.0455	.0433	.0416	.0410
	(II)	.0864	.0144	.0046	.0014	$\underline{.0000}$	$\underline{.0021}$	$\underline{.0026}$	$\underline{.0029}$	$\underline{.0030}$	$\underline{.0031}$
	(III)	.1068	.0231	.0138	.0107	.0093	.0069	.0062	.0057	.0054	.0052
	(IV)	$\underline{.0484}$	$\underline{.0301}$	$\underline{.0198}$	$\underline{.0156}$	$\underline{.0134}$	$\underline{.0096}$	$\underline{.0085}$	$\underline{.0077}$	$\underline{.0072}$	$\underline{.0069}$
2	(I)	1.080	.4800	.3333	.2700	.2352	.1728	.1541	.1399	.1298	.1248
	(II)	1.636	.2616	.1065	.0608	.0409	.0148	.0093	.0058	.0036	.0027
	(III)	2.703	.1881	.0579	.0303	.0203	.0091	.0070	.0057	.0049	.0046
	(IV)	4.526	.0690	$\underline{.0022}$	$\underline{.0074}$	$\underline{.0075}$	$\underline{.0053}$	$\underline{.0044}$	$\underline{.0038}$	$\underline{.0033}$	$\underline{.0031}$
3	(I)	3.840	1.500	.9600	.7350	.6144	.4056	.3456	.3011	.2696	.2546
	(II)	10.55	1.420	.5248	.2824	.1835	.0627	.0396	.0258	.0176	.0142
	(III)	29.94	1.427	.3229	.1317	.0723	.0189	.0118	.0080	.0061	.0053
	(IV)	95.65	1.457	.1648	.0368	.0103	$\underline{.0028}$	$\underline{.0029}$	$\underline{.0026}$	$\underline{.0022}$	$\underline{.0021}$
4	(I)	10.00	3.600	2.178	1.600	1.296	.7840	.6418	.5382	.4666	.4326
	(II)	42.83	5.085	1.728	.8760	.5445	.1653	.0993	.0617	.0407	.0323
	(III)	188.9	7.337	1.414	.5047	.2474	.0447	.0229	.0131	.0085	.0068
	(IV)	952.4	11.70	1.205	.2773	.0955	.0035	$\underline{.0007}$	$\underline{.0016}$	$\underline{.0016}$	$\underline{.0015}$
5	(I)	21.60	7.350	4.267	3.038	2.400	1.350	1.067	.8640	.7260	.6615
	(II)	132.5	14.35	4.564	2.201	1.315	.3564	.2024	.1190	.0748	.0576
	(III)	839.2	28.43	4.924	1.609	.7321	.1024	.0453	.0220	.0124	.0092
	(IV)	6113	63.61	5.839	1.247	.4121	.0206	.0040	$\underline{.0001}$	$\underline{.0010}$	$\underline{.0011}$
10	(I)	266.2	79.20	41.31	26.95	19.80	8.800	6.111	4.312	3.168	2.662
	(II)	5364	461.0	122.1	50.69	26.75	4.830	2.188	1.004	.4913	.3229
	(III)	1135^{2}	2769	366.8	95.71	35.89	2.545	.7406	.2192	.0722	.0380
	(IV)	2785^{3}	1914^{1}	1260	205.2	54.27	1.450	.2582	.0441	.0075	.0021
15	(I)	1229	346.8	172.8	108.3	76.80	30.00	19.20	12.29	8.112	6.348
	(II)	5180^{1}	4050	988.0	381.7	189.0	27.04	10.55	4.033	1.603	.9123
	(III)	2311^{3}	4938^{1}	5826	1373	469.9	23.67	5.480	1.212	.2802	.1137
	(IV)	1199^{5}	6964^{2}	3956^{1}	5658	1333	23.15	3.113	.3808	.0461	.0114
25	(I)	8788	2369	1132	683.3	468.0	159.3	92.44	52.00	29.25	20.31
	(II)	9681^{2}	6961^{1}	1574^{1}	5675	2637	292.8	95.30	28.58	8.330	3.273
	(III)	1137^{5}	2161^{3}	2292^{2}	4897^{1}	1532^{1}	537.5	95.62	14.71	2.009	.5729
	(IV)	1555^{7}	7789^{4}	3865^{3}	4881^{2}	1025^{2}	1121	107.7	8.316	.5528	.0855
50	(I)	1327^{2}	3448^{1}	1592^{1}	9295	6171	1836	957.7	459.0	204.0	114.8
	(II)	5574^{4}	3745^{3}	7944^{2}	2695^{2}	1182^{2}	1023^{1}	2727	605.1	112.2	32.98
	(III)	2510^{7}	4340^{5}	4209^{4}	8260^{3}	2384^{3}	5880^{1}	7870	784.2	57.39	8.259
	(IV)	1319^{10}	5861^{7}	2595^{6}	2941^{5}	5575^{4}	3892^{2}	2598^{1}	1148	32.29	2.181
100	(I)	2061^{3}	5254^{2}	2381^{2}	1366^{2}	8908^{1}	2444^{1}	1187^{1}	5050	1818	808.0
	(II)	3379^{6}	2192^{5}	4490^{4}	1473^{4}	6254^{3}	4652^{2}	1086^{2}	1924^{1}	2389	444.3
	(III)	5958^{9}	9803^{7}	9057^{6}	1696^{6}	4674^{5}	9322^{3}	1034^{3}	7490^{1}	3084	226.7
	(IV)	1227^{13}	5119^{10}	2131^{9}	2275^{8}	4068^{7}	2167^{5}	1138^{4}	3346^{2}	4496	127.3

Key to Table A.2. (a) (I), (II), (III), (IV) refer to terms in Var $\hat{\alpha}$ when n = 100.

(b) Underlined entries are negative.

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

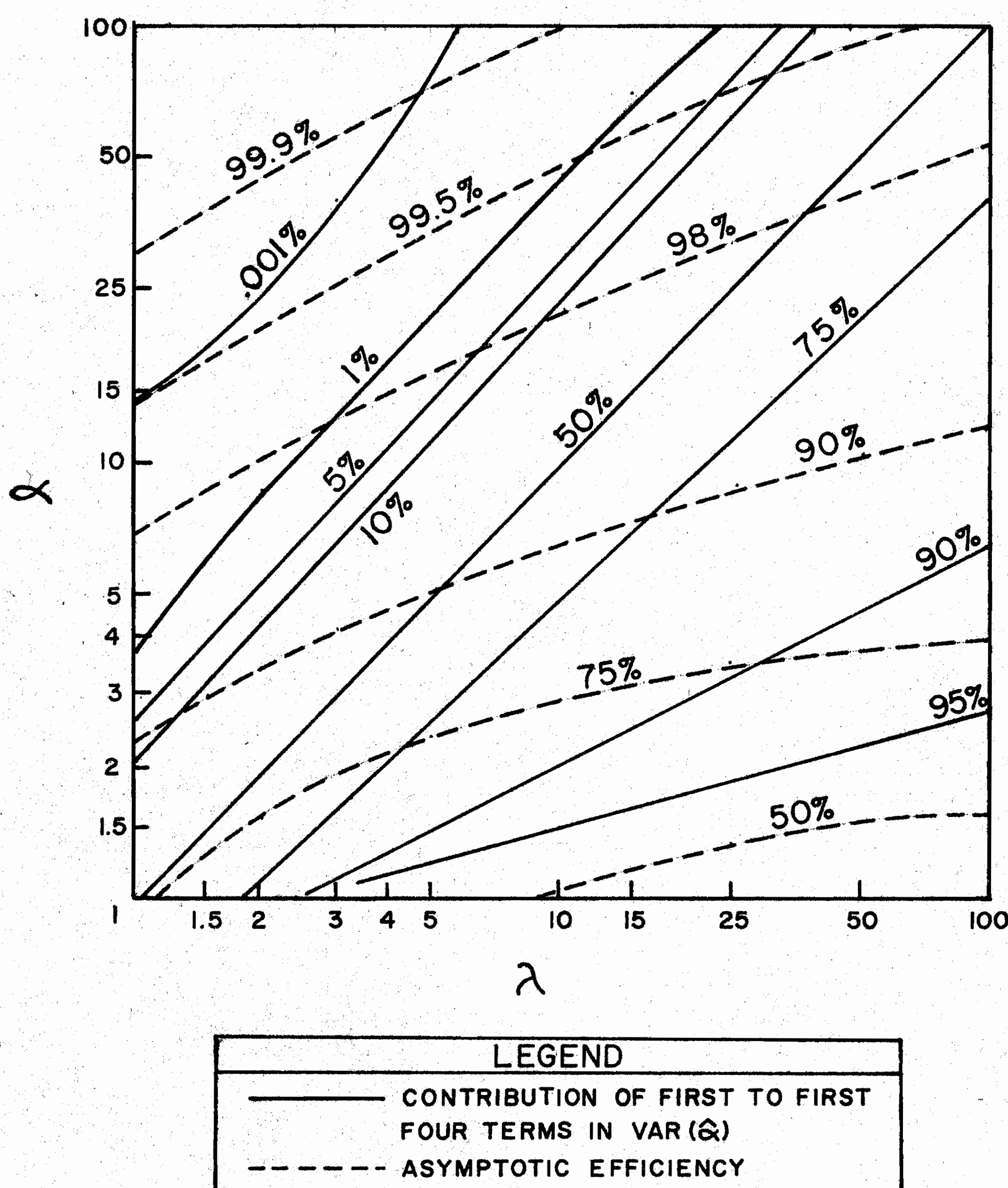

Figure 3. Variance of $\hat{\alpha} = m_1^2/(m_2-m_1)$.

(a) Ratio (First term)/(Sum of first four terms), as a percentage, in asymptotic expansion of Var $\hat{\alpha}$ when sample size n = 100.

(b) Asymptotic Efficiency of $\hat{\alpha}$ using n^{-1} terms only.

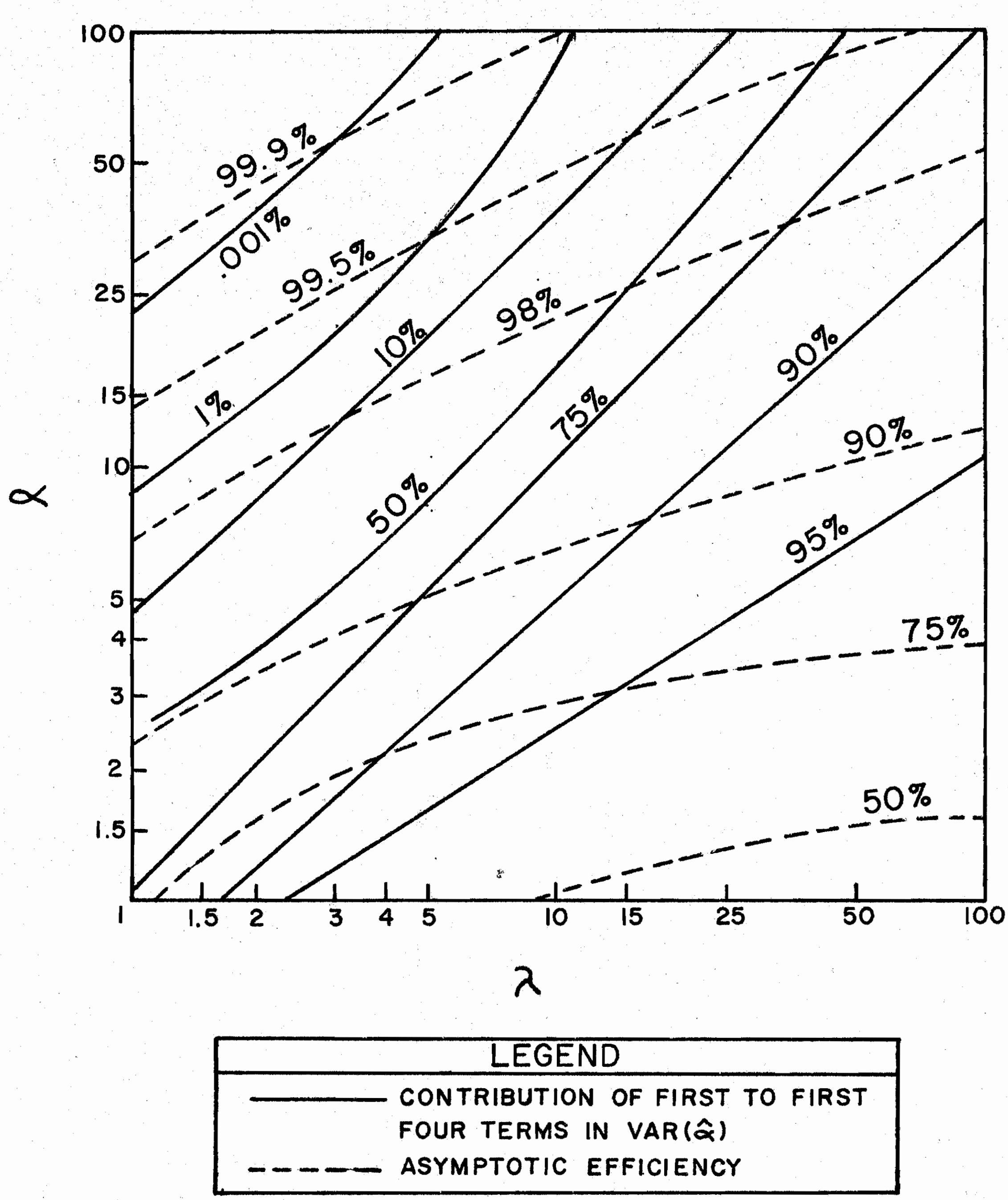

Figure 4. Variance of $\hat{\alpha} = m_1^2/(m_2-m_1)$.

(a) Ratio (First term)/(Sum of first four terms), as a percentage, in asymptotic expansion of Var $\hat{\alpha}$ when sample size n = 250.

(b) Asymptotic Efficiency of $\hat{\alpha}$ using n^{-1} terms only.

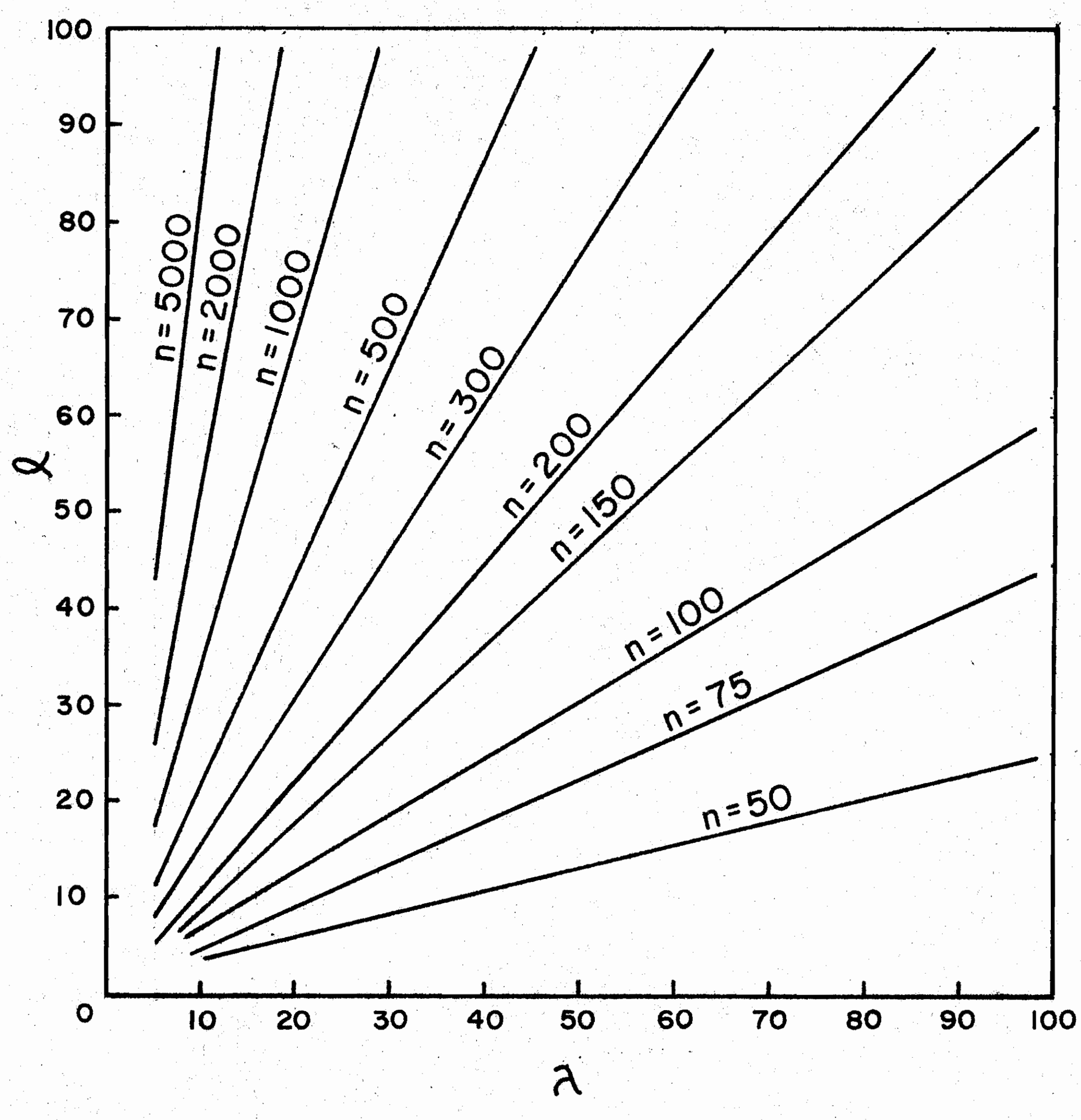

Figure 5. Variance of $\hat{\alpha} = m_1^2/(m_2-m_1)$.

Least sample size n required to make the ratio (Fourth term)/(First term) = 0.05 approx. in asymptotic expansion of Var $\hat{\alpha}$.

of estimators in general and $\hat{\alpha}$ in particular. Note that the very (λ,α) region in which one might expect in practice to be able to use the estimator, i.e., the area of high asymptotic efficiency, is the region where the first term of Var $\hat{\alpha}$ is least effective. Likewise, in the area where the behavior of $\hat{\alpha}$ is relatively mild the asymptotic efficiency is low. In other words one might easily be mislead into concluding that high asymptotic efficiency necessarily means that the variance of the estimator in question (in particular, $\hat{\alpha}$) does not exceed, to a great degree, that of the maximum likelihood estimator, and hence the former is a satisfactory estimator. However, we have here an example for which this is a complete paradox, since the first term in the variance represents only a small fraction of the true variance (except, of course, for extremely large sample sizes). Let us note in Figure 5 the large sample sizes that are needed in order that the expansion through four terms represents a good approximation of the true variance. Considering this situation, one might expect that astronomical sample sizes would be needed <u>in order that the asymptotic variance be allowed to be used when $\alpha \gg \lambda$</u>.

It can be said that in much of the work with the properties of estimators, the word "asymptotic" is used to "camouflage" the embarrassment of the ignorance about the succeeding terms in the expansion. It is assumed, of course, that there is some n which will make terms negligible beyond the first. However, this n cannot be known unless further terms are investigated. Even when further terms are known we can still only make a perspicacious guess at the required value of n. Thus it is impossible to be certain that one's assessment of say, the variance of an estimator is close to the true value unless he has investigated further terms. This case of $\hat{\alpha}$ is a perfect example in which the above is true.

3. Covariance Determinant of $(\hat{\alpha},\hat{\lambda})$

For Anscombe's form the moment estimators are asymptotically uncorrelated. For this reason, the asymptotic efficiency of $\hat{\alpha}$ was discussed by Anscombe [1] without reference to joint estimation. If joint estimation were to be considered, then the true efficiency of the estimators $(\hat{\alpha},\hat{\lambda})$ would be given as follows:

$$E = \begin{vmatrix} \text{Var } \alpha^*, & \text{Cov}(\alpha^*,\lambda^*) \\ \text{Cov}(\alpha^*,\lambda^*), & \text{Var } \lambda^* \end{vmatrix} \Bigg/ \begin{vmatrix} \text{Var } \hat{\alpha}, & \text{Cov}(\hat{\alpha},\hat{\lambda}) \\ \text{Cov}(\hat{\alpha},\hat{\lambda}), & \text{Var } \hat{\lambda} \end{vmatrix} , \quad (5\text{-}7)$$

where, of course, α^* and λ^* are the maximum likelihood estimators. As for the case of the bias and variance it was considered worthwhile to investigate further terms in the expansion of the covariance and hence the covariance determinant for $\hat{\alpha}$ and $\hat{\lambda}$. A similar study was conducted by Bowman [7] for α^* and λ^* and the actual discussion of a comparison of these results is given in Chapter VI.

Denote by D_A the denominator of (5-7), i.e., the determinant of the covariance matrix of the estimators $\hat{\alpha}$ and $\hat{\lambda}$. The purpose of the subscript notation here is to differentiate between the three forms discussed in the sequel, i.e., Anscombe, Evans, and Fisher (for example, the corresponding covariance determinant in Evans' notation will be D_E). Orthogonal statistics were used to obtain the following expansion for D_A:

$$D_A = \frac{2(\alpha+1)(\lambda+\alpha)^3}{\lambda n^2} + \frac{(\alpha+1)(\lambda+\alpha)^2}{n^3\lambda^3\alpha}[(22\alpha-38)\lambda^3 + (78\alpha-26)\lambda^2\alpha + (88\alpha+36)\lambda\alpha^2 + (32\alpha+24)\alpha^3] + \ldots \quad . \qquad (5\text{-}8)$$

Table A.3 displays a tabulation of the first and second term in the expansion of the covariance determinant for a sample of size 100. Note that, as in the case of the variance, as α becomes larger than λ the covariance determinant becomes larger and the n^{-3} term becomes more and more important. Of course the practical use of asymptotic efficiency assumes that the latter term is negligible. Figure 6 is a chart of the sample size for different values of λ and α that is needed in order that (5-8) is a good approximation to the covariance determinant. (A criterion was selected here which remained in keeping with that mentioned before for the bias and the variance, i.e., n was selected such that the ratio of the second term to the first was $\sqrt[3]{1/20}$.)

4. <u>Discussion of the "Distribution" of $\hat{\alpha}$</u>

Further work was conducted with the object of attempting to verify the "chaotic" nature of the distribution of $\hat{\alpha}$ or at least to lead to further implications of this. The coefficients of skewness $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$ were approximated for various values of α and λ. These were calculated from the moment expansions through terms in n^{-4}. Here orthogonal statistics were used to obtain $\mu_3(\hat{\alpha})$ and $\mu_4(\hat{\alpha})$. The actual expressions for these moments are given in Appendix C.

Table A.3. First and Second Terms in Covariance Determinant for Estimators of Anscombe's Form for n = 100

α \ λ	1	2	3	4	5	10	15	25	50	100
1	0.00320	0.00540	0.00853	0.01250	0.01728	0.05324	0.10920	0.28120	1.06120	4.12120
	0.00173	0.00086	0.00054	0.00028	$-0.0^{4}14$*	-0.0023	-0.0061	-0.0185	-0.0776	-0.3156
2	0.01620	0.01920	0.02500	0.03240	0.04116	0.10370	0.19650	0.47240	1.68730	6.36720
	0.02450	0.01046	0.00799	0.00729	0.00717	0.00890	0.01184	0.01947	0.04683	0.13555
3	0.05120	0.05000	0.05760	0.06860	0.08192	0.17580	0.31104	0.70250	2.38200	8.74180
	0.14063	0.04733	0.03149	0.02635	0.02447	0.02718	0.03567	0.06010	0.15575	0.48820
4	0.12500	0.10800	0.11433	0.12800	0.14580	0.27440	0.45730	0.97560	3.14930	11.2486
	0.53530	0.15255	0.09074	0.07008	0.06125	0.05787	0.07073	0.11176	0.27498	0.83920
5	0.25920	0.20580	0.20480	0.21870	0.24000	0.40500	0.64000	1.29600	3.99300	13.8915
	1.58940	0.40172	0.21909	0.15845	0.13152	0.10692	0.12146	0.17854	0.41144	1.20940
10	2.92820	1.90080	1.61110	1.50920	1.48500	1.76000	2.29160	3.77300	9.50400	29.2820
	59.0060	11.0650	4.76160	2.83880	2.00624	0.96624	0.82067	0.87881	1.47390	3.55250
15	13.1072	7.86080	6.22080	5.48720	5.12000	5.00000	5.76000	8.19200	17.5760	48.6680
	552.565	91.7930	35.5664	19.3415	12.6020	4.50670	3.16410	2.68870	3.47400	6.99420
25	91.3952	51.1760	38.0501	31.7057	28.0800	22.2950	22.1867	26.0000	43.8750	101.563
	10068.3	1503.67	528.904	263.313	158.214	40.9871	22.8710	14.2896	12.4956	18.6164
50	1353.04	717.100	506.182	401.533	339.405	220.320	186.745	172.125	204.000	344.250
	5.6850^{5}	77903.1	25260.5	11642.9	6502.39	1227.67	531.846	226.895	112.159	98.9375
100	20812.8	10718.2	7357.69	5680.56	467618.	2688.62	2048.11	1578.13	1363.50	1616.00
	3.4135^{7}	4.4709^{6}	1.3875^{6}	612779.	328312.	51175.4	18726.3	6012.18	1791.77	888.638

*Refers to 4 zeros between the decimal and first digit.

Other indices refer to the power of 10 to which the number in question is multiplied by, e.g., $3.4135^{7} = 3.4135 \times 10^{7}$.

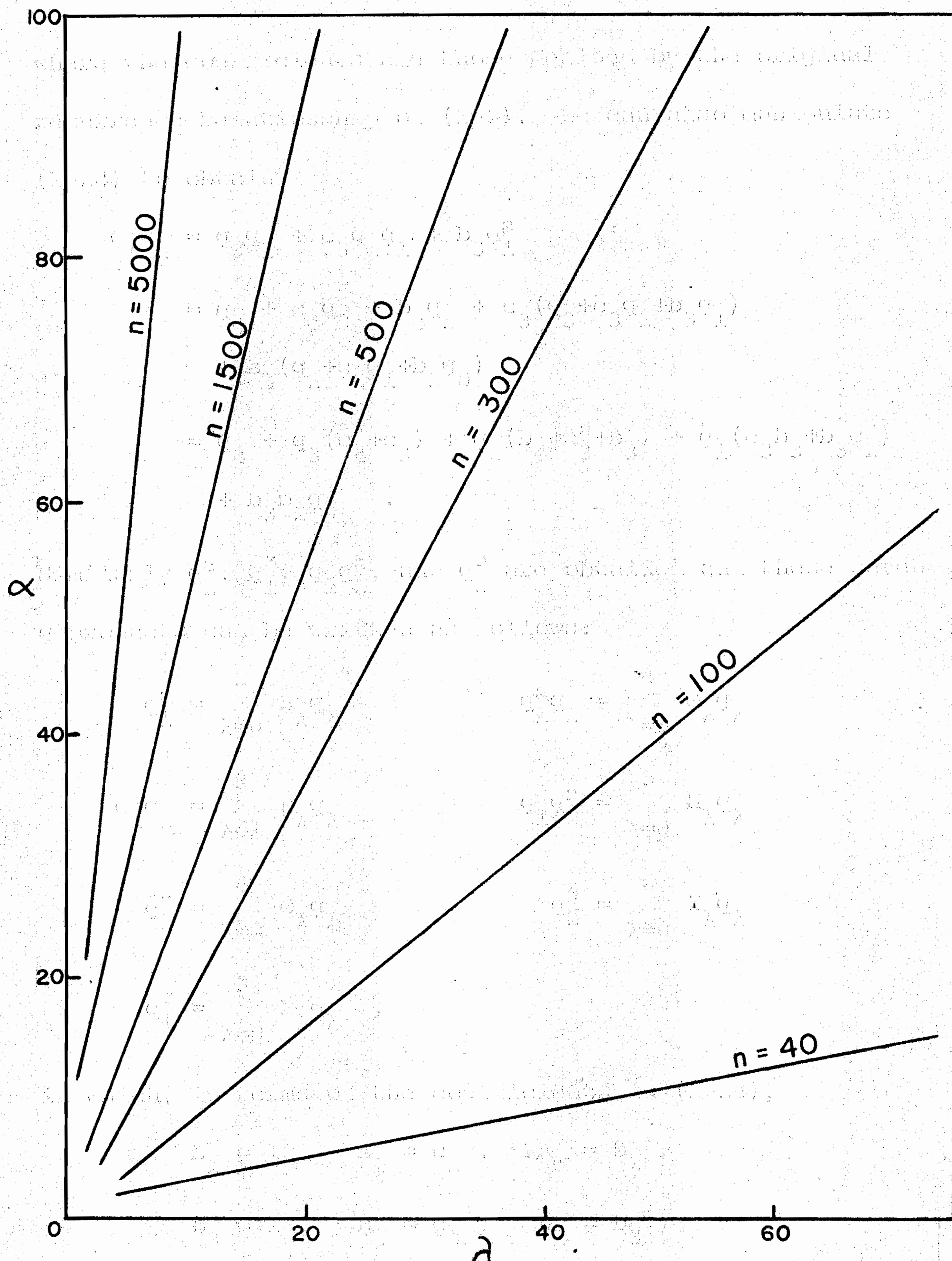

Figure 6. Covariance determinant for the moment estimators of the parameters λ and α.
Covariance determinant = $A_2/n^2 + A_3/n^3 + \ldots$.
Sample size n to make $(A_3/n^3)/(A_2/n^2) = 0.4$ approx.

A problem involved with this study was the difficulty in being forced to use such large sample sizes to "damp off" the asymptotic series for $\mu_3(\hat{\alpha})$ and $\mu_4(\hat{\alpha})$. For instance, for the most extreme case, i.e., when $\lambda = 1$, $\alpha = 100$, only sample sizes of greater than 2,000,000 could be considered. At any rate, one might certainly expect that for a sample size so large that a moment estimator would be at least fairly close to normality, i.e., $\beta_1 = 0$ and $\beta_2 = 3$. This is not the case. Table 3 shows for various values of α and λ approximate values of β_1 and β_2 for the sampling distribution of $\hat{\alpha}$. The sample sizes that were used were also given. Note here that normality is approached in a fairly restricted portion of the "good behavior" region of the (λ,α) plane, i.e., where $\lambda \gg \alpha$. For example, for $\lambda = 50$, $\alpha = 1$, and $n = 750$, $\beta_1 = 0.01$ and $\beta_2 = 3.04$. We note here with interest that there appears to be a certain stability for $\lambda = \alpha$ (>1). That is, for a constant sample size of 500 (large enough to give safe approximations for μ_2, μ_3, and μ_4) β_1 and β_2 remain approximately constant.

The unusual behavior of the distribution of $\hat{\alpha}$ warrants further discussion. After a close examination of the denominator of $\hat{\alpha}$ one might expect the estimator to have a large

Table 3. Measures of Skewness for the Sampling Distribution of $\hat{\alpha}$

α \ λ	1		2		5		10		50		100	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
1	0.55	4.19	0.38	3.97	0.10	3.25	0.03	3.08	0.01	3.04	0.08	3.02
	(500)		(200)		(200)		(500)		(750)		(750)	
2	0.72	4.55	0.46	3.89	0.37	3.74	0.07	3.16	0.03	3.11	0.03	3.00
	(1000)		(500)		(200)		(500)		(500)		(750)	
5	0.76	4.67	0.16	3.32	0.46	3.87	0.18	3.38	0.07	3.14	0.05	3.07
	(5000)		(5000)		(500)		(500)		(500)		(500)	
10	0.52	3.98	0.59	4.09	0.66	4.25	0.46	3.87	0.10	3.20	0.07	3.15
	(20,000)		(5000)		(1000)		(500)		(500)		(500)	
50	0.44	3.84	0.95	4.09	0.49	3.94	0.58	4.06	0.46	3.88	0.21	3.41
	(500,000)		(100,000)		(20,000)		(5000)		(500)		(500)	
100	0.50	4.55	0.44	3.84	0.35	3.69	0.49	3.93	0.64	4.15	0.48	3.88
	(2,000,000)		(500,000)		(100,000)		(20,000)		(1000)		(500)	

The number in parentheses in this table is n the sample size.

$\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$

variance for $\alpha \gg \lambda$. Call this denominator $t = m_2 - m_1'$, where $E(t) = \lambda^2/\alpha - 1/n[\lambda + \lambda^2/\alpha]$. The situation under which one would expect a large variance of $\hat{\alpha}$ would be that for which there is an appreciable probability that the denominator is in the vicinity of 0 or precisely at 0. Letting λ grow very small in relation to α will cause $E(t)$ to approach 0. At the same time, if we look at the variance of the denominator we see that

$$\mathrm{Var}(t) = \frac{2\lambda^2(\lambda+\alpha)(\alpha+1)(\alpha+3\lambda)}{n\alpha^3} - \frac{\lambda^2(\lambda+\alpha)(12\lambda+8\alpha+2\lambda\alpha+2\alpha^2)}{n^2\alpha^3} + \frac{\lambda(\lambda+\alpha)(\alpha^2+6\lambda\alpha+6\lambda^2)}{\alpha^3 n^3} + \ldots . \qquad (5\text{-}9)$$

Note from this that if we impose the same limiting condition as before, i.e., $\alpha \gg \lambda$, $\mathrm{Var}(t)$ grows smaller. This, in addition to the above argument on $E(t)$, indicates that under this limiting condition a "bulk" of the probability will be in the vicinity of 0. Thus the explosive nature of the distribution of $\hat{\alpha}$ results. Unfortunately, however, this is not reflected in the first term in the expansion of the bias or variance but in the latter terms which, until now, have not been investigated.

The expressions for E(t) and Var(t) also indicate the nature of the region of the (n,α,λ) space in which $Var(\hat{\alpha})$ is not quite so large. Obviously as λ increases in proportion to α the distribution of t moves away from the "trouble" area, that is, the vicinity of $t = 0$. This reduces the chance of obtaining a sample for which $t = 0$. Then an increase in n will bring about a corresponding decrease in Var(t) which even further reduces this chance. Thus while only one sample for which $t = 0$ would theoretically cause the bias and variance to become infinite, λ, α, and n can be chosen in such a way that the chance of this occurrence is practically nil.

Further Remarks

Note that in order that t be 0, m_2 must equal m_1'. It can be shown that for example when $n = 2$ there are an infinite number of samples that will result in $m_2 = m_1'$, and thus result in an infinite value for $\hat{\alpha}$. In fact, when $n = 2$ solutions of $m_2 = m_1'$ are

$$\begin{aligned} x_1 &= b(b-1) \\ x_2 &= b(b+1) \end{aligned} \qquad b = 1,2,3,4,\ldots \quad .$$

(We don't consider $b = 0$, since this generates the null sample.)

For the sample above, m_1' and m_2 are:

$$m_1' = \frac{b^2 + b^2}{2}$$

$$= b^2 \quad ,$$

$$m_2 = \frac{(-b)^2 + (b)^2}{2}$$

$$= b^2 \quad .$$

Thus $$m_1' = m_2 = b^2 \quad .$$

For an arbitrary sample size n there are a number of classes of samples $(x_1, x_2, \ldots, x_n)$ which are such that $m_2 = m_1'$. For example, consider the following sample for any positive integer p and sample size n:

$$\begin{aligned} x_1 &= (n-1)p^2 + p \\ x_2 &= (n-1)p^2 + p \\ &\vdots \\ x_{n-1} &= (n-1)p^2 + p \\ x_n &= (n-1)p^2 - (n-1)p \quad . \end{aligned}$$

It can easily be shown that for this sample, $m_2 = m_1'$.

Consider also the class of samples,

$$
\begin{aligned}
x_1 &= 3p^2(n-1)(n+1)(3n-2) + 6p \\
x_2 &= \quad " \qquad " \qquad " \quad + 12p \\
x_3 &= \quad " \qquad " \qquad " \quad + 18p \\
&\vdots \\
x_{n-1} &= \quad " \qquad " \qquad " \quad + 6(n-1)p \\
x_n &= \quad " \qquad " \qquad " \quad + 3n(n-1)p \quad ,
\end{aligned}
$$

where once again we make the restriction only that p is a positive integer. Here again any such sample gives $m_2 = m_1'$. Thus one could say in general that there will be a finite probability of obtaining an infinite value for $\hat{\alpha}$ for a sample of size n. This being the case, one would certainly expect the distribution of $\hat{\alpha}$ to be explosive in the region where the chance of an infinite value is non-negligible.

(b) Evans' Estimators

In Table 1 of Chapter III it was noted that the moment estimators for the notation of Evans with parameters a and m are $\hat{a} = (m_2-m_1')/m'$ and $\hat{m} = m_1'$. It turns out that the behavior of $\hat{a}$ is in marked contrast with that of $\hat{\alpha}$. A preliminary investigation of the denominator of this statistic might lead one to suspect this conclusion. Only a null sample will result in a zero denominator.

1. <u>First Two Moments of $\hat{a}$</u>

The following expressions were obtained for the mean and variance of $\hat{a}$; (To avoid complication, we express the results in terms of λ and α, the parameters used previously.)

$$E(\hat{a}) = \lambda/\alpha - \frac{(\lambda+\alpha)(\alpha+1)}{\alpha(n\alpha+1)} \quad \text{(approx.)}, \tag{5-10}$$

$$\mathrm{Var}(\hat{a}) = \Lambda_1^{(2)}/n + \Lambda_2^{(2)}/n^2 + \Lambda_3^{(2)}/n^3 + \Lambda_4^{(2)}/n^4 + \ldots . \tag{5-11}$$

Here the coefficients $\Lambda_1^{(2)}$, $\Lambda_2^{(2)}$, $\Lambda_3^{(2)}$, and $\Lambda_4^{(2)}$ are as follows:

$$\Lambda_1^{(2)} = (\lambda+\alpha)\{\lambda(2\alpha+3) + 2\alpha(\alpha+1)\}/\alpha^3 \quad ,$$

$$\Lambda_2^{(2)} = -\ (\lambda+\alpha)(\alpha+1)\{\lambda^2(2\alpha+14) + \lambda\alpha(2\alpha+14) + 2\alpha^2\}/\lambda\alpha^4 \quad ,$$

$$\Lambda_3^{(2)} = (\lambda+\alpha)(\alpha+1)\{\lambda^3(13\alpha+53) + \lambda^2\alpha(14\alpha+64) + \lambda\alpha^2(2\alpha+12) - 2\alpha^3\}/\lambda^2\alpha^5 \quad ,$$

$$\Lambda_4^{(2)} = -\ (\lambda+\alpha)(\alpha+1)\{\lambda^4(52\alpha+184) + \lambda^3\alpha(64\alpha+244) + \lambda^2\alpha^2(12\alpha+50) - \lambda\alpha^3(2\alpha+10) + 4\alpha^4\}/\lambda^3\alpha^6 \quad .$$

The individual terms are given in Table A.4 for a sample size of 100. Note that in all cases the first term is definitely dominant. (We restrict our study here to cases

Table A.4. Terms in Var $\hat{a}$ n = 100

α \ λ		1	2	3	4	5	10	15	25	50	100
1	(I)	.1800	.4200	.7600	1.200	1.740	5.940	12.64	33.54	129.5	509.0
	(II)	$\underline{.0136}$	$\underline{.0294}$	$\underline{.0517}$	$\underline{.0805}$	$\underline{.1157}$	$\underline{.3876}$	$\underline{.8196}$	$\underline{2.164}$	$\underline{8.324}$	$\underline{32.64}$
	(III)	6240^{3}	1299^{2}	2244^{2}	3454^{2}	4929^{2}	1626^{1}	3421^{1}	8989^{1}	.3446	1.349
	(IV)	$\underline{2390}^{4}$	$\underline{4850}^{4}$	$\underline{8284}^{4}$	$\underline{1269}^{3}$	$\underline{1799}^{3}$	$\underline{5883}^{3}$	$\underline{1233}^{2}$	$\underline{3229}^{2}$	$\underline{1235}^{1}$	$\underline{4830}^{1}$
2	(I)	.0713	.1300	.2063	.3000	.4113	1.230	2.486	6.311	23.53	90.78
	(II)	$\underline{3488}^{2}$	$\underline{5700}^{2}$	$\underline{8688}^{2}$	$\underline{1238}^{1}$	$\underline{1675}^{1}$	$\underline{4878}^{1}$	$\underline{9771}^{1}$	$\underline{.2462}$	$\underline{.9127}$	$\underline{3.512}$
	(III)	8747^{4}	1388^{3}	2065^{3}	2897^{3}	3880^{3}	1103^{2}	2188^{2}	5471^{2}	2016^{1}	7731^{1}
	(IV)	$\underline{1800}^{5}$	$\underline{2715}^{5}$	$\underline{3976}^{5}$	$\underline{5524}^{5}$	$\underline{7347}^{5}$	$\underline{2055}^{4}$	$\underline{4051}^{4}$	$\underline{1007}^{3}$	$\underline{3693}^{3}$	$\underline{1413}^{2}$
3	(I)	.0489	.0778	.1133	.1556	.2044	.5489	1.060	2.582	9.304	35.25
	(II)	$\underline{1936}^{2}$	$\underline{2691}^{2}$	$\underline{3733}^{2}$	$\underline{4995}^{2}$	$\underline{6463}^{2}$	$\underline{1681}^{1}$	$\underline{3210}^{1}$	$\underline{7753}^{1}$	$\underline{2775}$	$\underline{1.048}$
	(III)	3411^{4}	4687^{4}	6341^{4}	8332^{4}	1064^{3}	2683^{3}	5060^{3}	1210^{2}	4293^{2}	1614^{1}
	(IV)	$\underline{5080}^{6}$	$\underline{6330}^{6}$	$\underline{8395}^{6}$	$\underline{1091}^{5}$	$\underline{1382}^{5}$	$\underline{3410}^{5}$	$\underline{6378}^{5}$	$\underline{1512}^{4}$	$\underline{5329}^{4}$	$\underline{1996}^{3}$
4	(I)	.0398	.0581	.0798	.1050	.1336	.3281	.6086	1.427	4.978	18.52
	(II)	$\underline{1387}^{2}$	$\underline{1734}^{2}$	$\underline{2251}^{2}$	$\underline{2875}^{2}$	$\underline{3593}^{2}$	$\underline{8509}^{2}$	$\underline{1559}^{1}$	$\underline{3621}^{1}$	$\underline{.1254}$	$\underline{.4648}$
	(III)	1897^{4}	2397^{4}	3033^{4}	3797^{4}	4675^{4}	1067^{3}	1926^{3}	4414^{3}	1513^{2}	5577^{2}
	(IV)	$\underline{2339}^{6}$	$\underline{2496}^{6}$	$\underline{3084}^{6}$	$\underline{3813}^{6}$	$\underline{4650}^{6}$	$\underline{1037}^{5}$	$\underline{1851}^{5}$	$\underline{4199}^{5}$	$\underline{1426}^{4}$	$\underline{5232}^{4}$
5	(I)	.0350	.0482	.0634	.0806	.1000	.2280	.4080	.9240	3.124	11.42
	(II)	$\underline{1117}^{2}$	$\underline{1297}^{2}$	$\underline{1603}^{2}$	$\underline{1974}^{2}$	$\underline{2400}^{2}$	$\underline{5256}^{2}$	$\underline{9280}^{2}$	$\underline{2070}^{1}$	$\underline{6975}^{1}$	$\underline{.2541}$
	(III)	1253^{4}	1503^{4}	1812^{4}	2184^{4}	2600^{4}	5479^{4}	9510^{4}	2090^{3}	6949^{3}	2515^{2}
	(IV)	$\underline{1386}^{6}$	$\underline{1282}^{6}$	$\underline{1500}^{6}$	$\underline{1780}^{6}$	$\underline{2116}^{6}$	$\underline{4330}^{6}$	$\underline{7410}^{6}$	$\underline{1616}^{5}$	$\underline{5296}^{5}$	$\underline{1905}^{4}$
10	(I)	.0267	.0319	.0376	.0437	.0503	.0900	.1412	.2783	.8220	2.772
	(II)	$\underline{6945}^{3}$	$\underline{6706}^{3}$	$\underline{7239}^{3}$	$\underline{8100}^{3}$	$\underline{9075}^{3}$	$\underline{1540}^{2}$	$\underline{2374}^{2}$	$\underline{4612}^{2}$	$\underline{1349}^{1}$	$\underline{4528}^{1}$
	(III)	4142^{5}	4628^{5}	4900^{5}	5308^{5}	5800^{5}	9174^{5}	1372^{4}	2595^{4}	7427^{4}	2460^{3}
	(IV)	$\underline{4420}^{7}$	$\underline{2145}^{7}$	$\underline{2111}^{7}$	$\underline{2257}^{7}$	$\underline{2455}^{7}$	$\underline{3180}^{7}$	$\underline{5613}^{7}$	$\underline{1042}^{6}$	$\underline{2928}^{6}$	$\underline{9608}^{6}$
15	(I)	.0243	.0275	.0309	.0345	.0382	.0600	.0867	.1547	.4102	1.288
	(II)	$\underline{5836}^{3}$	$\underline{5228}^{3}$	$\underline{5359}^{3}$	$\underline{5696}^{3}$	$\underline{6131}^{3}$	$\underline{9047}^{3}$	$\underline{1280}^{2}$	$\underline{2248}^{2}$	$\underline{5894}^{2}$	$\underline{1841}^{1}$
	(III)	2379^{5}	2738^{5}	2751^{5}	2819^{5}	2937^{5}	3934^{5}	5329^{5}	8999^{5}	2287^{4}	7028^{4}
	(IV)	$\underline{3108}^{7}$	$\underline{8937}^{8}$	$\underline{7760}^{8}$	$\underline{7895}^{8}$	$\underline{8273}^{8}$	$\underline{1114}^{7}$	$\underline{1493}^{7}$	$\underline{2474}^{7}$	$\underline{6139}^{7}$	$\underline{1857}^{6}$
25	(I)	.0225	.0243	.0262	.0281	.0300	.0410	.0536	.0840	.1896	.5280
	(II)	$\underline{5043}^{3}$	$\underline{4229}^{3}$	$\underline{4116}^{3}$	$\underline{4186}^{3}$	$\underline{4333}^{3}$	$\underline{5509}^{3}$	$\underline{7038}^{3}$	$\underline{1082}^{2}$	$\underline{2409}^{2}$	$\underline{6666}^{2}$
	(III)	1262^{5}	1630^{5}	1560^{5}	1513^{5}	1497^{5}	1649^{5}	1966^{5}	2835^{5}	5993^{5}	1615^{4}
	(IV)	$\underline{2468}^{6}$	$\underline{3433}^{8}$	$\underline{2301}^{8}$	$\underline{2240}^{8}$	$\underline{2309}^{8}$	$\underline{2796}^{8}$	$\underline{3375}^{8}$	$\underline{4822}^{8}$	$\underline{9929}^{8}$	$\underline{2617}^{7}$
50	(I)	.0213	.0221	.0229	.0238	.0247	.0294	.0346	.0461	.0820	.1848
	(II)	$\underline{4500}^{3}$	$\underline{3576}^{3}$	$\underline{3338}^{3}$	$\underline{3263}^{3}$	$\underline{3263}^{3}$	$\underline{3594}^{3}$	$\underline{4107}^{3}$	$\underline{5355}^{3}$	$\underline{9384}^{3}$	$\underline{2099}^{2}$
	(III)	5730^{6}	9938^{6}	9157^{6}	8406^{6}	7873^{6}	6926^{6}	7033^{6}	8149^{6}	1287^{5}	2724^{5}
	(IV)	$\underline{2172}^{7}$	$\underline{1150}^{8}$	$\underline{2082}^{9}$	$\underline{2130}^{9}$	$\underline{2892}^{9}$	$\underline{4897}^{9}$	$\underline{5698}^{9}$	$\underline{7011}^{8}$	$\underline{1105}^{8}$	$\underline{2279}^{8}$
100	(I)	.0206	.0210	.0214	.0219	.0223	.0245	.0267	.0316	.0455	.0810
	(II)	$\underline{4250}^{3}$	$\underline{3279}^{3}$	$\underline{2987}^{3}$	$\underline{2863}^{3}$	$\underline{2807}^{3}$	$\underline{2837}^{3}$	$\underline{3013}^{3}$	$\underline{3478}^{3}$	$\underline{4924}^{3}$	$\underline{8686}^{3}$
	(III)	2730^{6}	7318^{6}	6605^{6}	5849^{6}	5270^{6}	3903^{6}	3475^{6}	3306^{6}	3873^{6}	6115^{6}
	(IV)	$\underline{2073}^{7}$	4559^{9}	3820^{9}	3219^{9}	2130^{9}	$\underline{2980}^{10}$	$\underline{8870}^{10}$	$\underline{1248}^{9}$	$\underline{1671}^{9}$	$\underline{2641}^{9}$

Key to Table A.4. (a) An underlined entry is negative.

(b) (I) - (IV) refer to the four terms in Var $\hat{a}$.

(c) The index is the number of zeros preceeding the first digit to the decimal point.

in which λ and α both exceed 1. One might note that if λ and α are less than 1, terms beyond the asymptotic variance will be non-negligible.) The expression in (5-11) is certainly more useful than the corresponding variance expression for $\hat{\alpha}$. In fact, it can safely be used (the ratio of the fourth term to the first being less than 0.05) when n is 15 or greater. Moreover, if $n \geq 60$ the usual term for the asymptotic variance can be used safely as an approximation for the true variance. It can not be said for sure whether or not the approximation given by the asymptotic variance is an underestimate or an overestimate of the true variance. This is due to the alternating signs in (5-11). It is evident that for some points in the (n,α,λ) space, this n^{-1} term might be in excess of $\mathrm{Var}(\hat{a})$.

2. <u>Covariance Determinant</u>

An expression was obtained using orthogonal statistics for the covariance determinant

$$D_E = \begin{vmatrix} \mathrm{Var}\ \hat{a}, & \mathrm{Cov}(\hat{a},\hat{m}) \\ \mathrm{Cov}(\hat{a},\hat{m}), & \mathrm{Var}\ \hat{m} \end{vmatrix} ,$$

through terms in n^{-3}. This expression is as follows:

$$D_E = \frac{2\lambda(\lambda+\alpha)^3(\alpha+1)}{\alpha^4 n^2} - \frac{2(\lambda+\alpha)^3(\alpha+1)[\lambda(\alpha+6)+\alpha]}{\alpha^5 n^3} + \ldots . \tag{5-12}$$

As was done for the estimators of Anscombe's form, the first two terms of this expansion were tabulated assuming $n = 100$. They are shown in Table A.5.

Note from Table A.5 that D_E is relatively well-behaved and can be approximated well by (5-12) for reasonable sample sizes. For example, if $\lambda = 5$ and $\alpha = 1$ and $n = 100$, D_E, term by term, is 0.432 - 0.031104. By contrast for a case in which $\alpha \gg \lambda$, e.g., $\lambda = 4$, $\alpha = 100$, and $n = 100$, the successive terms are 0.00090889 - 0.000011906. Note that it can generally be stated that, unlike the case of simultaneous estimation for Anscombe's parameters, the covariance determinant D_E is small in the area of low λ and high α, whereas it is relatively high for high λ and low α. Nevertheless, despite the region of the parameter space in which one is working, reasonable sample sizes, e.g., $n > 50$, result in good approximations using both the n^{-2} and n^{-3} term. Actually the table shows clearly that for $\alpha > \lambda$ one cannot be too far wrong in using only the asymptotic expansion. Note also that the n^{-3} term is negative which indicates that

Table A.5. First and Second Terms in Covariance Determinant for Estimators of Evan's Form using n = 100

α \ λ	1	2	3	4	5	10	15	25	50	100
1	32^{2} $\underline{2560}^{3}$	216^{1} $\underline{162}^{2}$	768^{1} $\underline{5632}^{2}$	2000^{0} $\underline{14500}^{1}$	43200^{0} $\underline{31104}^{1}$	5.3240 $\underline{37800}^{0}$	24.576 $\underline{1.7367}$	175.76 $\underline{12.374}$	2653.02 $\underline{186.24}$	41212.0 $\underline{2888.96}$
2	10125^{2} $\underline{50625}^{4}$	48^{2} $\underline{216}^{3}$	14063^{1} $\underline{60937}^{3}$	32400^{1} $\underline{13770}^{2}$	64313^{1} $\underline{27011}^{2}$	64800^{0} $\underline{26568}^{1}$	2.7636 $\underline{11238}^{0}$	18.453 $\underline{74549}^{0}$	263.64 $\underline{10.598}$	3979.53 $\underline{159.579}$
3	63210^{3} $\underline{25284}^{4}$	24691^{2} $\underline{86491}^{4}$	6400^{2} $\underline{21330}^{3}$	13551^{1} $\underline{44040}^{3}$	25284^{1} $\underline{80909}^{3}$	21699^{0} $\underline{67266}^{2}$	86400^{0} $\underline{26496}^{1}$	5.4202 $\underline{16478}^{0}$	73.520 $\underline{2.220}$	1079.23 $\underline{32.485}$
4	48828^{3} $\underline{17090}^{4}$	16875^{2} $\underline{50625}^{4}$	40195^{2} $\underline{11389}^{3}$	8000^{2} $\underline{22000}^{3}$	14238^{1} $\underline{38443}^{3}$	10719^{0} $\underline{27869}^{2}$	40189^{0} $\underline{10315}^{1}$	2.3817 $\underline{60496}^{1}$	30.755 $\underline{77507}^{0}$	439.4 $\underline{11.029}$
5	41472^{3} $\underline{13271}^{4}$	13171^{2} $\underline{35562}^{4}$	29491^{2} $\underline{74711}^{4}$	55987^{2} $\underline{13717}^{3}$	96000^{2} $\underline{23040}^{3}$	64800^{1} $\underline{14904}^{2}$	23040^{0} $\underline{52224}^{2}$	1.2960 $\underline{29030}^{1}$	15.972 $\underline{35458}^{0}$	222.26 $\underline{4.912}$
10	29282^{3} $\underline{76133}^{5}$	76032^{3} $\underline{15967}^{4}$	14500^{2} $\underline{28030}^{4}$	24147^{2} $\underline{44672}^{4}$	37125^{2} $\underline{66825}^{4}$	17600^{1} $\underline{29920}^{3}$	31563^{1} $\underline{85938}^{3}$	23581^{0} $\underline{38673}^{2}$	2.376 $\underline{38491}^{1}$	29.282 $\underline{47144}^{0}$
15	25891^{3} $\underline{62138}^{5}$	62110^{3} $\underline{11801}^{4}$	11059^{2} $\underline{19169}^{4}$	17342^{2} $\underline{28615}^{4}$	25284^{2} $\underline{40454}^{4}$	98765^{2} $\underline{14815}^{3}$	25600^{1} $\underline{37547}^{3}$	10114^{0} $\underline{14564}^{2}$	86795^{0} $\underline{12325}^{1}$	9.613 $\underline{13555}^{0}$
25	23397^{3} $\underline{52409}^{5}$	52404^{3} $\underline{91183}^{5}$	87668^{3} $\underline{13793}^{4}$	12987^{2} $\underline{19350}^{4}$	17971^{2} $\underline{25879}^{4}$	57075^{2} $\underline{76481}^{4}$	12780^{1} $\underline{16699}^{3}$	41600^{1} $\underline{53248}^{3}$	28080^{0} $\underline{35381}^{2}$	2.600 $\underline{32500}^{1}$
50	21649^{3} $\underline{45895}^{5}$	45894^{3} $\underline{74349}^{5}$	72890^{3} $\underline{10593}^{4}$	10279^{2} $\underline{14083}^{4}$	13576^{2} $\underline{17920}^{4}$	35251^{2} $\underline{43006}^{4}$	67228^{2} $\underline{79777}^{4}$	17213^{1} $\underline{19967}^{3}$	81600^{1} $\underline{93024}^{3}$	55080^{0} $\underline{62240}^{2}$
100	20812^{3} $\underline{42873}^{5}$	42873^{3} $\underline{66882}^{5}$	66219^{3} $\underline{92265}^{5}$	90889^{3} $\underline{11906}^{4}$	11692^{2} $\underline{14732}^{4}$	26886^{2} $\underline{31188}^{4}$	46083^{2} $\underline{51920}^{4}$	98633^{2} $\underline{10850}^{3}$	34088^{1} $\underline{36815}^{3}$	16160^{0} $\underline{17291}^{2}$

The indices refer to the number of zeros preceeding the first digit to the decimal point. Underlined number refers to negative number.

the asymptotic covariance determinant may, in some cases, be in excess of the true determinant.

3. Remarks on the Distribution of $\hat{a}$

There was every indication, from the results of the study of the variance, that the behavior of $\hat{a}$ is very stable. To verify this further, expansions were derived for $\mu_3(\hat{a})$ and $\mu_4(\hat{a})$ through terms in n^{-4}. The actual expressions are found in Appendix D. As was done for $\hat{\alpha}$, "safe" sample sizes were chosen for a small set of values of α and λ, and $\beta_1(\hat{a})$ and $\beta_2(\hat{a})$ calculated for these combinations of λ, α, and n. The purpose of this was to go a bit further in showing the contrast between $\hat{\alpha}$ and $\hat{a}$. The following is a table of these results:

n	λ	α	$\beta_1(\hat{a})$	$\beta_2(\hat{a})$
100	1	1	0.91	4.93
200	1	1	0.48	4.10
100	1	2	0.53	4.46
200	1	2	0.28	3.71
100	2	2	0.55	4.09
200	2	2	0.27	3.58
100	2	1	0.69	4.86
200	2	1	0.45	4.03

The above table indicates that as λ and α become greater than unity, the distribution of a shows less departure from normality. The most marked change in this direction comes about with an increase in α. One striking comparison between these results and those for $\hat{\alpha}$ is the safe sample size used in obtaining $\beta_1(\hat{a})$ and $\beta_2(\hat{a})$. We safely used $n = 100$ and 200 here in order to "damp off" terms beyond n^{-4}, while for $\hat{\alpha}$ we were never able to use a sample size quite so small.

It can certainly be said from all of our results regarding $\hat{a}$ that the latter does not behave in a manner similar to $\hat{\alpha}$ and that under most practical conditions we can determine such properties as the bias and variance of $\hat{a}$, while due to the chaotic behavior of the distribution of $\hat{\alpha}$, this cannot be done.

4. Examples — $\mu_3(\hat{a})$ and $\mu_4(\hat{a})$

In order to give the reader an idea of the amount of work involved in obtaining the higher moments of one of these estimators, we shall proceed to show some of the details in obtaining $\mu_3(\hat{a})$ and $\mu_4(\hat{a})$.

4.1 $\underline{\mu_3(\hat{a})}$

In obtaining $\mu_3(\hat{a})$, we first expand $E(\hat{a}-a)^3$ in ascending powers of Q_1 and Q_2 as follows:

$$E(\hat{a}-a)^3 = E\{\frac{(\lambda/\alpha)Q_1+Q_2-Q_1^2}{\lambda+Q_1}\}^3$$

$$= E\{\frac{Q_1^3}{\alpha^3} + \frac{3Q_1^2Q_2}{\alpha^2\lambda} + \frac{3Q_1Q_2^2}{\lambda^2\alpha} + \frac{Q_2^3}{\lambda^3} - \frac{3Q_1^4}{\lambda\alpha^2} - \frac{6Q_1^3Q_2}{\lambda^2\alpha}$$

$$- \frac{3Q_1^2Q_2^2}{\lambda^3} + \frac{3Q_1^5}{\lambda^2\alpha} + \frac{3Q_1^4Q_2}{\lambda^3} - Q_1^6/\lambda^3\}\cdot\{1 - \frac{3Q_1}{\lambda} + \frac{6Q_1^2}{\lambda^2}$$

$$- \frac{10Q_1^3}{\lambda^3} + \frac{15Q_1^4}{\lambda^4} - \frac{21Q_1^5}{\lambda^5} + \ldots.\} \quad .$$

We will proceed to expand through order 8 in the Q's so as to obtain the results through terms in n^{-4}. Letting G_k denote that part of $(\hat{a}-a)^3$ which gives terms of order k in the Q's, we have:

$$G_3 = \frac{Q_1^3}{\alpha^3} + \frac{3Q_1^2Q_2}{\lambda\alpha^2} + \frac{3Q_1Q_2^2}{\lambda^2\alpha} + \frac{Q_2^3}{\lambda^3} \quad , \tag{5-13}$$

$$G_4 = -\frac{Q_1^4}{\lambda\alpha^3}(3+3\alpha) - \frac{Q_1^3Q_2}{\lambda^2\alpha^2}(9+6\alpha) - \frac{Q_1^2Q_2^2}{\alpha\lambda^3}(9+3\alpha) - \frac{3Q_1Q_2^3}{\lambda^4} \quad , \tag{5-13-a}$$

$$G_5 = \frac{Q_1^5}{\lambda^2\alpha^3}(3\alpha^2+9\alpha+6) + \frac{Q_1^4Q_2}{\alpha^2\lambda^3}(3\alpha^2+18\alpha+18) + \frac{Q_1^3Q_2^2}{\lambda^4\alpha}(18+9\alpha)$$

$$+ \frac{6Q_1^2Q_2^3}{\lambda^5} \quad , \tag{5-13-b}$$

$$G_6 = -\frac{Q_1^6}{\lambda^3\alpha^3}(\alpha^3+9\alpha^2+18\alpha+10) - \frac{Q_1^5Q_2}{\lambda^4\alpha^2}(9\alpha^2+36\alpha+30)$$

$$-\frac{Q_1^4Q_2^2}{\lambda^5\alpha}(30+18\alpha) - \frac{10Q_1^3Q_2^3}{\lambda^6} \quad , \qquad (5\text{-}13\text{-}c)$$

$$G_7 = \frac{Q_1^7}{\lambda^4\alpha^3}(3\alpha^3+18\alpha^2+30\alpha+15) + \frac{Q_1^6Q_2}{\lambda^5\alpha^2}(18\alpha^2+60\alpha+45)$$

$$+\frac{Q_1^5Q_2^2}{\lambda^6\alpha}(30\alpha+45) + \frac{15Q_1^4Q_2^3}{\lambda^7} \quad , \text{ and} \qquad (5\text{-}13\text{-}d)$$

$$G_8 = -\frac{Q_1^8}{\lambda^5\alpha^3}(6\alpha^3+30\alpha^2+45\alpha+21) - \frac{Q_1^7Q_2}{\lambda^6\alpha^2}(30\alpha^2+90\alpha+63)$$

$$-\frac{Q_1^6Q_2^2}{\lambda^7\alpha}(45\alpha+63) - \frac{21Q_1^5Q_2^3}{\lambda^8} \quad , \qquad (5\text{-}13\text{-}e)$$

We shall first obtain the n^{-2} term in the expansion,

$$E(\hat{a}-a)^3 = E_2^{(3)}/n^2 + E_3^{(3)}/n^3 + E_4^{(3)}/n^4 + \ldots \quad .$$

For this we use G_3 and G_4. From (5-13) and (5-13-a) we have:

$$E_2^{(3)} = (1^3)_2 + \frac{3(1^22)_2}{\lambda\alpha^2} + \frac{3(12^2)_2}{\lambda^2\alpha} + \frac{(2^3)_2}{\lambda^3} - \frac{(1^4)_2}{\lambda\alpha^3}(3+3\alpha)$$

$$-\frac{(1^32)_2}{\lambda^2\alpha^2}(9+6\alpha) - \frac{(1^22^2)_2}{\alpha\lambda^3}(9+3\alpha) - \frac{3(12^3)_2}{\lambda^4} \quad .$$

From Appendix B (using the same notation as in the appendix) we can write:

$$E_2^{(3)} = \frac{\lambda\chi\psi}{\alpha^3} + \frac{\lambda\chi^2 a}{\alpha^3} + \frac{12\chi^2\psi a}{\alpha^5} + \frac{\chi^2 a(8\chi b+\psi^2)}{\alpha^5}$$

$$-\frac{6\lambda\chi^2(\alpha+1)}{\alpha^5} - \frac{6\lambda\chi^3 a(3+2\alpha)}{\alpha^6}$$

After simplification, the above yields:

$$E_2^{(3)} = \frac{(\lambda+\alpha)}{\lambda\alpha^5}\{\lambda^3(2\alpha^2+37\alpha+37) + \lambda^2\alpha(4\alpha^2+65\alpha+62)$$

$$+ \lambda\alpha^2(2\alpha^2+32\alpha+30) + 4\alpha^3(\alpha+1)\} \quad .$$

We can now find $E_3^{(3)}$ by considering G_4, G_5, and G_6. Thus from (5-13-a), (5-13-b), and (5-13-c), we have:

$$E_3^{(3)} = -\frac{3(1^4)_3(\alpha+1)}{\lambda\alpha^3} - \frac{3(1^3 2)_3(2\alpha+3)}{\lambda^2\alpha^2} - \frac{3(1^2 2^2)_3(\alpha+3)}{\alpha\lambda^3}$$

$$- \frac{3(12^3)_3}{\lambda^4} + \frac{3(1^5)_3(\alpha^2+3\alpha+2)}{\lambda^2\alpha^3} + \frac{3(1^4 2)_3(\alpha^2+6\alpha+6)}{\alpha^2\lambda^3}$$

$$+ \frac{9(1^3 2^2)_3(\alpha+2)}{\lambda^4\alpha} + \frac{6(1^2 2^3)_3}{\lambda^5} - \frac{(1^6)_3(\alpha^3+9\alpha^2+18\alpha+10)}{\lambda^3\alpha^3}$$

$$- \frac{3(1^5 2)_3(3\alpha^2+12\alpha+10)}{\lambda^4\alpha^2} - \frac{6(1^4 2^2)_3(3\alpha+5)}{\lambda^5\alpha} - \frac{10(1^3 2^3)_3}{\lambda^6} \quad .$$

$$E_3^{(3)} = -\frac{3\chi(\alpha^2+6\lambda\alpha+6\lambda^2)(\alpha+1)}{\alpha^6} - \frac{18\psi\chi^2 a(2\alpha+3)}{\alpha^6}$$

$$- \frac{24\chi^2 a(\lambda\chi b+\psi^2)(\alpha+3)}{\lambda\alpha^6} - \frac{24\chi^2\psi a[\psi^2+\lambda\chi(6\alpha+11)]}{\alpha^5\lambda^2}$$

$$+ \frac{30\chi^2\psi ab}{\alpha^6\lambda^2} + \frac{36\chi^3 a(\alpha^2+6\alpha+6)}{\alpha^6} + \frac{126\chi^3\psi ab}{\alpha^6\lambda}$$

$$+ \frac{24\chi^3 a[(5\alpha+7)\lambda\chi+\psi^2]}{\alpha^6\lambda^2} - \frac{15\chi^3(\alpha^3+9\alpha^2+18\alpha+10)}{\alpha^6}$$

$$- \frac{36a\chi^4(3\alpha+5)}{\alpha^6\lambda} \quad .$$

This can then be simplified algebraically to yield:

$$E_3^{(3)} = -\frac{3(\alpha+1)(\lambda+\alpha)}{\lambda\alpha^6}\{\lambda^3(\alpha^2+52\alpha+168) + \lambda^2\alpha(2\alpha^2+104\alpha+340) + \lambda\alpha^2(\alpha^2+62\alpha+213) + \alpha^3(10\alpha+40)\} \quad .$$

For $E_4^{(3)}$ we have, using G_5, G_6, G_7, and G_8 from (5-13-b), (5-13-c), (5-13-d), and (5-13-e):

$$\begin{aligned} E_4^{(3)} = {} & \frac{3(1^5)_4(\alpha^2+3\alpha+2)}{\lambda^2\alpha^3} + \frac{3(1^4 2)_4(\alpha^2+6\alpha+6)}{\lambda^3\alpha^2} \\ & + \frac{9(1^3 2^2)_4(\alpha+2)}{\lambda^4\alpha} + \frac{6(1^2 2^3)_4}{\lambda^5} - \frac{(1^6)_4(\alpha^3+9\alpha^2+18\alpha+10)}{\lambda^3\alpha^3} \\ & - \frac{3(1^5 2)_4(3\alpha^2+12\alpha+10)}{\lambda^4\alpha^2} - \frac{6(1^4 2^2)_4(3\alpha+5)}{\lambda^5\alpha} \\ & - \frac{10(1^3 2^3)_4}{\lambda^6} + \frac{3(1^7)_4(\alpha^3+6\alpha^2+10\alpha+5)}{\lambda^4\alpha^3} \\ & + \frac{3(1^6 2)_4(6\alpha^2+20\alpha+15)}{\lambda^5\alpha^2} + \frac{15(1^5 2^2)_4(2\alpha+3)}{\lambda^6\alpha} \\ & + \frac{15(1^4 2^3)_4}{\lambda^7} - \frac{3(1^8)_4(2\alpha^3+10\alpha^2+15\alpha+7)}{\lambda^5\alpha^3} \\ & - \frac{3(1^7 2)_4(10\alpha^2+30\alpha+21)}{\lambda^6\alpha^2} - \frac{9(1^6 2^2)_4(5\alpha+7)}{\lambda^7\alpha} \\ & - \frac{21(1^5 2^3)_4}{\lambda^8} \quad . \end{aligned}$$

If we use Appendix B and insert the expressions for the "round bracket" terms in terms of the parameters λ and α and simplify, we have:

$$E_4^{(3)} = \frac{(\lambda+\alpha)(\alpha+1)}{\lambda^3\alpha^7}\{\lambda^5(122\alpha^2+1944\alpha+4381)$$

$$+ \lambda^4\alpha(253\alpha^2+4287\alpha+9962)+ \lambda^3\alpha^2(157\alpha^2+2978\alpha+7241)$$

$$+ \lambda^2\alpha^3(26\alpha^2+630\alpha+1640) + \lambda\alpha^4(-2\alpha-18) - 4\alpha^5\} \quad .$$

These $E_k^{(3)}$'s represent terms in the expansion of the crude moment. The correction is then applied in order to obtain $\mu_3(\hat{a})$. In applying this correction we have:

$$\mu_3(\hat{a}) = E[(\hat{a}-a) - E(\hat{a}-a)]^3 \quad ,$$

$$= E(\hat{a}-a)^3 - 3[E(\hat{a}-a)^2]\cdot E(\hat{a}-a) + 2[E(\hat{a}-a)]^3 \quad . \tag{5-14}$$

In order to evaluate $\mu_3(\hat{a})$, we need expansions through terms in n^{-4} for $E(\hat{a}-a)$ and $E(\hat{a}-a)^2$. From (5-10) we have:

$$E(\hat{a}-a) = -\frac{(\lambda+\alpha)(\alpha+1)}{n\alpha^2}\{1 - \frac{1}{n\alpha} + \frac{1}{n^2\alpha^2} - \frac{1}{n^3\alpha^3} + \ldots.\} \quad .$$

$E(\hat{a}-a)^2$ was originally used in obtaining $Var(\hat{a})$. The expansion is as follows:

$$E(\hat{a}-a)^2 = \frac{(\lambda+\alpha)}{n\alpha^3\lambda}\{\lambda(2\alpha+3)+2\alpha(\alpha+1)\} - \frac{(\alpha+1)(\lambda+\alpha)}{\lambda\alpha^4n^2}\{\lambda^2(\alpha+13)$$

$$+ \lambda\alpha(\alpha+13)+2\alpha^2\} + \frac{(\lambda+\alpha)(\alpha+1)}{\lambda^2\alpha^5n^3}\{\lambda^3(11\alpha+51)$$

$$+ \lambda^2\alpha(12\alpha+62)+ \alpha^2(2\alpha+12)-2\alpha^3\}$$

$$- \frac{(\lambda+\alpha)(\alpha+1)}{\lambda^3\alpha^6n^4}\{\lambda^4(49\alpha+181)+\lambda^3\alpha(61\alpha+241)$$

$$+ \lambda^2\alpha^2(12\alpha+50)-\lambda\alpha^3(2\alpha+10)+4\alpha^4\} \quad .$$

Substituting these expressions into (5-14) yields the result given in Appendix D for $\mu_3(\hat{a})$.

4.2 $\underline{\mu_4(\hat{a})}$

We shall proceed here to show some details in obtaining $\mu_4(\hat{a})$. When we expand $(\hat{a}-a)^4$ and take expectations we have:

$$E(\hat{a}-a)^4 = \lambda^{-4}E\{(Q_2 + \frac{\lambda Q_1}{\alpha})^4 - 4(Q_2 + \frac{\lambda Q_1}{\alpha})^3 Q_1^2$$

$$+ 6(Q_2 + \frac{\lambda Q_1}{\alpha})^2 Q_1^4 - 4(Q_2 + \frac{\lambda Q_1}{\alpha}) Q_1^6 + Q_1^8\}$$

$$\cdot \{1 - \frac{4Q_1}{\lambda} + \frac{10Q_1^2}{\lambda^2} - \frac{20Q_1^3}{\lambda^3} + \frac{35Q_1^4}{\lambda^4} + \ldots .\} \quad .$$

Simplification of the above expression yields:

$$\lambda^4 E(\hat{a}-a)^4 = E\{Q_2^4 + \frac{4Q_1Q_2^3\lambda}{\alpha} + \frac{6Q_1^2Q_2^2\lambda^2}{\alpha^2} + \frac{4Q_1^3Q_2\lambda^3}{\alpha^3} + \frac{Q_1^4\lambda^4}{\alpha^4} - \frac{4Q_1Q_2^4}{\lambda}$$

$$- \frac{4Q_1^2Q_2^3(\alpha+4)}{\alpha} - \frac{12\lambda(\alpha+2)Q_1^3Q_2^2}{\alpha^2} - \frac{4(3\alpha+4)Q_1^4Q_2\lambda^2}{\alpha^3}$$

$$- \frac{4\lambda^3(\alpha+1)Q_1^5}{\alpha^4} + \frac{10Q_1^2Q_2^4}{\lambda^2} + \frac{8(2\alpha+5)Q_1^3Q_2^3}{\lambda\alpha}$$

$$+ \frac{6(\alpha^2+8\alpha+10)Q_1^4Q_2^2}{\alpha^2} + \frac{4\lambda(3\alpha^2+12\alpha+10)Q_1^5Q_2}{\alpha^3}$$

$$+ \frac{\lambda^2(6\alpha+10)(\alpha+1)Q_1^6}{\alpha^4} - \frac{20\ Q_1^3Q_2^4}{\lambda^3} - \frac{40(\alpha+2)Q_1^4Q_2^3}{\lambda^2\alpha}$$

$$- \frac{24(\alpha^2+5\alpha+5)Q_1^5Q_2^2}{\lambda\alpha^2} - \frac{4(\alpha^3+12\alpha^2+30\alpha+20)Q_1^6Q_2}{\alpha^3}$$

$$+ \frac{4\lambda(\alpha^3+6\alpha^2+10\alpha+5)Q_1^7}{\alpha^4} + \frac{35Q_1^4Q_2^4}{\lambda^4} + \frac{(80\alpha+140)Q_1^5Q_2^3}{\lambda^3\alpha}$$

$$+ \frac{30(2\alpha^2+8\alpha+7)Q_1^6Q_2^2}{\lambda^2\alpha^2} + \frac{4(4\alpha^3+30\alpha^2+60\alpha+35)Q_1^7Q_2}{\lambda\alpha^3}$$

$$+ \frac{(\alpha^4+16\alpha^3+60\alpha^2+80\alpha+35)Q_1^8}{\alpha^4}\} \quad .$$

As was demonstrated in the case of $\mu_3(\hat{a})$, we then refer to Appendix B to obtain the expected Q's and simplify. If we write

$$E(\hat{a}-a)^4 = E_2^{(4)}/n^2 + E_3^{(4)}/n^3 + E_4^{(4)}/n^4 + \ldots \quad ,$$

then the $E_k^{(4)}$'s are as follows:

$$E_2^{(4)} = \frac{3(\lambda+\alpha)^2}{\alpha^6}\{\lambda(2\alpha+3)+2\alpha(\alpha+1)\}^2 ,$$

$$\begin{aligned}E_3^{(4)} = \frac{\lambda+\alpha}{\lambda^2\alpha^7}\{&(4\alpha^3+390\alpha^2+1768\alpha+1388)\lambda^5+(12\alpha^3+1080\alpha^2\\&+4800\alpha+3738)\lambda^4\alpha+(12\alpha^3+1046\alpha^2+4648\alpha+3615)\lambda^3\alpha^2\\&+\lambda^2\alpha^3(4\alpha^3+412\alpha^2+1904\alpha+1496)+(56\alpha^2+296\alpha+240)\lambda\alpha^4\\&+(8\alpha+8)\alpha^5\} ,\end{aligned}$$

$$\begin{aligned}E_4^{(4)} = \frac{(\lambda+\alpha)(\alpha+1)}{\lambda^3\alpha^8}\{&\lambda^6(-15\alpha^3-1921\alpha^2-19595\alpha-37121)\\&+\lambda^5\alpha(-45\alpha^3-5751\alpha^2-59025\alpha-112623)+\lambda^4\alpha^2(-45\alpha^3\\&-6135\alpha^2-65007\alpha-126273)+\lambda^3\alpha^3(-15\alpha^3-2701\alpha^2-31305\alpha\\&-63131)+\lambda^2\alpha^4(-396\alpha^2-5988\alpha-13056)+\lambda\alpha^5(-260\alpha-676)\\&+16\alpha^6\} .\end{aligned}$$

The correction to obtain the fourth central moment is:

$$\begin{aligned}\mu_4(\hat{a}) &= E[(\hat{a}-a) - E(\hat{a}-a)]^4\\&= E(\hat{a}-a)^4 - 4E(\hat{a}-a)^3 \cdot E(\hat{a}-a) + 6E(\hat{a}-a)^2\\&\quad\cdot [E(\hat{a}-a)]^2 - 3[E(\hat{a}-a)]^4 . \qquad (5\text{-}15)\end{aligned}$$

The expansions for these terms are substituted into (5-15) and, after considerable simplification, the results are as given in Appendix D.

(c) Fisher's Form for the Negative Binomial Distribution

We shall consider now Fisher's form of the negative binomial distribution with parameters p and k. From Table 1 one can see that the moment estimators are as follows:

$$\hat{p} \text{ (Fisher)} = \hat{a} \text{ (Evans)}$$

$$\hat{k} \text{ (Fisher)} = \hat{\alpha} \text{ (Anscombe)} \quad .$$

Thus the remarks concerning the meaning and variance of $\hat{\alpha}$ and $\hat{a}$ (found in sections a.1, a.2, and b.1 of this chapter) apply also to the moment estimators of Fisher's form of the distribution.

Fisher [13] points out that these estimators are inefficient in some areas of the (p,k) space. However, he says that for $p < 1/9$ and for any value of k, the asymptotic efficiency is 90% or more. Since $p = \lambda/\alpha$, it can be seen that this would exactly correspond to the area in the (λ,α) space in which one is completely inaccurate by using only the usual asymptotic variance of $\hat{\alpha} = \hat{k}$ for any reasonable sample size. Thus the region that gives highest asymptotic efficiency is the region where the variance of the estimator is not only quite large, but in some instances cannot be approximated by an expansion through terms in n^{-4}.

Fisher gives a practical example in which he is considering a sample of 60 sheep classified according to the number of ticks found on each. For this example, $\hat{p} = 0.821382$ and $\hat{k} = 3.956746$. Fisher finds the asymptotic efficiency to be 89.37%. One would certainly not consider the moment estimators in this case to be overly inefficient and might use the estimator based on the asymptotic efficiency. In fact, however, the asymptotic variance is in error, for this case, by more than 95%. Actually it cannot be said how much the asymptotic variance is in error, since for n = 60 the expression (5-6) for Var $\hat{\alpha}$ = Var $\hat{k}$ cannot safely be used.

1. Covariance Determinant

Let us denote the covariance determinant for the moment estimators of Fisher's p and k by

$$D_F = \begin{vmatrix} \text{Var } \hat{p}, & \text{Cov}(\hat{p},\hat{k}) \\ \text{Cov}(\hat{p},\hat{k}), & \text{Var } \hat{k} \end{vmatrix} .$$

The expression through terms in n^{-3} is as follows:

$$D_F = \frac{2(\lambda+\alpha)^3(\alpha+1)}{n^2\alpha^2\lambda} + \frac{(\lambda+\alpha)^3(\alpha+1)}{n^3\lambda^4\alpha^3} \sum_{i=0}^{3} c_i\lambda^i\alpha^{3-i} + \ldots , \tag{5-16}$$

where

$$C_3 = 16\alpha^2 + 54\alpha - 22$$

$$C_2 = 48\alpha^2 + 140\alpha + 48$$

$$C_1 = 48\alpha^2 + 116\alpha + 60$$

$$C_0 = 16\alpha^2 + 32\alpha + 16 \quad .$$

An idea of the behavior of this expression is at least partially explained by Figure 7, which contains the sample size required to make the ratio of the second term to the first term approximately $\sqrt[3]{1/20}$. Note from this figure that in the case of Fisher's example with the sheep, it would be almost impossible to evaluate the error in the covariance determinant by using only the first term in this expression since the sample size used was so small ($n = 60$). However, it would certainly be safe to say that the error involved would be more than 100%. Of course, a true assessment of how inaccurate the asymptotic efficiency really is depends on the corresponding behavior of the covariance determinant for the maximum likelihood estimators. Since the latter determinant was not worked out, one can only speculate on this point. There is every indication that the maximum likelihood covariance form will be very complicated. Nevertheless, even though the behavior of the maximum

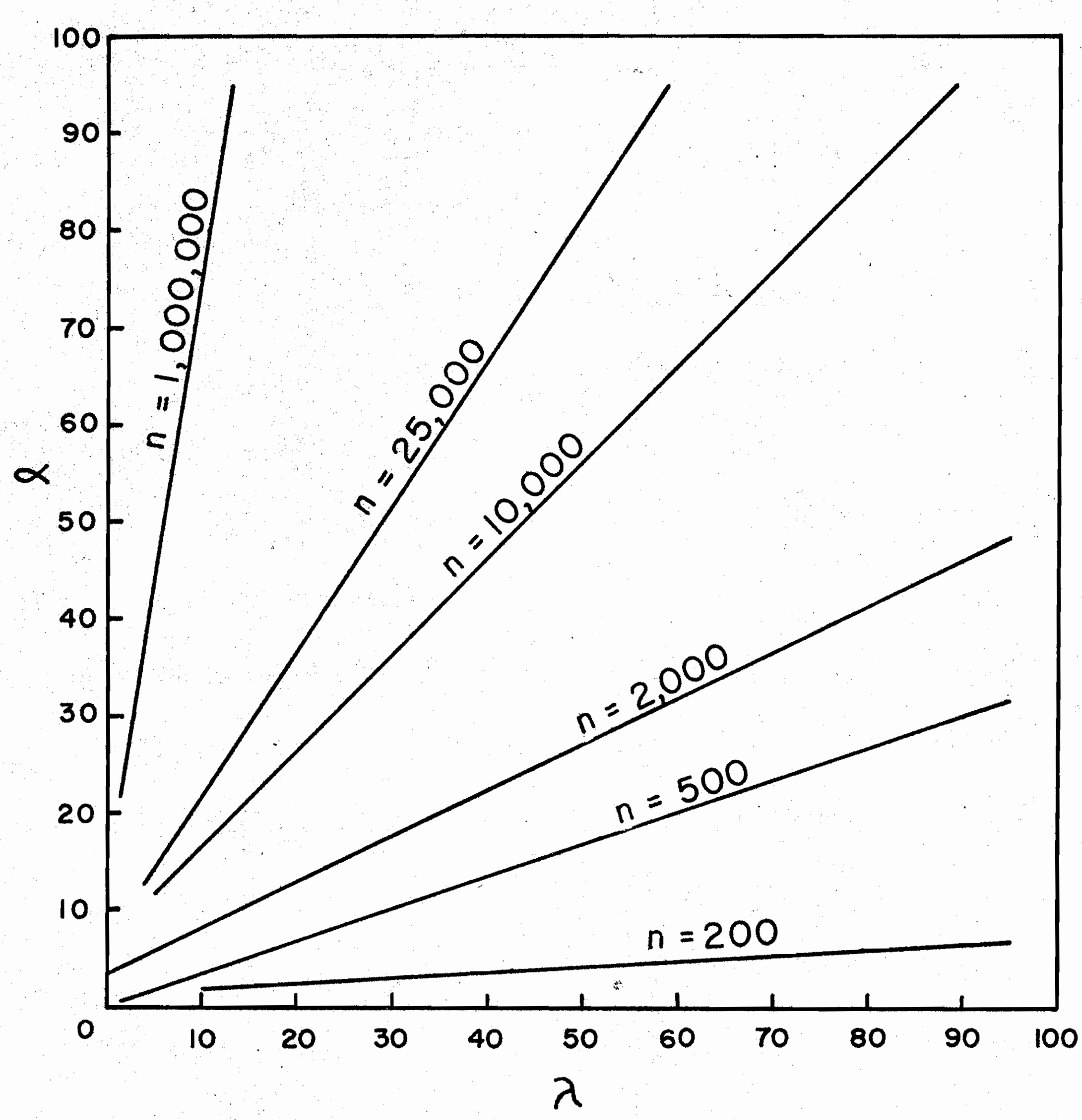

Figure 7. Covariance determinant for the moment estimators of (p,k) in the notation of R. A. Fisher.

Covariance determinant $= A_2/n^2 + A_3/n^3 + \ldots$.

Sample size n to make $(A_3/n^3)/(A_2/n^2) = 0.4$ approx.

likelihood form for this determinant is not known, just the knowledge of the large n^{-3} term for the determinant for the moment estimators is enough to generate some question as to the validity of the use of asymptotic efficiency when one has not considered the possibility that the sample size wasn't sufficiently large to make further terms insignificant in the expansion of both covariance determinants.

(d) Verification of Expansions

It was thought that a certain amount of checking was needed on the validity of expansions given in this chapter on the population moments and covariance determinants of the moment estimators discussed. This is partly due to the heavy amount of algebra involved in the derivations of the expansions. At the same time, nothing is known about the complex nature of the expansions beyond the n^{-4} term given, particularly for the case of the estimator $\hat{\alpha}$. Hence, both algebraic errors and errors in judgement are possible here.

Two types of "checks" were used on the expansions, one for $\hat{a}$ and one for $\hat{\alpha}$. The first was a Monte Carlo Simulation study in which 1000 negative binomial samples were generated for a combination of λ, α, and n. The moments and measures of skewness were calculated for $\hat{a}$ using an IBM 1620 and the results compared with those given by the expansions. The results of this study are shown below for $\lambda = \alpha = 1$ and $n = 100$:

<u>Monte Carlo Results</u>

$\mu_2(\hat{a})$	$\mu_3(\hat{a})$	$\mu_4(\hat{a})$	$\beta_1(\hat{a})$	$\beta_2(\hat{a})$
0.1508	0.0597	0.1154	1.04	5.07

<u>Results From Theoretical Expansions</u>

$\mu_2(\hat{a})$	$\mu_3(\hat{a})$	$\mu_4(\hat{a})$	$\beta_1(\hat{a})$	$\beta_2(\hat{a})$
0.1670	0.0651	0.1375	0.911	4.93

A histogram of the Monte Carlo Results was constructed to show the deviation from normality. This can be found in Figure 8. Note the reasonably good agreement between the Monte Carlo results and the results from the expansions of the moments of $\hat{a}$.

The second method certainly has more theoretical appeal and is a more realistic approach for checking the expansions of $\hat{\alpha}$. Due to the rather erratic nature of $\hat{\alpha}$, it was felt that the Monte Carlo procedure would not be quite so successful. The method used involved a linearization of the

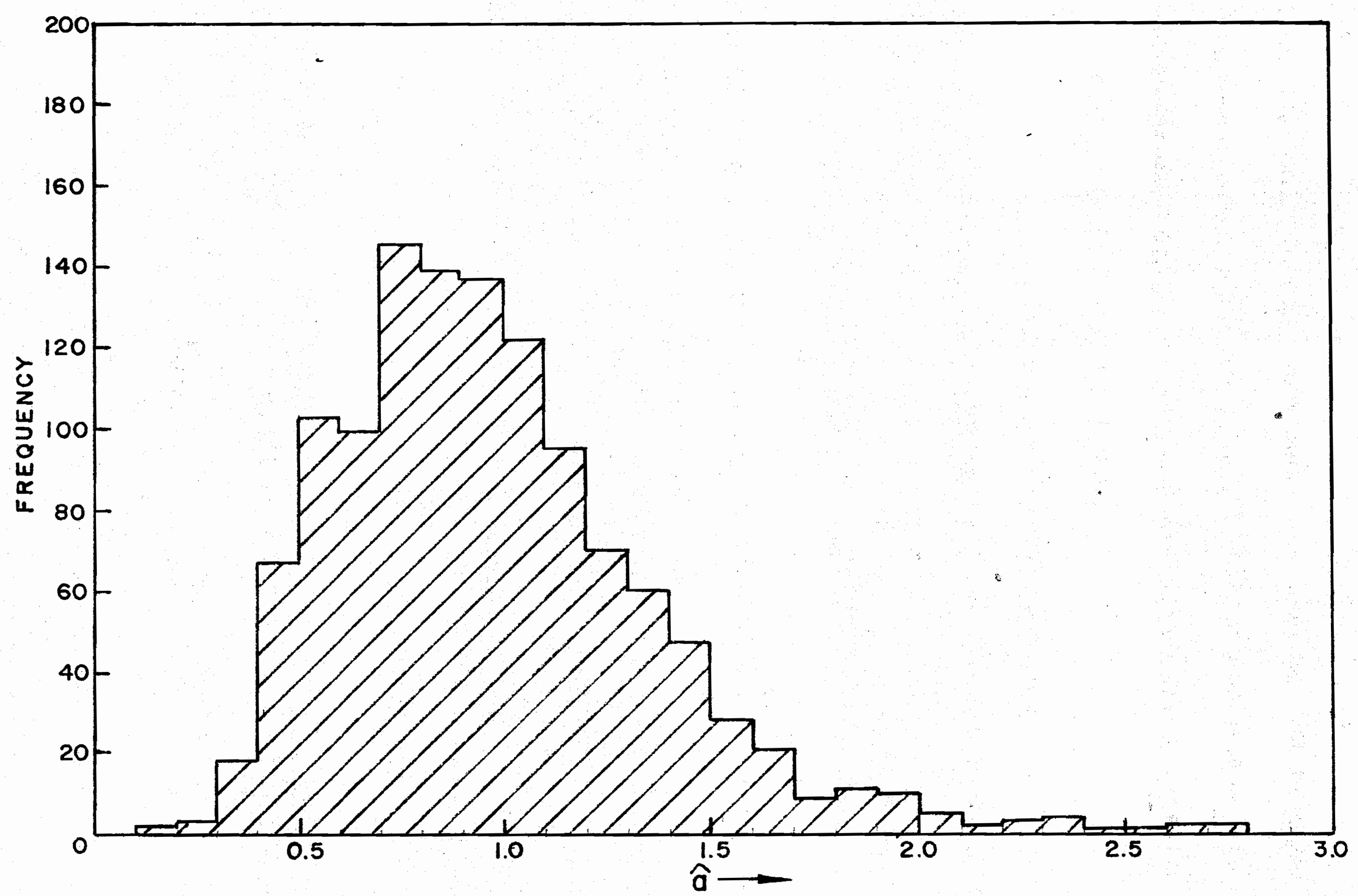

Figure 8. Histogram of frequencies for $\hat{a}$.

statistic $\hat{\alpha} = m_1'^2/(m_2-m_1')$. For examples on the use of this method, see Kendall and Stuart [20].

If we denote $\hat{\alpha}$ by the ratio U/V where $U = m_1'^2$ and $V = m_2-m_1'$, and the upper 5% point of the standardized statistic $\hat{\alpha}$ by $t_{.05}$, then

$$Pr\left(\frac{U/V - \mu_{\hat{\alpha}}}{\sigma_{\hat{\alpha}}} > t_{.05}\right) = 0.05 \quad . \tag{5-17}$$

This t value can be found for particular values of λ, α, and n by using β_1 and β_2 as previously tabulated in Table 3 and entering into the tables of Pearson and Merrington [22]. Thus denoting a particular value of $t_{.05}$ by t, then

$$Pr\left(\frac{U}{V} > t\sigma_{\hat{\alpha}} + \mu_{\hat{\alpha}}\right) = 0.05 \quad ,$$

or

$$Pr\left(\frac{U}{V} > \tau\right) = 0.05 \quad . \tag{5-18}$$

τ, of course, can be calculated using the approximate mean and variance from the expansions. Let us define a new variable

$$Z = U - V\tau \quad .$$

Then ignoring samples for which $m_2 \leq m_1'$, (5-18) can be written

$$Pr(Z > 0) = 0.05 \quad , \tag{5-19}$$

where Z is the statistic

$$m_1'^2 - (m_2 - m_1')\tau \quad .$$

τ is a known constant for a particular case, i.e., for any combination of values of λ, α, and n. The moments of the statistic Z can then be found using orthogonal statistics. Standardizing Z, we have (5.19),

$$\Pr\{\frac{Z - \mu_Z}{\sigma_Z} > - \mu_Z/\sigma_Z\} \doteq 0.05 \quad .$$

Using the first four moments of Z, one can then enter the Pearson Tables with $\beta_1(Z) = \mu_3(Z)^2/\mu_2(Z)^3$, and $\beta_2 = \mu_4(Z)/\mu_2(Z)^2$ and find the corresponding 5% deviate for the standardized Z. This should theoretically be close to $-\mu_Z/\sigma_Z$. It is subject only to the degree of approximation in the moments of $\hat{\alpha}$ and the slight inaccuracy in the attainment of the Pearson deviates. (Accuracy of this four-moment representation is discussed in a recent paper by E. S. Pearson.[1])

Some of the percentage points on the lower side of the distribution of $\hat{\alpha}$ were used in a similar check. The linearization process for the lower percentage points would be

---

[1] E. S. Pearson, Biometrika, 50, (1963). Some problems arising in approximating to probability distributions using moments.

similar to that explained above. We now define $Y = U - V\tau^*$, where τ^* is as follows:

$$\tau^* = t^* \sigma_{\hat{\alpha}} + \mu_{\hat{\alpha}} \quad ,$$

where t^* is the lower-tail Pearson deviate of $\hat{\alpha}$. Thus for a particular τ^*,

$$\Pr(\frac{U}{V} < \tau^*) = 0.05 \quad ,$$

or

$$\Pr(Y < 0) = 0.05 \quad .$$

Thus here, after standardization, it is obvious that the lower-tail Pearson percentage point for the standardized variable Y theoretically corresponds to $-\mu_Y/\sigma_Y$.

Below is a table which shows the results for some values of λ, α, and "safe" sample sizes of the comparison between the Pearson deviate d_p and the calculated percentage point d_c.

Lower Tail Percentage Points

%	λ	α	n	d_p	d_c
5	1	2	1,000	-1.56	-1.58
1	1	2	1,000	-2.10	-2.14
5	5	5	1,000	-1.68	-1.68
5	1	100	2×10^6	-1.645	-1.645

Upper Tail Percentage Points

%	λ	α	n	d_p	d_c
5	1	2	1,000	1.70	1.68
1	1	2	1,000	2.44	2.44
5	5	5	1,000	1.59	1.60
5	1	100	2×10^6	1.640	1.625
5	2	50	100,000	1.645	1.646
5	5	100	100,000	1.645	1.645

Note the good agreement in the results.

CHAPTER VI

COMPARISON WITH RESULTS OF MAXIMUM LIKELIHOOD ESTIMATORS

The results of the bias, variance, and the determinant of the covariance matrix for the maximum likelihood estimators were obtained by Bowman [7] for the form of the negative binomial distribution by Anscombe. The bias and variance were found through terms in n^{-2} and, correspondingly the covariance determinant through terms in n^{-3}. It was believed that actual objective conclusions about the moment estimators could not be made without pointing out similar properties of the maximum likelihood estimators. Fortunately we have access to properties in the form of expansions beyond the asymptotic properties for both the moment and maximum likelihood estimators.

Let us denote

$$\mathrm{Var}(\alpha^*) = V_1^*/n + V_2^*/n^2 + \dots \quad . \tag{6-1}$$

As we have stated before, α^* is the maximum likelihood estimator for α. Table A.6 contains the terms V_1^*/n and V_2^*/n^2 for a few values of α and λ and $n = 100$. If we compare this table with A.2, which is a similar table for $\hat{\alpha}$, we see that the results are strikingly similar. Note the

Table A.6. Terms in Var(α^*), the Maximum Likelihood Estimator of α for n = 100 (taken from Bowman [7])

α \ λ	1	10	50	100
1	0.12159 0.08178	0.02414 0.00294	0.01764 0.00178	0.01670 0.00167
10	263.4957 5293.9366	8.2941 4.6248	2.85583 0.48866	2.37379 0.32890
50	132,584.28 55,697,288.91	1,828.13 10,177.55	201.36 111.04	112.77 32.6998
100	2,060,335. 3.379134×10^9	24,413 4.64517×10^5	1810.06 2378.46	802.69 442.05

Key to Table A.6.

(a) The first term represents the n^{-1} term in (6-1) for n = 100.

(b) The second term represents the n^{-2} term in (6-1) for n = 100.

extremely large contribution of the n^{-2} term for the case in which $\alpha > \lambda$. This is also true for $\hat{\alpha}$. Moreover the rate of increase and decrease of the terms seem to be similar for the two estimators. For example, it can be seen that for say, $\lambda = 1$ and $\alpha = 10$ the comparative terms are $263.4957 + 5293.9366$ for maximum likelihood and $266.2 + 5364$ for the moment estimator. This similarity remains throughout most of the (λ,α) space but is less in evidence for $\lambda \gg \alpha$.

The results for α^* are further evidence of the possible inadequacy of asymptotic expansions. It also illustrates an extremely interesting point, namely that at least as far as the variance is concerned the moment and maximum likelihood estimators of the parameter α behave in an amazingly similar fashion. The behavior of α^* appears to be just as chaotic as that of $\hat{\alpha}$ when $\alpha \gg \lambda$. Thus, while a comparison of these two estimators via efficiency is meaningful, it might be emphasized that both methods of estimation yield estimators with explosive variances for $\alpha \gg \lambda$.

Table A.7 shows the results of n^{-2} and n^{-3} term in the expansion of the determinant of the covariance matrix of α^* and λ^*, i.e., the expansion of

$$D_A^* = \begin{vmatrix} \text{Var } \alpha^*, & \text{Cov}(\alpha^*,\lambda^*) \\ \text{Cov}(\alpha^*,\lambda^*), & \text{Var } \lambda^* \end{vmatrix} = C_2^*/n^2 + C_3^*/n^3 + \ldots . \quad (6\text{-}2)$$

As in the case of the moment estimators, α^* and λ^* are asymptotically uncorrelated, and hence for terms through n^{-3} this expansion does not actually involve the covariance. A sample size of 100 was considered here. To obtain an idea of the true efficiency of moment estimation of the parameters λ and α, one should compare these results with those of Table A.3 which contains the terms of the covariance determinant for the moment estimators. Notice that as one would expect there is a certain similarity in behavior. Note the large contribution of the n^{-3} term when $\alpha > \lambda$.

To get an idea of the effect of including the higher order term on the efficiency of the estimators $(\hat{\alpha},\hat{\lambda})$, we have constructed Table A.8 which shows a comparison of the asymptotic efficiency (using only the first term in the covariance determinants) with the actual efficiencies obtained by including the n^{-3} term. Various sample sizes were used. It can be seen that although in many cases the usual asymptotic covariance determinants for both the moment and maximum likelihood estimators do not nearly describe the actual determinants, the behavior beyond the n^{-2} term is so

Table A.7. Terms in Covariance Determinant of α^*, the Maximum Likelihood Estimator of α for n = 100 (taken from Bowman [7])

α \ λ	1	10	50	100
1	0.002432 0.001636	0.02655 0.00323	0.44975 0.04530	1.6862 0.1684
10	2.8985 58.2333	1.6588 0.9250	8.5675 1.46599	26.112 3.618
50	1352.4 568,112.0	219.38 1221.31	201.362 111.036	338.297 98.099
100	20809.4 34.129×10^6	2685.42 51097.00	1357.5 1783.9	1605.390 884.099

Key to Table A.7. (a) The first term represents the n^{-2} term in (6-2) for n = 100.

(b) The second term represents the n^{-3} term in (6-2) for n = 100.

Table A.8. Efficiencies of $\hat{\alpha} = m_1'^2/(m_2-m_1')$ for Various Sample Sizes

α \ λ	1	10	50	100
1	*76.00% 82.59% (100) 77.82% (500)	*49.87% 58.46% (100) 51.51% (500)	*42.38% 50.33% (100) 43.88% (500)	*40.92% 48.73% (100) 42.38% (500)
10	*98.89% 98.71% (100) 98.75% (500)	*94.25% 94.78% (100) 94.40% (500)	*90.15% 91.40% (100) 90.43% (500)	*89.17% 90.54% (100) 89.47% (500)
50	*99.95% 99.93% (100) 99.93% (1,000) 99.94% (10,000)	*99.57% 99.45% (100) 99.54% (1,000)	*98.71% 98.81% (100) 98.72% (500)	*98.27% 98.47% (100) 98.32% (500)
100	*99.98% 99.98% (100) 99.98% (1,000,000)	*99.88% 99.85% (100) 99.88% (10,000)	*99.55% 99.56% (100) 99.56% (1,000)	*99.34% 99.40% (100) 99.34% (1,000)

Key to Table A.8. (a) "Starred" entries represent asymptotic efficiencies.

(b) Other entries represent efficiencies using n^{-3} terms in covariance determinants for the sample sizes given in parentheses.

similar for the two estimation procedures that the asymptotic efficiency and the "true" efficiency (explained by using the n^{-3} terms) do not differ appreciably, particularly when $\alpha > \lambda$. Note the striking similarity of the true efficiencies to the asymptotic efficiencies in this region, even when the sample size is changed.

There are philosophical implications here regarding the results on the efficiency of $\hat{\alpha}$. It can certainly be said that _if_ the sample size is large enough to "damp off" terms in the covariance determinants of $\hat{\alpha}$ and α^* beyond the n^{-3} term, then the true efficiency does not differ appreciably from the asymptotic efficiency. However, if n is not large enough, the expansions through n^{-3} do not truly describe the determinants and thus what occurs beyond this term is unknown. Thus one cannot say for sure how inaccurate the asymptotic efficiency is in this case. One might suspect that the behavior of further terms in the expansions for the two estimators are also very similar, but this is mere speculation.

The question of what constitutes a large enough sample size is very simple. In view of the close agreement between Tables A.7 and A.3, a "rule of thumb" would be to use the "safe" sample size given by Figure 6. Thus for a particular (λ,α) combination, if the sample size is at least as great as that described by Figure 6, we would assume the asymptotic efficiency to be accurate. However, it is of interest to note the limitations here. For if $\alpha > \lambda$, only fairly large sample sizes are considered to be "safe". For example, even for $\lambda = \alpha = 10$ (Table A.8), we can safely say that the true efficiency is very close to that given by the asymptotic efficiency, namely about 94.3% if the sample size is greater than about 200. However, for $n = 100$, or say 50, the closeness of the asymptotic efficiency to the true efficiency is questionable.

CHAPTER VII

SUMMARY

We have essentially considered here two problems:
(a) The development and use of orthogonal statistics for finding moments of say, a function $t = f(m_1, m_2, \ldots, m_k)$, where the m's represent the sample moments, and (b) a discussion of the sampling properties of moment estimators of the parameters of the negative binomial distribution with special emphasis on a study of the forms of the distribution due to Anscombe, Evans, and Fisher.

(a) General Remarks on Orthogonal Statistics

The development of the technique of orthogonal statistics originated from the concept of existence of an infinite set $\{q_r(x)\}$ of orthogonal polynomials associated with a particular distribution. The r'th orthogonal statistic Q_r was defined over the sample space as $Q_r = \sum_{j=1}^{n} q_r(x_j)$. From the use of the joint moment generating function of the Q's, tables of expectations of powers and products of these Q's are given in terms of expectations of powers and products of the q's. The expectations involve expansions in ascending powers of 1/n. The tabulations are given generally for powers and products of any four orthogonal statistics and are expanded through terms in n^{-5}. The technique of orthogonal statistics involves expanding powers of the statistic t, i.e., expanding t, t^2, etc. in powers and products of the Q's and taking expectations term by term to obtain the sampling moments.

The emphasis for this study was put on the applicability of these orthogonal statistics to the negative binomial distribution. Tables are given showing the expected values of powers and products of the Q's (using Q_1 and Q_2) in terms of the parameters of the distribution. Examples are given showing the use of the technique and the tables in the case of the negative binomial.

(b) General Remarks on Sampling Properties of Moment Estimators for the Negative Binomial Distribution

We have considered properties of the following joint moment estimators: $\hat{\alpha} = m_1'^2/(m_2-m_1')$ (assuming a non-null sample), $\hat{\lambda} = m_1'$ for Anscombe's form of the distribution; $\hat{a} = (m_2-m_1')/m_1'$, $\hat{m} = m_1'$ for Evans' form of the distribution; and $\hat{p} = (m_2-m_1')/m_1'$, $\hat{k} = m_1'^2/(m_2-m_1')$ for Fisher's form. The general forms of the distribution and other important properties of each are found in Table 1, Chapter III. Properties such as bias, variance, higher moments, and joint efficiency are discussed at length for these estimators in Chapter V.

Chapter VI contains a comparison of the results of the properties discussed for the moment estimators of the form of Anscombe with those of the maximum likelihood estimators. It was noted here that there was a striking similarity in the behavior of the two estimators for some of the higher order terms of the bias and variance. The maximum likelihood results were taken from work done by Bowman [7].

(c) <u>Concluding Remarks on the Moment Estimators for the Negative Binomial Distribution</u>

Some studies at the outset indicated that the distribution of $\hat{\alpha}$ might have moments which are much larger than what is reflected by first term approximations of the bias and variance. Further evidence of this was shown by the tabulation of the individual terms in the expansion of the variance of $\hat{\alpha}$ for $n = 100$. For $\alpha > \lambda$, the values of higher order terms are very large (see Table A.2) and require extremely large sample sizes to "damp off" succeeding terms. Thus for an ordinary sample size, the usual asymptotic variance does not even begin to tell the complete story. In the range where the asymptotic efficiency is high, namely $\alpha > \lambda$, the variance of $\hat{\alpha}$ is chaotic and unless the sample size is exceedingly large, one is completely mislead by using only the approximation given by the first term.

This same explosive property of $\hat{\alpha}$ (and hence Fisher's $\hat{k}$) was reflected in the bias. As with the case of the variance, extremely large sample sizes were required to "calm" the expansion of the bias of $\hat{\alpha}$. Illustrations are given throughout the text which reflect the nature of $\hat{\alpha}$. Other illustrations are shown which give the reader an idea of the inadequacy of first term approximations with expansions of the bias, variance, and covariance determinant of $\hat{\alpha}$.

In the comparison of the results for $\hat{\alpha}$ with the properties of the maximum likelihood estimator α^*, one finds that the expansions of the covariance determinants behave in a very similar fashion beyond the first term, i.e., the n^{-2} term. Hence for $\alpha > \lambda$, although the first term in the expansion of $|\mathrm{Cov}(\hat{\alpha},\hat{\lambda})|$ and $|\mathrm{Cov}(\alpha^*,\lambda^*)|$ in some cases does not begin to give the complete determinant, the ratio of the two determinants expanded for a "safe" sample size is strikingly close to the asymptotic efficiency. Thus when $\alpha > \lambda$ and for very large sample sizes (see Chapter VI for the interpretation of "how large?") the asymptotic efficiency is reliable. However, one cannot be sure of the behavior of further terms in the covariance determinants when n is not large enough to cause these terms to be negligible. This then makes asymptotic efficiency questionable for $\alpha > \lambda$ and for say, a sample size of 100.

The results here concerning $\hat{\alpha}$ represent strong evidence which leads to two practical conclusions; (a) in view of the explosive nature of the bias and variance of $\hat{\alpha}$, some doubt is cast on $\hat{\alpha}$ as an estimator, and (b) one should be cautious toward the interpretation of asymptotic properties in general without further investigation.

The results found for the moment estimators of Evans' form (considering only that portion of the parameter space for which $\alpha > 1$ and $\lambda > 1$) for the negative binomial indicated that the behavior of $\hat{a}$ is in sharp contrast to that of $\hat{\alpha}$. In fact, tables of the properties of $\hat{a}$ for $n = 100$ show extremely small variance and bias in the very range of λ and α in which $\mathrm{Var}(\hat{\alpha})$ is so large. Also one is better able to use asymptotic approximations in the case of $\hat{a}$. For example, a sample size of 30 is large enough to justify the use of asymptotic variance. To show the contrast in the case of $\hat{\alpha}$, for $\alpha \gg \lambda$, for example, when $\lambda = 2$ and $\alpha = 50$, n must be approximately 2,000,000 before one is justified in using the asymptotic variance.

ACKNOWLEDGEMENTS

The author is indebted to Dr. L. R. Shenton of the Department of Statistics for his help and encouragement in directing this problem.

Gratitude is also expressed to Dr. Boyd Harshbarger, Head of the Statistics Department, for his constant interest in this problem.

The author would also like to express his deepest appreciation to Mrs. Lela Barnhisel for her untiring efforts in typing this thesis.

REFERENCES

[1] Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. Biometrika, 37, 358-362.

[2] Anscombe, F. J. (1949). The statistical analysis of insect counts based on the negative binomial distribution. Biometrics, 5, 165-173.

[3] Arbous, A. G. and Kerrich, J. E. (1951). Accident statistics and the concept of accident-proneness. Biometrics, 7, 340-342.

[4] Barankin, E. W. and Gurland, J. (1951). On asymptotically efficient estimators. University of California Publications in Statistics, 1, 89-129.

[5] Bartko, J. J. (1961). The negative binomial distribution: A review of properties and applications. The Virginia Journal of Science, 12(1), 18-37.

[6] Bliss, C. I. and Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. Biometrics, 9(2), 176-200.

[7] Bowman, K. (1963). Higher moments of maximum likelihood estimators. Thesis to be submitted for Ph.D. degree, Virginia Polytechnic Institute.

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

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

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

[11] Feller, W. (1943). On a general class of contagious distributions. _Ann_. _Math_. _Stat_, _14_, 389-400.

[12] Ferguson, T. S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. _Ann_. _Math_. _Stat_, _29_, 1046-1062.

[13] Fisher, R. A. (1941). The negative binomial distribution. _Annals of Eugenics_, _11_, 182-187.

[14] Fisher, R. A. (1929). Moments and product moments of sampling distributions. _Proc_. _Lond_. _Math_. _Soc_, _30_, 199.

[15] Greenwood, M. and Yule, G. V. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. _J_. _Royal Stat_. _Soc_, _83_, 255-279.

[16] Gurland, J. (1957). Some applications of the negative binomial and other contagious distributions. _American Journal of Public Health_, _49_, 1388-1399.

[17] Haldane, J.B.S. and Smith, Shelia Maynard (1956). The sampling distribution of a maximum-likelihood estimate. _Biometrika_, _43_, 96-103.

[18] Katti, S. K. and Gurland, J. (1962). Efficiency of certain methods of estimation for the negative binomial and the Neyman Type A distributions. _Biometrika_, _49_, 215-226.

[19] Kendall, D. B. (1949). _Symposium on Stochastic Processes_, _J_. _Royal Stat_. _Soc_. (Series B), _11_, 242.

[20] Kendall, M. G. and Stuart, A. (1958). _The Advanced Theory of Statistics_. Charles Griffin: London.

[21] Kendall, M. G. (1940). The derivation of multivariate sampling formulae from univariate formulae by symbolic operation. Annals of Eugenics, 10, 392.

[22] Pearson, E. S. and Merrington, M. (1951). Tables of the 5% and 0.5% points of Pearson curves. Biometrika, 38, 4.

[23] Pearson, E. S. (1930). A further development of tests for normality. Biometrika, 22, 239-249.

[24] 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.

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

[26] Shenton, L. R. (1949). On the efficiency of the method of moments and Neyman's Type A distribution. Biometrika, 36, 450-454.

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

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

VITA

Raymond H. Myers was born October 13, 1937, in Charleston, West Virginia. He attended Stonewall Jackson High School in Charleston where he was graduated in June 1955. In September 1955, he enrolled at Virginia Polytechnic Institute in the Department of Chemical Engineering. While in undergraduate school he did part-time work with Union Carbide Chemical Company in South Charleston, West Virginia. In June 1959, he completed the requirements for a Bachelor of Science degree in Chemical Engineering.

In September 1959, the author enrolled in graduate school at Virginia Polytechnic Institute in the Department of Statistics while holding the position of graduate assistant. In the summer of 1960, he was employed as a statistician for Hercules Powder Company in Radford, Virginia. In June 1961, he received his Master of Science degree and immediately began work toward his Ph.D. degree. He joined the faculty of the Virginia Polytechnic Institute in September 1962, as Assistant Professor of Statistics and continued studying toward a degree.

He was elected into Phi Lambda Upsilon Chemical Honorary, Pi Mu Epsilon for Mathematics, Phi Kappa Phi, and Sigma Xi. He is a member of the American Institute of Chemical Engineers, Virginia Academy of Science, and American Statistical Association.

The author is married, and he and his wife Patsy have one child, Billy, born June 1, 1963.

Raymond Harold Myers

A P P E N D I C E S

APPENDIX A

Expectation of Q-Products (through terms in n^{-5} for any four orthogonal statistics)

Products of two (n^{-1} only)

$(r^2)_1 = [r^2]$

$(rs)_1 = 0$

Products of three (n^{-2} only)

$(r^3)_2 = [r^3]$

$(r^2s)_2 = [r^2s]$

$(rst)_2 = [rst]$

Products of four (n^{-2} to n^{-3})

$(r^4)_2 = 3[r^2]^2$

$(r^3s)_2 = 0$

$(r^2st)_2 = 0$

$(rstu)_2 = 0$

$(r^2s^2)_2 = [r^2][s^2]$

* Note how n^{-3} terms are found

Products of five (n^{-3} to n^{-4})

$(r^5)_3 = 10[r^2][r^3]$

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

$(r^3s^2)_3 = [r^3][s^2]+3[r^2][rs^2]$

$(r^3st)_3 = 3[r^2][rst]$

$(r^2stu)_3 = [r^2][stu]$

$(r^2s^2t)_3 = [r^2t][s^2]+[r^2][s^2t]$

* Note how n^{-4} terms are found

<u>Products of six</u> (n^{-3} to n^{-5})

$$(r^6) = 15[r^2]^3/n^3+(10[r^3]^2+15[r^2][r^4]-45[r^2]^3)/n^4 +(r^6)_5/n^5$$

$$(r^5s) = (10[r^3][r^2s]+10[r^3s][r^2])/n^4+(r^5s)_5/n^5$$

$$(r^4st) = (6[r^2s][r^2t]+4[r^3][rst]+6[r^2][r^2st])/n^4+(r^4st)_5/n^5$$

$$(r^4s^2) = 3[r^2]^2[s^2]/n^3+(6[r^2s]^2+4[r^3][rs^2]+[r^4][s^2] +6[r^2][r^2s^2]-9[r^2]^2[s^2])/n^4+(r^4s^2)_5/n^5$$

$$(r^3stu) = (3[rtu][r^2s]+3[rsu][r^2t]+3[rst][r^2u]+[r^3][stu] +3[r^2][rstu])/n^4+(r^3stu)_5/n^5$$

$$(r^3s^3) = (9[rs^2][r^2s]+[r^3][s^3]+3[s^2][r^3s]+3[r^2][rs^3])/n^4 +(r^3s^3)_5/n^5$$

$$(r^2s^2t^2) = [r^2][s^2][t^2]/n^3+(4[rst]^2+2[r^2s][st^2]+2[rt^2][rs^2] +2[r^2t][s^2t]+[r^2t^2][s^2]+[r^2s^2][t^2]+[r^2][s^2t^2] -3[r^2][s^2][t^2])/n^4+(r^2s^2t^2)_5/n^5$$

$$(r^2s^2tu) = (4[rst][rsu]+2[r^2s][stu]+2[rs^2][rtu]+[r^2u][s^2t] +[r^2t][s^2u]+[r^2tu][s^2]+[r^2][s^2tu])/n^4 +(r^2s^2tu)_5/n^5$$

* Note how n^{-5} terms are found

<u>Products of seven</u> <u>Order</u> n^{-4}

$$(r^7)_4 = 105[r^2]^2[r^3]$$

$$(r^6s)_4 = 45[r^2]^2[r^2s]$$

$$(r^5st)_4 = 15[r^2]^2[rst]$$

$$(r^5s^2)_4 = 10[r^2][s^2][r^3]+15[r^2]^2[rs^2]$$

$$(r^4stu)_4 = 3[r^2]^2[stu]$$

$$(r^4s^2t)_4 = 6[r^2][s^2][r^2t]+3[r^2]^2[s^2t]$$

$$(r^4s^3)_4 = 18[r^2][s^2][r^2s]+3[r^2]^2[s^3]$$

$$(r^3s^2tu)_4 = 3[r^2][s^2][rtu]$$

$$(r^3s^2t^2)_4 = [r^3][s^2][t^2]+3[r^2][s^2][rt^2]+3[r^2][t^2][rs^2]$$

$$(r^3s^3t)_4 = 9[r^2][s^2][rst]$$

$$(r^2s^2t^2u)_4 = [r^2u][s^2][t^2]+[t^2u][r^2][s^2]+[s^2u][t^2][r^2]$$

<u>Products of seven</u> <u>Order</u> n^{-5}

$$(r^7)_5 = 21[r^2][r^5]+35[r^3][r^4]-315[r^2]^2[r^3]$$

$$(r^6s)_5 = 15[r^2][r^4s]+15[r^2s][r^4]+20[r^3][r^3s]-135[r^2]^2[r^2s]$$

$$(r^5st)_5 = 10[r^2][r^3st]+5[rst][r^4]+10[r^2s][r^3t]$$
$$+10[r^2t][r^3s]+10[r^3][r^2st]-45[r^2]^2[rst]$$

$$(r^5s^2)_5 = 10[r^2][r^3s^2]+[r^5][s^2]+10[r^3][r^2s^2]+20[r^2s][r^3s]$$
$$+5[rs^2][r^4]-45[r^2]^2[rs^2]-30[r^2][s^2][r^3]$$

$$(r^4stu)_5 = 6[r^2][r^2stu]+[r^4][stu]+4[r^3u][rst]+4[r^3t][rsu]$$
$$+6[r^2s][r^2tu]+4[r^3s][rtu]+6[r^2su][r^2t]$$
$$+6[r^2st][r^2u]+4[r^3][rstu]-9[r^2]^2[stu]$$

$$(r^4s^2t)_5 = [r^4t][s^2]+6[r^2][r^2s^2t]+[r^4][s^2t]+4[rs^2][r^3t]$$
$$+8[rst][r^3s]+12[r^2s][r^2st]+6[r^2t][r^2s^2]$$
$$+4[r^3][rs^2t]-18[r^2][s^2][r^2t]-9[r^2]^2[s^2t]$$

$$(r^4s^3)_5 = 3[r^4s][s^2]+6[r^2][r^2s^3]+[r^4][s^3]+12[rs^2][r^3s]$$
$$+18[r^2s][r^2s^2]+4[r^3][rs^3]-54[r^2][s^2][r^2s]$$
$$-9[r^2]^2[s^3]$$

$$(r^3s^2tu)_5 = [r^3tu][s^2]+3[r^2][rs^2tu]+2[r^3s][stu]+[r^3u][s^2t]$$
$$+6[rst][r^2su]+[r^3t][s^2u]+6[rsu][r^2st]$$
$$+3[rs^2][r^2tu]+6[r^2s][rstu]+3[rtu][r^2s^2]$$
$$+3[r^2t][rs^2u]+3[r^2u][rs^2t]+[r^3][s^2tu]$$
$$-9[r^2][s^2][rtu]$$

$$(r^3s^2t^2)_5 = [r^3t^2][s^2]+[r^3s^2][t^2]+3[r^2][rs^2t^2]+2[r^3s][st^2]$$
$$+2[r^3t][s^2t]+12[rst][r^2st]+3[rs^2][r^2t^2]$$
$$+3[rt^2][r^2s^2]+6[r^2t][rs^2t]+6[r^2s][rst^2]$$
$$+[r^3][s^2t^2]-3[r^3][s^2][t^2]-9[r^2][s^2][rt^2]$$
$$-9[r^2][t^2][rs^2]$$

$$(r^3s^3t)_5 = 3[r^3st][s^2]+3[r^2][rs^3t]+3[r^3s][s^2t]+[r^3t][s^3]$$
$$+3[r^2t][rs^3]+9[rst][r^2s^2]+9[rs^2][r^2st]$$
$$+9[r^2s][rs^2t]+[r^3][s^3t]-27[r^2][s^2][rst]$$

$$(r^2s^2t^2u)_5 = [r^2][s^2t^2u]+[s^2][r^2t^2u]+[t^2][r^2s^2u]+4[rsu][rst^2]$$
$$+4[r^2st][stu]+4[rtu][rs^2t]+8[rst][rstu]+2[r^2s][st^2u]$$
$$+2[rs^2][rt^2u]+2[s^2t][r^2tu]+2[st^2][r^2su]+2[r^2t][s^2tu]$$
$$+[r^2t^2][s^2u]+[r^2s^2][t^2u]+[s^2t^2][r^2r]-3[r^2][s^2][t^2u]$$
$$-3[r^2][t^2][s^2u]-3[s^2][t^2][r^2u]+2[rt^2][rs^2u]$$

<u>Products of eight</u> <u>Order</u> n^{-4}

<u>These are zero with the exception of</u>

$(r^8)_4 = 105[r^2]^4$ $(r^6s^2)_4 = 15[r^2]^3[s^2]$

$(r^4s^2t^2)_4 = 3[r^2]^2[s^2][t^2]$ $(r^4s^4)_4 = 9[r^2]^2[s^2]^2$

$(r^2s^2t^2u^2)_4 = [r^2][s^2][t^2][u^2]$

<u>Products of eight</u> <u>Order</u> n^{-5}

$$(r^8)_5 = 280[r^3]^2[r^2]+210[r^2]^2[r^4]-630[r^2]^4$$

$$(r^7s)_5 = 210[r^3][r^2s][r^2]+105[r^2]^2[r^3s]$$

$$(r^6st)_5 = 90[r^2t][r^2s][r^2]+60[r^3][r^2][rst]+45[r^2]^2[r^2st]$$

$$(r^6s^2)_5 = 90[r^2s]^2[r^2]+60[r^3][rs^2][r^2]+10[r^3]^2[s^2]$$
$$+15[r^2][s^2][r^4]+45[r^2]^2[r^2s^2]-90[r^2]^3[s^2]$$

$$(r^5stu)_5 = 30[r^2s][rtu][r^2]+30[r^2u][rst][r^2]+30[r^2t][rsu][r^2]$$
$$+10[r^3][stu][r^2]+15[r^2]^2[rstu]$$

$$(r^5s^2t)_5 = 60[rst][r^2s][r^2]+30[r^2t][rs^2][r^2]+10[r^3][r^2t][s^2]$$
$$+15[r^2]^2[rs^2t]+10[r^3][s^2t][r^2]+10[r^2][s^2][r^3t]$$

$$(r^5s^3)_5 = 15[r^2]^2[rs^3]+30[r^2][s^2][r^3s]+90[r^2s][rs^2][r^2]$$
$$+30[r^2s][r^3][s^2]+10[r^2][r^3][s^3]$$

$$(r^4s^2tu)_5 = 3[r^2]^2[s^2tu]+6[r^2][s^2][r^2tu]+6[r^2][r^2t][s^2u]$$
$$+6[r^2][r^2u][s^2t]+24[r^2][rst][rsu]+12[r^2][rs^2][rtu]$$
$$+4[r^3][s^2][rtu]+6[s^2][r^2t][r^2u]+12[r^2][r^2s][stu]$$

$$\begin{aligned}(r^4s^3t)_5 = {} & 3[r^2]^2[s^3t]+18[r^2s][r^2t][s^2]+36[rs^2][rst][r^2]\\ & +18[r^2][s^2][r^2st]+6[r^2t][r^2][s^3]\\ & +12[r^3][rst][s^2]+18[r^2s][s^2t][r^2]\end{aligned}$$

$$\begin{aligned}(r^4s^4)_5 = {} & 3[r^2]^2[s^4]+3[r^4][s^2]^2+36[r^2s]^2[s^2]+36[rs^2]^2[r^2]\\ & +36[r^2][r^2s^2][s^2]+24[r^2s][r^2][s^3]\\ & +24[r^3][rs^2][s^2]-54[r^2]^2[s^2]^2\end{aligned}$$

$$\begin{aligned}(r^3s^3tu)_5 = {} & 9[r^2t][rsu][s^2]+9[r^2s][rtu][s^2]+9[r^2][rst][s^2u]\\ & +9[r^2][rs^2][stu]+9[r^2][rstu][s^2]+3[r^2][rtu][s^3]\\ & +9[r^2u][rst][s^2]+3[r^3][stu][s^2]+9[r^2][rsu][s^2t]\end{aligned}$$

$$\begin{aligned}(r^3s^3r^2)_5 = {} & 18[r^2t][rst][s^2]+9[rt^2][r^2s][s^2]+18[r^2][rst][s^2t]\\ & +9[r^2][rs^2][st^2]+9[r^2][rst^2][s^2]+3[r^2][rt^2][s^3]\\ & +3[r^3][s^2][st^2]+3[r^2][rs^3][t^2]+3[r^3s][s^2][t^2]\\ & +[r^3][s^3][t^2]+9[rs^2][r^2s][t^2]\end{aligned}$$

$$\begin{aligned}(r^4s^2t^2)_5 = {} & 12[r^2][r^2s][st^2]+24[r^2][rst]^2+12[r^2][r^2t][s^2t]\\ & +4[r^3][rs^2][t^2]+6[r^2s]^2[t^2]+12[r^2][rt^2][rs^2]\\ & +6[r^2t]^2[s^2]+4[r^3][rt^2][s^2]+[r^2][s^2][t^2]\\ & +6[r^2][s^2][r^2t^2]+6[r^2][r^2s^2][t^2]+3[r^2]^2[s^2t^2]\\ & -18[r^2]^2[s^2][t^2]\end{aligned}$$

$$\begin{aligned}(r^3s^2t^2u)_5 = {} & 6[r^2t][rtu][s^2]+12[r^2][rst][stu]+3[r^2][rs^2u][t^2]\\ & +3[r^2u][rt^2][s^2]+6[r^2][rsu][st^2]+[r^3][t^2u][s^2]\\ & +6[r^2s][rsu][t^2]+[r^3u][s^2][t^2]+3[r^2][rs^2][t^2u]\\ & +3[r^2][rt^2][s^2u]+3[r^2][rt^2u][s^2]+6[r^2][rtu][s^2t]\\ & +3[r^2u][rs^2][t^2]+[r^3][s^2u][t^2]\end{aligned}$$

$$(r^2s^2t^2u^2)_5 = [r^2][s^2][t^2u^2]+[r^2][t^2][s^2u^2]+[r^2][u^2][s^2t^2]$$
$$+[s^2][t^2][r^2u^2]+[s^2][u^2][r^2t^2]+[t^2][u^2][r^2s^2]$$
$$+4[r^2][stu]^2+4[s^2][rtu]^2+4[t^2][rsu]^2+4[u^2][rst]^2$$
$$+2[r^2]\{[st^2][su^2]+[tu^2][ts^2]+[ut^2][us^2]\}$$
$$+2[s^2]\{[ru^2][rt^2]+[tu^2][tr^2]+[ur^2][ut^2]\}$$
$$+2[t^2]\{[rs^2][ru^2]+[sr^2][su^2]+[ur^2][us^2]\}$$
$$+2[u^2]\{[rs^2][rt^2]+[st^2][sr^2]+[ts^2][tr^2]\}$$
$$-6[r^2][s^2][t^2][u^2]$$

<u>Products of nine</u> <u>Order</u> n^{-5}

$$(r^9)_5 = 1260[r^2]^3[r^3]$$

$$(r^8s)_5 = 420[r^2]^3[r^2s]$$

$$(r^7st)_5 = 105[r^2]^3[rst]$$

$$(r^7s^2)_5 = 105[r^2]^2[r^3][s^2]+105[r^2]^3[rs^2]$$

$$(r^6stu)_5 = 15[r^2]^3[stu]$$

$$(r^6s^2t)_5 = 45[r^2]^2[r^2t][s^2]+15[r^2]^3[s^2t]$$

$$(r^6s^3)_5 = 135[r^2]^2[r^2s][s^2]+15[r^2]^3[s^3]$$

$$(r^5s^2tu)_5 = 15[r^2]^2[rtu][s^2]$$

$$(r^5s^3t)_5 = 45[r^2]^2[rst][s^2]$$

$$(r^5s^4)_5 = 90[r^2]^2[rs^2][s^2]+30[r^2][r^3][s^2]^2$$

$$(r^5s^2t^2)_5 = 15[r^2]^2[rt^2][s^2]+15[r^2]^2[rs^2][t^2]$$
$$+10[r^2][r^3][s^2][t^2]$$

$$(r^4s^4t)_5 = 18[r^2]^2[s^2t][s^2]+18[r^2][r^2t][s^2]^2$$

$$(r^4s^3tu)_5 = 9[r^2]^2[stu][s^2]$$

$$(r^4s^3t^2)_5 = 9[r^2]^2[st^2][s^2]+3[r^2]^2[s^3][t^2] +18[r^2][s^2][t^2][r^2s]$$

$$(r^4s^2t^2u)_5 = 3[r^2]^2[s^2][t^2u]+3[r^2]^2[s^2u][t^2] +6[r^2][s^2][t^2][r^2u]$$

$$(r^3s^3t^3)_5 = 27[r^2][s^2][t^2][rst]$$

$$(r^3s^3t^2u)_5 = 9[r^2][s^2][t^2][rsu]$$

$$(r^3s^2t^2u^2)_5 = 3[r^2][s^2][t^2][ru^2]+3[r^2][t^2][u^2][rs^2] +3[r^2][s^2][u^2][rt^2]$$

__Products of ten__ __Order__ n^{-5}

__These are zero with the exception of__

$$(r^{10})_5 = 945[r^2]^5$$

$$(r^8s^2)_5 = 105[r^2]^4[s^2]$$

$$(r^6s^4)_5 = 45[r^2]^3[s^2]^2$$

$$(r^6s^2t^2)_5 = 15[r^2]^3[s^2][t^2]$$

$$(r^4s^4t^2)_5 = 9[r^2]^2[s^2]^2[t^2]$$

$$(r^4s^2t^2u^2)_5 = 3[r^2]^2[s^2][t^2][u^2]$$

---

*The highest order term in products of four, five, and six is found by:

$$(r^\alpha s^\beta t^\gamma \ldots)_b = [r^\alpha s^\beta t^\gamma \ldots] - \sum_{\lambda=a}^{b-1} (r^\alpha s^\beta t^\gamma \ldots)_\lambda \quad .$$

APPENDIX B

Table of Expected Values of Q-Products for the Negative Binomial Distribution in Terms of λ and α

Notation

$\lambda + \alpha = \chi$

$2\lambda + \alpha = \psi$

$\alpha + 1 = a$

$\alpha + 2 = b$

Products of order 2

$(1^2)_1 = \lambda\chi/\alpha$ $\quad$ $(2^2)_1 = 2\lambda^2\chi^2 a/\alpha^3$

Products of order 3

$(1^3)_2 = \lambda\chi\psi/\alpha^2$

$(1^2 2)_2 = 2\lambda^2\chi^2 a/\alpha^3$

$(12^2)_2 = 4\lambda^2\chi^2\psi a/\alpha^4$

$(2^3)_2 = 4(2\lambda\chi b+\psi^2)\lambda^2\chi^2 a/\alpha^5$

Products of order 4

$(1^4)_2 = 3\lambda^2\chi^2/\alpha^2$ $\quad$ $(1^3 2)_2 = 0$

$(1^2 2^2)_2 = 2\lambda^3\chi^3 a/\alpha^4$ $\quad$ $(12^3)_2 = 0$

$(2^4)_2 = 12\lambda^4\chi^4 a^2/\alpha^6$

$$(1^4)_3 = \lambda\chi(\alpha^2+6\lambda\alpha+6\lambda^2)/\alpha^3$$

$$(1^3 2)_3 = 6\psi\lambda^2\chi^2 a/\alpha^4$$

$$(1^2 2^2)_3 = 8\lambda^2\chi^2 a(\lambda\chi b+\psi^2)/\alpha^5$$

$$(12^3)_3 = 8\lambda^2\chi^2\psi a\{\psi^2+\lambda\chi(6\alpha+11)\}/\alpha^5$$

$$(2^4)_3 = 8a\lambda^2\chi^2\{(6\alpha^2+30\alpha+34)\lambda^2\chi^2+(18\alpha+34)\lambda\chi\psi^2+\psi^4\}/\alpha^7$$

<u>Products of order 5</u>

$$(1^5)_3 = 10\lambda^2\chi^2\psi/\alpha^3$$

$$(1^4 2)_3 = 12\lambda^3\chi^3 a/\alpha^4$$

$$(1^3 2^2)_3 = 14\lambda^3\chi^3\psi a/\alpha^5$$

$$(1^2 2^3)_3 = 4\lambda^3\chi^3 a\{(5\alpha+7)\lambda\chi+\psi^2\}/\alpha^6$$

$$(12^4)_3 = 48\lambda^4\chi^4\psi a^2/\alpha^7$$

$$(2^5)_3 = 80\lambda^4\chi^4 a^2\{2\lambda\chi b+\psi^2\}/\alpha^8$$

$$(1^5)_4 = \lambda\chi\psi(8\lambda\chi+\psi^2)/\alpha^4$$

$$(1^4 2)_4 = 2\lambda^2\chi^2 a(8\lambda\chi+7\psi^2)/\alpha^5$$

$$(1^3 2^2)_4 = 8\lambda^2\chi^2\psi a\{(6\alpha+13)\lambda\chi+2\psi^2\}/\alpha^6$$

$$(1^2 2^3)_4 = 8\lambda^2\chi^2 a\{(6\alpha^2+30\alpha+34)\lambda^2\chi^2+(24\alpha+45)\lambda\chi\psi^2+2\psi^4\}/\alpha^7$$

$$(12^4)_4 = 16\lambda^2\chi^2 a\psi\{(36\alpha^2+162\alpha+180)\lambda^2\chi^2+(31\alpha+59)\lambda\chi\psi^2+\psi^4\}/\alpha^8$$

$$(2^5)_4 = 16\lambda^2\chi^2 a\{b(24\alpha^2+168\alpha+248)\lambda^3\chi^3$$
$$+(210\alpha^2+960\alpha+1074)\lambda^2\chi^2\psi^2+(70\alpha+136)\lambda\chi\psi^4+\psi^6\}/\alpha^9$$

<u>Products of order 6</u>

$(1^6)_3 = 15\lambda^3\chi^3/\alpha^3$ $\qquad (1^52)_3 = 0$

$(1^42^2)_3 = 6a\lambda^4\chi^4/\alpha^5$ $\qquad (1^32^3)_3 = 0$

$(1^22^4)_3 = 12\lambda^5\chi^5a^2/\alpha^7$ $\qquad (12^5)_3 = 0$

$(2^6)_3 = 120\lambda^6\chi^6a^3/\alpha^9$

$(1^6)_4 = 5\lambda^2\chi^2(6\lambda\chi+5\psi^2)/\alpha^4$

$(1^52)_4 = 80a\psi\chi^3\lambda^3/\alpha^5$

$(1^42^2)_4 = \lambda^3\chi^3a\{(72\alpha+124)\lambda\chi+66\psi^2\}/\alpha^6$

$(1^32^3)_4 = 4\lambda^3\chi^3a\psi\{(65\alpha+97)\lambda\chi+7\psi^2\}/\alpha^7$

$(1^22^4)_4 = 8\lambda^3\chi^3a\{(26\alpha^2+90\alpha+74)\lambda^2\chi^2+(46\alpha+62)\lambda\chi\psi^2+\psi^4\}/\alpha^8$

$(12^5)_4 = 160\lambda^4\chi^4a^2\psi\{(8\alpha+15)\lambda\chi+2\psi^2\}/\alpha^9$

$(2^6)_4 = 40\lambda^4\chi^4a^2\{(52\alpha^2+244\alpha+268)\lambda^2\chi^2+(124\alpha+236)\lambda\chi\psi^2+10\psi^4\}/\alpha^{10}$

<u>Products of order 7</u>

$(1^7)_4 = 105\lambda^3\chi^3\psi/\alpha^4$

$(1^62)_4 = 90\lambda^4\chi^4a/\alpha^5$

$(1^52^2)_4 = 80\lambda^4\chi^4\psi a/\alpha^6$

$(1^42^3)_4 = 12\lambda^4\chi^4a\{8\alpha+10)\lambda\chi+\psi^2\}/\alpha^7$

$(1^32^4)_4 = 156\lambda^5\chi^5\psi a^2/\alpha^8$

$(1^22^5)_4 = 40\lambda^5\chi^5a^2\{(7\alpha+11)\lambda\chi+2\psi^2\}/\alpha^9$

$(12^6)_4 = 720\lambda^6\chi^6\psi a^3/\alpha^{10}$

$(2^7)_4 = 1680\lambda^6\chi^6a^3\{2\lambda\chi b+\psi^2\}/\alpha^{11}$

<u>Products of order 8</u>

$(1^8)_4 = 105\lambda^4\chi^4/\alpha^4$

$(1^7 2)_4 = 0$

$(1^6 2^2)_4 = 30\lambda^5\chi^5 a/\alpha^6$

$(1^5 2^3)_4 = 0$

$(1^4 2^4)_4 = 36\lambda^6\chi^6 a^2/\alpha^8$

$(1^3 2^5)_4 = 0$

$(1^2 2^6)_4 = 120\lambda^7\chi^7 a^3/\alpha^{10}$

$(12^7)_4 = 0$

$(2^8)_4 = 1680\lambda^8\chi^8 a^4/\alpha^{12}$

APPENDIX C (a)

Expression for the Third Moment of $\hat{\alpha} = m_1^2/(m_2-m_1)$ Through Terms in n^{-4}

$$\mu_3(\hat{\alpha}) = \Lambda_2^{(3)}/n^2 + \Lambda_3^{(3)}/n^3 + \Lambda_4^{(3)}/n^4 + \ldots ,$$

where:

$$\Lambda_2^{(3)} = \frac{\alpha(\alpha+1)(\lambda+\alpha)^2}{\lambda^4}\{(16\alpha-8)\lambda^2+(40\alpha+16)\lambda\alpha+(24\alpha+20\alpha^2)\}$$

$$\Lambda_3^{(3)} = \frac{(\lambda+\alpha)^2(\alpha+1)}{\lambda^6}\{A_4^{(3)}\lambda^4+A_3^{(3)}\lambda^3\alpha+A_2^{(3)}\lambda^2\alpha^2+A_1^{(3)}\lambda\alpha^3+A_0^{(3)}\alpha^4\} ,$$ where the $A_k^{(3)}$ are given by

$$A_k^{(3)} = \sum_{i=0}^{2} a_i^{(3)}\alpha^i .$$

The $a_i^{(3)}$ are found in the body of the following table for each $A_k^{(3)}$:

$A_k^{(3)}$ \ α^i	Const.	α	α^2
$A_0^{(3)}$	544	1448	928
$A_1^{(3)}$	844	3368	3064
$A_2^{(3)}$	348	1344	3672
$A_3^{(3)}$	1704	-1032	1872
$A_4^{(3)}$	1752	-432	336

$$\Lambda_4^{(3)} = \frac{(\lambda+\alpha)(\alpha+1)}{\alpha\lambda^8}\{B_7^{(3)}\lambda^7+B_6^{(3)}\lambda^6\alpha+B_5^{(3)}\lambda^5\alpha^2+\ldots$$

$$+B_1^{(3)}\lambda\alpha^6+B_0^{(3)}\alpha^7\} \quad , \quad \text{where}$$

$$B_k^{(3)} = \sum_{i=0}^{3} b_i^{(3)}\alpha^i \quad .$$

The $b_i^{(3)}$ are found in the following table for each $B_k^{(3)}$:

$B_k^{(3)}$ \ α^i	Const.	α	α^2	α^3
$B_0^{(3)}$	11,392	49,216	67,680	29,952
$B_1^{(3)}$	39,904	198,464	325,120	173,664
$B_2^{(3)}$	56,068	306,008	579,388	423,616
$B_3^{(3)}$	-30,072	333,548	431,484	561,272
$B_4^{(3)}$	-474,576	408,840	49,496	433,816
$B_5^{(3)}$	-1,086,388	318,036	-103,488	194,008
$B_6^{(3)}$	-991,312	59,456	-49,528	45,896
$B_7^{(3)}$	-321,872	-26,528	-6,080	4,336

APPENDIX C (b)

Expressions for the Fourth Moment of $\hat{\alpha} = m_1^2/(m_2-m_1)$ Through Terms in n^{-4}

$$\mu_4(\hat{\alpha}) = \Lambda_2^{(4)}/n^2 + \Lambda_3^{(4)}/n^3 + \Lambda_4^{(4)}/n^4 + \ldots ,$$

where:

$$\Lambda_2^{(4)} = 12\alpha^2(\alpha+1)^2(\lambda+\alpha)^4/\lambda^4$$

$$\Lambda_3^{(4)} = \frac{(\lambda+\alpha)^2\alpha(\alpha+1)}{\lambda^6}\{A_4^{(4)}\lambda^4+A_3^{(4)}\lambda^3\alpha+A_2^{(4)}\lambda^2\alpha^2\ A_1^{(4)}\lambda\alpha^3 + A_0^{(4)}\alpha^4\} ,$$

where the $A_k^{(4)}$ are given by $A_k^{(4)} = \sum_{i=0}^{2} a_i^{(4)}\alpha^i$.

The $a_i^{(4)}$ are found in the body of the following table for each $A_k^{(4)}$:

$A_k^{(4)}$ \ α^i	Const.	α	α^2
$A_0^{(4)}$	680	1632	960
$A_1^{(4)}$	1440	4608	3360
$A_2^{(4)}$	408	3840	4344
$A_3^{(4)}$	-480	528	2448
$A_4^{(4)}$	-120	-336	504

$$\Lambda_4^{(4)} = \frac{(\lambda+\alpha)^2(\alpha+1)}{\lambda^8}\{B_6^{(4)}\lambda^6+B_5^{(4)}\lambda^5\alpha+B_4^{(4)}\lambda^4\alpha^2+\dots$$
$$+B_1^{(4)}\lambda\alpha^5+B_0^{(4)}\alpha^6\} \quad , \quad \text{where}$$

$$B_k^{(4)} = \sum_{i=0}^{3} b_i^{(4)}\alpha^i \quad .$$

The $b_i^{(4)}$ are found in the body of the following table for each $B_k^{(4)}$:

$B_k^{(4)}$ \ α^i	Const.	α	α^2	α^3
$B_0^{(4)}$	23,168	91,584	117,216	48,864
$B_1^{(4)}$	64,544	312,736	487,392	244,032
$B_2^{(4)}$	63,736	356,312	731,280	498,960
$B_3^{(4)}$	42,380	222,392	439,020	532,608
$B_4^{(4)}$	-63,024	218,604	34,872	311,388
$B_5^{(4)}$	-199,608	171,528	-56,328	93,816
$B_6^{(4)}$	-116,820	40,092	-13,284	11,244

APPENDIX D (a)

Expressions for the Third Moment of $\hat{a} = (m_2 - m_1)/m_1$ Through Terms in n^{-4}

$$\mu_3(\hat{a}) = \Lambda_2^{(3)}/n^2 + \Lambda_3^{(3)}/n^3 + \Lambda_4^{(3)}/n^4 + \ldots ,$$

where:

$$\Lambda_2^{(3)} = \frac{(\lambda+\alpha)}{\lambda\alpha^5}\{\lambda^3(8\alpha^2+52\alpha+46)+\lambda^2\alpha(16\alpha^2+92\alpha+77) +\lambda\alpha^2(8\alpha^2+44\alpha+36)+4(\alpha+1)\alpha^3\} ,$$

$$\Lambda_3^{(3)} = -\frac{(\lambda+\alpha)(\alpha+1)}{\lambda\alpha^6}\{\lambda^3(8\alpha^2+208\alpha+554)+\lambda^2\alpha(16\alpha^2+416\alpha+1117) +\lambda\alpha^2(8\alpha^2+244\alpha+692)+(36\alpha+126)\alpha^3\},$$ and

$$\Lambda_4^{(3)} = \frac{(\lambda+\alpha)(\alpha+1)}{\lambda^3\alpha^7}\{\lambda^5(164\alpha^2+2190\alpha+4588) +\lambda^4\alpha(340\alpha^2+4815\alpha+10406) +\lambda^3\alpha^2(208\alpha^2+3308\alpha+7520) +\lambda^2\alpha^3(32\alpha^2+672\alpha+1676)-(8\alpha+24)\lambda\alpha^4-4\alpha^5\}.$$

APPENDIX D (b)

Expressions for the Fourth Moment of $\hat{a} = (m_2-m_1)/m_1$ Through Terms in n^{-4}

$$\mu_4(\hat{a}) = \Lambda_2^{(4)}/n^2 + \Lambda_3^{(4)}/n^3 + \Lambda_4^{(4)}/n^4 + \ldots ,$$

where:

$$\Lambda_2^{(4)} = \frac{3(\lambda+\alpha)^2}{\alpha^6}\{\lambda(2\alpha+3)+2\alpha(\alpha+1)\}^2 ,$$

$$\Lambda_3^{(4)} = \frac{(\lambda+\alpha)}{\lambda^2\alpha^7}\{A_5^{(4)}\lambda^5+A_4^{(4)}\lambda^4\alpha+A_3^{(4)}\lambda^3\alpha^2+A_2^{(4)}\lambda^2\alpha^3 +A_1^{(4)}\lambda\alpha^4+A_0^{(4)}\alpha^5\} ,$$

where the $A_k^{(4)}$ are given by $A_k^{(4)} = \sum_{i=0}^{3} a_i^{(4)}\alpha^i$.

The $a_i^{(4)}$ are found in the following table for each $A_k^{(4)}$:

$A_k^{(4)}$ \ α^i	Const.	α	α^2	α^3
$A_0^{(4)}$	8	8	0	0
$A_1^{(4)}$	256	328	72	0
$A_2^{(4)}$	1644	2220	600	24
$A_3^{(4)}$	4025	5524	1572	72
$A_4^{(4)}$	4182	5736	1632	72
$A_5^{(4)}$	1554	2112	588	24

$$\Lambda_4^{(4)} = \frac{(\lambda+\alpha)(\alpha+1)}{\lambda^3\alpha^8}\{B_6^{(4)}\lambda^6+B_5^{(4)}\lambda^5\alpha+B_4^{(4)}\lambda^4\alpha^2+ \dots + B_1^{(4)}\lambda\alpha^5+B_0^{(4)}\alpha^6\} \quad ,$$

where $B_k^{(4)} = \sum_{i=0}^{3} b_i^{(4)}\alpha^i$.

The $b_i^{(4)}$ are found in the following table for each $B_k^{(4)}$:

$B_k^{(4)}$ \ α^i	Const.	α	α^2	α^3
$B_0^{(4)}$	16	0	0	0
$B_1^{(4)}$	-676	-260	0	0
$B_2^{(4)}$	-13,564	-6,628	-528	0
$B_3^{(4)}$	-66,432	-35,616	-3,732	-36
$B_4^{(4)}$	-133,616	-74,716	-8,568	-108
$B_5^{(4)}$	-119,454	-68,082	-8,052	-108
$B_6^{(4)}$	-39,402	-22,614	-2,688	-36

ABSTRACT

This dissertation deals primarily with the development of the technique of orthogonal statistics and the use of this technique to investigate sampling properties of moment estimators of parameters of the negative binomial distribution.

The general technique of orthogonal statistics which is based on the existence of an infinite set $\{q_r(x)\}$ of orthogonal polynomials associated with a particular distribution, enables one to obtain expansions of sampling moments of statistics which are functions of say, the first k sample moments $m_1, m_2, \ldots, m_k$. The thesis describes the technique in general, and gives tables which facilitate the expansion through terms in n^{-5} of sampling moments of statistics which are functions of any four sample moments.

The need for the development of this technique resulted from an interest in the problem of investigating sampling properties of certain moment estimators for the case of the negative binomial distribution. Thus further work was done on the technique for this particular case. Tables are given in the thesis which simplify the procedure for moment statistics

which result from a sample taken from this particular distribution.

Sampling properties of moment estimators for the negative binomial distribution were investigated. The distribution forms considered in depth were due to Anscombe [<u>Biometrika</u>, <u>37</u> (1950), pp. 358-362] with parameters λ and α, Evans [<u>Biometrika</u>, <u>40</u>(1953), pp. 186-211] with parameters m and a, and Fisher [<u>Annals of Eugenics</u>, <u>11</u> (1941), pp. 182-187] with parameters p and k. The purpose of this study was to obtain an insight into the behavior of expansions through high powers of $1/n$ (e.g., terms in n^{-4}) of the bias, variance, and higher moments for these estimators. It was felt that the usual asymptotic properties described by the first term approximations might be misleading for practical cases (i.e., ordinary sample sizes).

The results verified what was suspected. For the moment estimators of Anscombe's form, when $\alpha > \lambda$ the sample sizes needed to make high order terms negligible for the expansion of the bias and variance were extremely large. (For one particular case, in order to use the usual asymptotic variance safely one would need an n of 2 million.) This then reveals the hazardous practice of using the first term approximation and resulting in a very serious under-assessment of the true variance of the estimate of α. Since for Fisher's form $\hat{k} = \hat{\alpha}$, the same applies. For Evans' form, the situation was in marked contrast. Higher order terms were "damped off" with much smaller sample sizes, and in most cases one is justified in using first term approximations. Studies for Evans' estimators were confined to the range $\lambda > 1$ and $\alpha > 1$.

The results for the estimators of Anscombe's form were compared with similar results for the maximum likelihood estimator of α in order to ascertain the effect on efficiency of the chaotic nature of the n^{-3} term in the expansion of the covariance determinant of $\hat{\alpha}$. The maximum likelihood results were taken from Bowman [Thesis submitted for Ph.D. degree, Virginia Polytechnic Institute, Moments to Higher Orders for Maximum Likelihood Estimators with an Application to the Negative Binomial Distribution]. This study revealed that there is a striking similarity in the n^{-3} term in the covariance determinant for the two estimators. This made the "true" efficiency almost identical to the asymptotic efficiency in cases when sufficiently large sample sizes are used to "sink" terms beyond n^{-3}. This statement cannot be generalized, however, to include any sample size, since for $\alpha > \lambda$ only relatively large sample sizes "damp off" further terms in the covariance determinants for both estimators. Hence one cannot be sure of the behavior of these determinants beyond n^{-3} unless these large sample sizes are used.

Tables and charts are given which display the nature of the expansions given in the text. In particular, charts are given of minimum sample size needed in order that the expansions given can safely be used as approximations.