Some aspects of statistical inference for contagious distributions

Shriniwas Keshavarao Katti

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Mathematics Commons

Recommended Citation


https://lib.dr.iastate.edu/rtd/2617

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
SOME ASPECTS OF STATISTICAL INFERENCE FOR CONTAGIOUS DISTRIBUTIONS

by

Shriniwas Keshevareo Katti

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University Of Science and Technology Ames, Iowa

1960
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION.</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF LITERATURE.</td>
<td>3</td>
</tr>
<tr>
<td>III. SOME USEFUL TECHNIQUES FOR EVALUATING EFFICIENCY.</td>
<td>6</td>
</tr>
<tr>
<td>A. Introductory Remarks.</td>
<td>6</td>
</tr>
<tr>
<td>B. Covariance of Certain Statistics for Arbitrary Distributions.</td>
<td>6</td>
</tr>
<tr>
<td>1. Notation</td>
<td>6</td>
</tr>
<tr>
<td>2. Cov($P_1$, $P_i$)</td>
<td>6</td>
</tr>
<tr>
<td>3. Cov($P_1$, $A_i$)</td>
<td>7</td>
</tr>
<tr>
<td>4. Cov($A_1$, $P_i$)</td>
<td>9</td>
</tr>
<tr>
<td>C. Large Sample Covariance of Certain Statistics for Arbitrary Distributions.</td>
<td>9</td>
</tr>
<tr>
<td>1. The method of statistical differentials for evaluating large sample covariance.</td>
<td>9</td>
</tr>
<tr>
<td>2. Cov($A_1$, $P_1/P_0$)</td>
<td>11</td>
</tr>
<tr>
<td>3. $V(P_1/P_0)$</td>
<td>12</td>
</tr>
<tr>
<td>D. Generalized Variance of Estimators as Functions of the Statistics Used.</td>
<td>13</td>
</tr>
<tr>
<td>1. Notation and preliminary remarks.</td>
<td>13</td>
</tr>
<tr>
<td>2. The case when the number of statistics is equal to the number of parameters.</td>
<td>13</td>
</tr>
<tr>
<td>3. The case when the number of statistics used is larger than the number of parameters.</td>
<td>16</td>
</tr>
<tr>
<td>E. Some Results on Generalized Variances for Some Particular Transformations</td>
<td>17</td>
</tr>
<tr>
<td>1. Lemma 1</td>
<td>17</td>
</tr>
<tr>
<td>2. Lemma 2</td>
<td>19</td>
</tr>
<tr>
<td>3. Corollary 1</td>
<td>19</td>
</tr>
<tr>
<td>4. Corollary 2</td>
<td>20</td>
</tr>
<tr>
<td>5. An extension of lemma 1</td>
<td>21</td>
</tr>
</tbody>
</table>
F. Relationship between Factorial Cumulants and Factorial Moments ... 23

1. Relationship between factorial cumulants and factorial moments ... 23
2. Relationship between factorial moments and moments about origin ... 24

IV. EFFICIENCY OF CERTAIN METHODS OF ESTIMATION FOR THE PASCAL AND THE NEYMAN TYPE A DISTRIBUTIONS ... 25

A. Introductory Remarks ... 25
B. Efficiency for the Pascal Distribution ... 26

1. Notation ... 26
2. Efficiency of the method of the first two moments relative to that of maximum likelihood ... 26
3. Efficiency of the method of the first moment and the first frequency ... 29
4. Efficiency of the method of the first moment and the value of the first two frequencies ... 37

C. Efficiency for the Neyman Type A Distribution ... 43

1. Preliminary notation ... 43
2. Efficiency of the method of the first two moments ... 43
3. Efficiency of the method of the first moment and the first frequency ... 52
4. The minimum chi square method of estimation ... 61

V. EFFICIENCY OF ESTIMATORS FOR CERTAIN THREE PARAMETER FAMILIES ... 89

A. Introductory Remarks ... 89
B. Asymptotic Efficiency for the Poisson v Pascal Distribution ... 91

1. Preliminary remarks ... 91
2. Information determinant ... 91
3. Generalized variance and efficiency for the method of three moments ... 97
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Generalized variance and efficiency of the method of the first two</td>
<td>10?</td>
</tr>
<tr>
<td>moments and the zero frequency.</td>
<td></td>
</tr>
<tr>
<td>5. Generalized variance and efficiency for the ratio of the first two</td>
<td>103</td>
</tr>
<tr>
<td>frequencies</td>
<td></td>
</tr>
<tr>
<td>C. Asymptotic Efficiency for the Pascal v</td>
<td></td>
</tr>
<tr>
<td>Poisson Distribution.</td>
<td>102</td>
</tr>
<tr>
<td>1. Preliminary remarks</td>
<td>109</td>
</tr>
<tr>
<td>2. Information determinant</td>
<td>110</td>
</tr>
<tr>
<td>3. Generalized variance and efficiency for the method of the first</td>
<td>113</td>
</tr>
<tr>
<td>three moments</td>
<td></td>
</tr>
<tr>
<td>4. Generalized variance and efficiency for the method of the first</td>
<td>116</td>
</tr>
<tr>
<td>two moments and the first frequency</td>
<td></td>
</tr>
<tr>
<td>5. Generalized variance and efficiency for the method of the first</td>
<td>118</td>
</tr>
<tr>
<td>two moments and the ratio of the first two frequencies</td>
<td></td>
</tr>
<tr>
<td>D. Asymptotic Efficiency for the Poisson v</td>
<td>120</td>
</tr>
<tr>
<td>Binomial Distribution</td>
<td></td>
</tr>
<tr>
<td>1. Preliminary remarks</td>
<td>120</td>
</tr>
<tr>
<td>2. Information determinant</td>
<td>121</td>
</tr>
<tr>
<td>3. Generalized variance and efficiency of the method of moments.</td>
<td>121</td>
</tr>
<tr>
<td>4. Generalized variance efficiency for the method of the first moment</td>
<td>123</td>
</tr>
<tr>
<td>and the first frequency</td>
<td></td>
</tr>
<tr>
<td>5. Generalized variance and efficiency for the method of the first</td>
<td>125</td>
</tr>
<tr>
<td>two moments and the ratio of the first two frequencies</td>
<td></td>
</tr>
<tr>
<td>VI. SELECTION OF AD HOC METHODS BY PRELIMINARY TESTS</td>
<td>122</td>
</tr>
<tr>
<td>A. Introductory Remarks</td>
<td>122</td>
</tr>
<tr>
<td>B. Definitions</td>
<td>128</td>
</tr>
<tr>
<td>1. Preliminary test</td>
<td>122</td>
</tr>
<tr>
<td>2. P.T. criterion</td>
<td>122</td>
</tr>
<tr>
<td>3. P.T. partition</td>
<td>122</td>
</tr>
<tr>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td></td>
</tr>
<tr>
<td>4. P.T. result</td>
<td>109</td>
</tr>
<tr>
<td>5. P.T. departure function</td>
<td>109</td>
</tr>
<tr>
<td>6. Efficiency of the P.T. relative to a null hypothesis</td>
<td>109</td>
</tr>
<tr>
<td>7. P.T. partition, locally optimal with respect to a given P.T. criterion and a given P.T. departure function for a given true hypothesis</td>
<td>130</td>
</tr>
<tr>
<td>C. A Preliminary Test for Selecting One of the Two Methods of Estimation</td>
<td>130</td>
</tr>
<tr>
<td>1. Notation and statement of the preliminary test</td>
<td>130</td>
</tr>
<tr>
<td>2. A property of the above preliminary test</td>
<td>131</td>
</tr>
<tr>
<td>D. Locally Optimal Critical Region for a Simple P.T. Departure Function</td>
<td>132</td>
</tr>
<tr>
<td>VII. A MEASURE OF RELIABILITY OF STATISTICS</td>
<td>138</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>138</td>
</tr>
<tr>
<td>B. A Definition of Reliability</td>
<td>138</td>
</tr>
<tr>
<td>C. Reliability of Certain Statistics</td>
<td>138</td>
</tr>
<tr>
<td>1. Reliability of the $i^{th}$ observed count</td>
<td>139</td>
</tr>
<tr>
<td>2. Reliability of the ratio of the $i^{th}$ count to the $j^{th}$ count for large samples</td>
<td>140</td>
</tr>
<tr>
<td>3. Reliability of the $j^{th}$ count relative to the ratio $r = \frac{P_i}{P_j}$</td>
<td>141</td>
</tr>
<tr>
<td>4. Reliability of the mean</td>
<td>142</td>
</tr>
<tr>
<td>D. Relation between Efficiency and Reliability</td>
<td>143</td>
</tr>
<tr>
<td>E. General Remarks</td>
<td>144</td>
</tr>
<tr>
<td>VIII. SOME DEVELOPMENTS ON CONTAGION</td>
<td>146</td>
</tr>
<tr>
<td>A. Introductory Remarks</td>
<td>146</td>
</tr>
<tr>
<td>B. Examples of Contagious Situations, wherein Variance Can Be Smaller than the Mean</td>
<td>147</td>
</tr>
<tr>
<td>1. Increasing mean</td>
<td>147</td>
</tr>
<tr>
<td>2. Decreasing variance</td>
<td>147</td>
</tr>
<tr>
<td>3. Finite size of the plot</td>
<td>148</td>
</tr>
<tr>
<td>4. Competition and social instincts</td>
<td>148</td>
</tr>
</tbody>
</table>
C. Distribution of Dental Caries in Children .......................... 149
   1. Statement of the problem ........................................ 149
   2. Model for the distribution of unilateral pitted surface
cavitation count .................................................. 150
   3. Application of the model ....................................... 153
   4. Conclusions regarding the contagion of caries ................. 157

D. Effect of Plot Size and Shape ..................................... 159

IX. CONCLUSIONS ..................................................... 164

X. LITERATURE CITED ................................................ 168

XI. ACKNOWLEDGEMENT ................................................ 171
I. INTRODUCTION

Contagious distributions have been widely used in biological and medical sciences. Two of the problems which are of importance in connection with contagion are (i) to discover the nature of the contagion in an attempt to gain an understanding of the structure of the population and (ii) to develop methods for analyzing data from populations wherein contagion is present.

As for (i), the general approach is to formulate a null hypothesis and test it on samples individually. Since an effect is generally caused by several factors, a simple family of distributions can represent the population only roughly while an accurate underlying distribution will involve such a large number of parameters that its utility is heavily restricted. One of the topics in the thesis is to indicate how a dynamic model can be formulated and advantage taken of the independent samples at various times that may be available in such cases. The effect of plot size and shape on the sample distribution will also be briefly discussed.

In dealing with problem (ii) distributions such as the Pascal and the Neyman Type A have been used by various authors (cf. Beall [2], Bliss and Fisher [4]). Some more general families have been formulated by Skellem [21], Beall and Rescia [3] and Gurland [13]. A major difficulty in applying these distributions to data is that efficient methods of estimation
such as maximum likelihood and minimum chi-square are usually cumbersome. The problem of estimation is attacked here along the following lines:

(1) Construct *ad hoc* methods of estimation for which the resulting equations are relatively simple to solve and study their relative efficiencies.

(2) Develop methods for selecting one of the *ad hoc* methods on the basis of the data.

Most of the results obtained are in respect to (1). Some ideas have been suggested for (2) which require further development.

It is hoped that the results developed for estimating the parameters in the contagious distribution as considered will be of some help in determining the possible applicability of a particular distribution.
II. REVIEW OF LITERATURE

Greenwood and Yule [12] studied families of distributions obtained by using models in which the occurrence of an event affects the probability of the occurrence of the subsequent events and derived the Pascal (or the Negative Binomial) distribution as a particular case. The Neyman Type A (cf. Neyman [17]), the Thomas (cf. Thomas [24]), the Polya Aeppli (cf. Skellem [21]) and the "Poisson generalized Binomial" (cf. Skellem [21]) distributions were later developed by using models similar to those of Greenwood and Yule. The ideas underlying these distributions were explained in mathematical terms by Feller [8]. He also gave a clear definition of the notions, "compounding" and "generalizing".

Some of the above distributions were fitted to ecological data by Evans [7], Beall [2], Skellem [21] and Bliss and Fisher [4]. The "Poisson generalized Binomial" distribution was fitted to data on corn-borers by McGuire et al. [16]. In all applications, the common conclusions were that there is "contagion" in the population and the contagion can be partially accounted for by the models of compounding and generalizing.

Some further families of compound and generalized distributions were developed by Gurland [13] and their shapes were studied by Katti [15]. Preliminary fitting indicates that they can be usefully employed in certain cases when none of the aforementioned distributions give adequate representa-
tion of the population.

One of the principal problems in fitting these distributions is that of estimating their parameters. In most of the cases, the maximus likelihood equations are highly cumbersome to solve. Attempts at simplifying the maximum likelihood equations for the Neyman Type A have been made by Douglas [6] and for the "Poisson generalized Binomial" by Shumway and Gurland [20]. Even in these simplified forms, the equations involve infinite series in the unknown parameters.

As alternatives to the maximum likelihood method, Neyman has developed methods which yield estimates that have the same asymptotic properties as maximum likelihood estimates. Neyman [8] suggests minimizing various "distance" functions for situations in which the underlying probability distributions have a multinomial structure. Barankin and Gurland [1] have generalized Neyman's distance functions for families of distributions belonging to a wider class. Further contributions to the development of such estimators have been made by Taylor [23], Chiang [5] and Ferguson [9].

Since the efficient methods are usually unwieldy, attention has been given to formulating methods which yield simple though not necessarily efficient estimates. Evans [7] has discussed the method of moments and the method of using the first moment and first frequency for the Negative Binomial
and Neyman Type A distributions. A study of the same two methods for the "Poisson generalized Binomial" distribution has been made by Sprott [22].
III. SOME USEFUL TECHNIQUES FOR EVALUATING EFFICIENCY

A. Introductory Remarks

In this chapter, it is intended to establish general formulae which will be used in the subsequent chapters in connection with finding the efficiency of various methods of estimation for certain distributions. Interesting properties of some of the formulae will also be discussed.

B. Covariance of Certain Statistics for Arbitrary Distributions

1. Notation

Let \( x_1, x_2, \ldots, x_N \) be a sample of size \( N \). We will denote by \( P_i \), the probability of obtaining a count \( i \) and by \( \hat{P}_i \), the observed proportional frequency of count \( i \). The \( i \)th population moment about the origin will be denoted by \( \mu_i \) and the \( i \)th sample moment about the origin by \( \hat{\mu}_i \). Throughout this thesis, \( \hat{P}_i \) and \( \hat{\mu}_i \) will be taken as the sample estimates of \( P_i \) and \( \mu_i \), respectively.

2. \( \text{Cov}(\hat{P}_i, \hat{P}_j) \)

It is clear that \( \hat{N}_P^i \) and \( \hat{N}_P^j \) have a marginal trivariate distribution with the frequency function

\[
\frac{N!}{(NP_i)!(NP_j)!(N-NP_i-NP_j)!} P_i^{(NP_i)} P_j^{(NP_j)} (1-P_i-P_j)^{(N-NP_i-NP_j)}.
\]
Then by standard statistical techniques, we derive
\[
\text{Cov}(\hat{P}_i, \hat{P}_j) = - \frac{P_i P_j}{N} \quad i \neq j \quad (2)
\]
and
\[
\text{V}(\hat{P}_i) = \frac{P_i (1 - P_i)}{N} \quad . \quad (3)
\]

3. Cov($\hat{P}_i, \hat{\mu}_j$)

With each sample $x_k$, associate a number $\delta(x_k, j)$ such that
\[
\delta(x_k, j) = 0 \quad \text{if} \quad x_k \neq j
\]
\[
= 1 \quad \text{if} \quad x_k = j
\]
Then $\hat{P}_j$ will be given by
\[
\hat{P}_j = \frac{1}{N} \sum_{k=1}^{N} \delta(x_k, j) \quad . \quad (4)
\]
Hence
\[
\text{Cov}(\hat{\mu}_i, \hat{P}_j) = \mathbb{E} (\hat{\mu}_i - \mu_i)(\hat{P}_j - P_j)
\]
\[
= \mathbb{E} (\hat{\mu}_i - \mu_i) \hat{P}_j
\]
\[
= \frac{1}{N^2} \mathbb{E} \sum_{k=1}^{N} (x_k^i - \mu_i^i) \sum_{m=1}^{N} \delta(x_m, j)
\]
\[
= (\mu_i^i - \mu_j^i) P_j / N \quad . \quad (5)
\]
a. An interesting property of $\hat{\mu}_i^i / \hat{\mu}_j^i$ and $\hat{P}_0$

It will be shown that $\hat{\mu}_i^i / \hat{\mu}_j^i$ is asymptotically independent of $\hat{P}_0$.

By setting $j = 0$ in (5), we obtain
\[
\text{Cov}(\hat{\kappa}'_1, \hat{P}_0) = -\frac{\kappa'_1 P_0}{N} \tag{6}
\]

and hence for large samples

\[
\text{Cov}(\log \hat{\kappa}'_1, \hat{P}_0) = \text{Cov}\left\{\log \kappa'_1 (1 + \frac{\hat{\kappa}'_1 - \kappa'_1}{\kappa'_1}), \hat{P}_0 \right\}
\]

\[
= \text{Cov}\left\{\log \kappa'_1 + \frac{\hat{\kappa}'_1 - \kappa'_1}{\kappa'_1}, \hat{P}_0 \right\} \tag{7}
\]

by expending \( \log \hat{\kappa}'_1 \) around \( \kappa'_1 \) and neglecting second and higher order terms in \( \hat{\kappa}'_1 - \kappa'_1 \). Equation (7) can then be further simplified to

\[
\text{Cov}(\log \hat{\kappa}'_1, \hat{P}_0) = \text{Cov}\left\{\frac{\hat{\kappa}'_1}{\kappa'_1}, \hat{P}_0 \right\}
\]

\[
= \frac{1}{\kappa'_1} \text{Cov}(\hat{\kappa}'_1, \hat{P}_0)
\]

\[
= -\frac{P_0}{N} . \tag{8}
\]

Note that formula (8) does not involve \( i \). By changing \( i \) to \( j \), we get

\[
\text{Cov}(\log \hat{\kappa}'_j, \hat{P}_0) = -\frac{P_0}{N} . \tag{9}
\]

On subtracting equation (8) from (9) therefore, we have

\[
\text{Cov}\left\{\log(\hat{\kappa}'_1/\hat{\kappa}'_j), \hat{P}_0 \right\} = 0 . \tag{10}
\]

Since for large samples \( \log(\hat{\kappa}'_1/\hat{\kappa}'_j) \) and \( \hat{P}_0 \) have a bivariate normal distribution, equation (10) implies that \( \log(\hat{\kappa}'_1/\hat{\kappa}'_j) \) is independent of \( \hat{P}_0 \) and consequently, that \( \hat{\kappa}'_1/\hat{\kappa}'_j \) is asymptotically independent of \( \hat{P}_0 \).
4. Cov(\(\hat{\kappa}_1^i, \hat{\kappa}_j^i\))

From the definition of \(\hat{\kappa}_1^i\) and \(\hat{\kappa}_j^i\), it is clear that

\[
\text{Cov}(\hat{\kappa}_1^i, \hat{\kappa}_j^i) = E\left((\hat{\kappa}_1^i - \kappa_1^i)(\hat{\kappa}_j^i - \kappa_j^i)\right)
\]

\[
= E\left((\hat{\kappa}_1^i - \kappa_1^i)\hat{\kappa}_j^i\right)
\]

\[
= \frac{1}{N^2} E \sum_{r=1}^{N} (x_r^1 - \kappa_1^i) \sum_{s=1}^{N} x_s^1
\]

\[
= \frac{1}{N}(\kappa_1^i + \kappa_j^i - \kappa_1^i \kappa_j^i) .
\]

(11)

C. Large Sample Covariance of Certain Statistics for Arbitrary Distributions

1. The method of statistical differentials for evaluating large sample covariances

Let \(t = (t_1, \ldots, t_k)\) be a consistent asymptotically normal estimator of \(\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_k)\) and \(v_1, v_2\), two functions of \(t_1, \ldots, t_k\) which are consistent estimates of the quantities \(\theta_1, \theta_2\) respectively. Then, we have

\[
v_1 = v_1(t_1, \ldots, t_k)
\]

(12)

and

\[
v_2 = v_2(t_1, \ldots, t_k)\text{ say}
\]

(13)

with

\[
\theta_1 = v_1(\gamma_1, \gamma_2, \ldots, \gamma_k)
\]

(14)

and
\[ \Theta_2 = v_2(\tau_1, \tau_2, \ldots, \tau_k). \] (15)

The problem here is to evaluate large sample covariance of \( v_1 \) and \( v_2 \) when the large sample covariance of \( t_i \) and \( t_j \) is known for all \( i \) and \( j \).

On subtracting equation (14) from (15), we obtain

\[ v_1 - \Theta_1 = v_1(t_1, \ldots, t_k) - v_1(\tau_1, \ldots, \tau_k). \] (16)

By expanding \( v_1(t_1, \ldots, t_k) \) in powers of \( (t_1 - \tau_1) \) \( i = 1, \ldots, k \) and neglecting the second and higher order terms in \( (t_i = \tau_i) \), we obtain as a large sample approximation,

\[ v_1 - \Theta_1 = \sum_{i=1}^{k} (t_i - \tau_i) \frac{\partial \Theta_1}{\partial \tau_i}. \] (17)

Similarly

\[ v_2 - \Theta_2 = \sum_{i=1}^{k} (t_i - \tau_i) \frac{\partial \Theta_2}{\partial \tau_i}. \] (18)

Hence

\[ \text{Cov}(v_1, v_2) = E(v_1 - \Theta_1)(v_2 - \Theta_2) \]

\[ = E \sum_{i=1}^{k} \sum_{j=1}^{k} (t_i - \tau_i)(t_j - \tau_j) \frac{\partial \Theta_1}{\partial \tau_i} \frac{\partial \Theta_2}{\partial \tau_j} \]

\[ = \sum_{i=1}^{k} \sum_{j=1}^{k} \text{Cov}(t_i, t_j) \frac{\partial \Theta_1}{\partial \tau_i} \frac{\partial \Theta_2}{\partial \tau_j}. \] (19)

The problem therefore reduces to that of evaluating the various derivatives of \( \Theta_1 \) and \( \Theta_2 \) and simplifying equation
In the forthcoming sections, we will evaluate the large sample covariances of certain functions of observed proportional frequencies and sample moments about the origin by using this method and the formulae derived in section B.

2. \( \text{Cov}(\hat{\kappa}_1', \hat{P}_1/\hat{P}_0) \)

\( \hat{\kappa}_1' \) and \( \hat{P}_1/\hat{P}_0 \) estimate \( \kappa_1' \) and \( P_1/P_0 \) consistently.

Hence, in order to use the method developed in section 1, we obtain

\[
\frac{3}{2P_0} \left( \frac{P_1}{P_0} \right) = - \frac{P_1}{P_0} \quad (20)
\]

\[
\frac{3}{2P_1} \left( \frac{P_1}{P_0} \right) = \frac{1}{P_0} \quad (21)
\]

and \( \frac{3}{\hat{\kappa}_1'} = 1 \).

Hence the asymptotic covariance of \( \hat{\kappa}_1' \) and \( \hat{P}_1/\hat{P}_0 \) is given by

\[
\text{Cov}(\hat{\kappa}_1', \hat{P}_1/\hat{P}_0) = - \left( \frac{P_1}{P_0} \right) \text{Cov}(\hat{P}_0, \hat{\kappa}_1') + \frac{1}{P_0} \text{Cov}(\hat{P}_1, \hat{\kappa}_1') .
\]

(23)

By making use of the formulae for covariances in section B, we can write equation (23) as

\[
\text{Cov}(\hat{\kappa}_1', \hat{P}_1/\hat{P}_0) = - \frac{P_1}{P_0^2} \frac{(- \kappa_1' P_0)}{N} + \frac{1}{P_0} \frac{(1 - \kappa_1') P_1}{N} = \frac{P_1}{P_0 N} .
\]

(24)

a. An interesting property of \( \hat{P}_1/\hat{P}_0 \) and \( \hat{\kappa}_1' - \hat{\kappa}_1' \)

Note that equation (24) does not involve \( i \). This implies that
\[
\text{Cov}(\hat{\mu}_1' - \hat{\mu}_j', \hat{P}_1/\hat{P}_0) - \text{Cov}(\hat{\mu}_1, \hat{P}_1/\hat{P}_0) = 0.
\]

(25)

Since asymptotically \((\hat{\mu}_1' - \hat{\mu}_j')\) and \(\hat{P}_1/\hat{P}_0\) have a bivariate normal distribution, equation (25) implies that \(\hat{\mu}_1' - \hat{\mu}_j'\) is asymptotically independent of the ratio \(\hat{P}_1/\hat{P}_0\).

Note that this property of \(\hat{P}_1/\hat{P}_0\) is due to the fact that for \(j = 0, 1\), \(j^1\) does not involve 1 and hence, referring to equation 3.4, the covariance between \(\hat{\mu}_1'\) and \(\hat{P}_j\) becomes a simple function of \(\hat{\mu}_1'\) and \(P_j\).

3. \(V(\hat{P}_1/\hat{P}_0)\)

\(\hat{P}_1/\hat{P}_0\) estimates \(P_1/P_0\) consistently. By using the derivatives of \(P_1/P_0\) as obtained in section 2, we get

\[
V(\hat{P}_1/\hat{P}_0) = \text{Cov}(\hat{P}_1/\hat{P}_0; \hat{P}_1/\hat{P}_0)
\]

\[
= (- P_1/P_0^2) V(\hat{P}_0) + (1/P_0) V(\hat{P}_1)
\]

\[
- 2(P_1 P_0^2)(1/P_0) \text{Cov}(\hat{P}_0, \hat{P}_1)
\]

\[
= (P_1 P_0^4) P_0(1 - P_0)/N + (1 P_0^2) P_1(1 - P_1) N
\]

\[
- 2(P_1 P_0^3)(- P_1 P_0) N
\]

\[
= (P_0 + P_1)\frac{P_1}{P_0^3}.
\]

(26)
D. Generalized Variance of Estimators as Functions of the Statistics Used

1. Notation and preliminary remarks

Let the vector \((\lambda_1, \ldots, \lambda_r) = \lambda\), say, denote the \(r\) unknown parameters of an arbitrary distribution and \(\gamma = (\gamma_1, \ldots, \gamma_k)\) a vector of \(k\) functions of \(\lambda_1, \ldots, \lambda_r\). Let \(t = (t_1, \ldots, t_k)\) be a consistent estimator of the vector \(\gamma\). Suppose that the parameters \(\lambda\) have been estimated as functions of \(t\). The problem here is to evaluate the generalized variance of these estimators of the parameters as a function of the covariances of the statistics \(t_1, \ldots, t_k\). Formulae for the generalized variance will also be given for the case when the statistics \(t_1, \ldots, t_k\) themselves can be expressed in terms of certain other statistics, say, \(s_1, \ldots, s_k\) for which the expressions for the mutual covariances are relatively simple.

2. The case when the number of statistics \(t\) is equal to the number of parameters

Let us assume that the estimation of the parameters is achieved by equating

\[ t = \gamma. \]  \hspace{1cm} (27)

Since \(\gamma_1\) is a function of the parameters \(\lambda_1, \ldots, \lambda_r\), we can write

\[ \gamma_1 = \gamma_1(\lambda_1, \ldots, \lambda_r) \quad 1 = 1, \ldots, r \]  \hspace{1cm} (28)
Then the relation between the statistics $t_1, \ldots, t_r$ and the estimates $\hat{\lambda}_1, \ldots, \hat{\lambda}_r$ of the parameters $\lambda_1, \ldots, \lambda_r$ is given by

$$t_i = \gamma_i (\lambda_1, \ldots, \lambda_r) \quad i = 1, \ldots, r \quad (29)$$

Hence, by the consistency of the various estimates, we obtain the following large sample relations by neglecting the second and higher order terms in $\hat{\lambda}_1 - \lambda_1$ for all $i$:

$$t_i - \gamma_i = \sum_{j=1}^{r} (\hat{\lambda}_j - \lambda_j) \frac{\delta \gamma_i}{\delta \lambda_j} \quad i = 1, \ldots, r \quad (30)$$

In the matrix form equation 30 can be rewritten as

$$\begin{bmatrix}
  t_1 - \gamma_1 \\
  t_2 - \gamma_2 \\
  \vdots \\
  t_r - \gamma_r
\end{bmatrix} =
\begin{bmatrix}
  \frac{\delta \gamma_1}{\delta \lambda_1} & \frac{\delta \gamma_1}{\delta \lambda_2} & \cdots & \frac{\delta \gamma_1}{\delta \lambda_r} \\
  \frac{\delta \gamma_2}{\delta \lambda_1} & \frac{\delta \gamma_2}{\delta \lambda_2} & \cdots & \frac{\delta \gamma_2}{\delta \lambda_r} \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{\delta \gamma_r}{\delta \lambda_1} & \frac{\delta \gamma_r}{\delta \lambda_2} & \cdots & \frac{\delta \gamma_r}{\delta \lambda_r}
\end{bmatrix}
\begin{bmatrix}
  \hat{\lambda}_1 - \lambda_1 \\
  \hat{\lambda}_2 - \lambda_2 \\
  \vdots \\
  \hat{\lambda}_r - \lambda_r
\end{bmatrix}$$

or in a more abridged form as

$$(t - \gamma)' = J(\gamma, \lambda)(\hat{\lambda} - \lambda)'$$

where
The asymptotic covariance matrix of \( \hat{\lambda} \) is, therefore, given by

\[
C(\hat{\lambda}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{\lambda}_i - \lambda)(\hat{\lambda}_i - \lambda)'
\]

and the generalized variance is given by

\[
|C(\hat{\lambda})| = \frac{|C(t)|}{|J(\tau, \lambda)|^2}.
\]  

(34)

If the \( r \)'s are expressible in terms of \( r \) other constants \((\theta_1, ..., \theta_r) = \Theta\), say, we can obtain the generalized variance for \( r \) in terms of the covariances of the estimators \((\hat{\theta}_1, ..., \hat{\theta}_r) = \hat{\Theta}\), say, as follows.

Let the relationship between the \( \tau \)'s and the \( \theta \)'s be:

\[
\tau_i = \tau_i(\theta_1, ..., \theta_r) \quad 1 = 1, ..., r.
\]

(35)

Then by using arguments similar to those above, we obtain

\[
(t - \tau)' = J(\tau, \theta)(\hat{\theta} - \theta)'
\]

(36)

and

\[
C(t) = J(\tau, \theta)C(\hat{\theta})J(\theta, \theta)'.
\]

(37)

Hence, from equation (34),
\[ |C(\hat{\lambda})| = \left| C(\hat{\theta}) \right| \frac{|J(\tau, \theta)|}{|J(\tau, \lambda)|}^\rho. \quad (38) \]

3. **The case when the number of statistics used is larger than the number of parameters**

In Chapter IV, we will be combining the information on a number of statistics while estimating the parameters. The formula for the generalized variance in the general case when \( r \) parameters are estimated with the help of \( k \) statistics \((k > r)\) is discussed below.

Since the equations for obtaining the estimates of parameters in such situations usually cannot be solved explicitly in terms of the statistics employed, we will assume that these equations have the general implicit form,

\[ f_1(\lambda_1, \ldots, \lambda_r; t_1, \ldots, t_m) = 0 \quad 1 = 1, \ldots, r, \quad \text{and} \quad m > r. \quad (39) \]

If \( \hat{\lambda}_1, \ldots, \hat{\lambda}_r \) are the estimators of \( \lambda_1, \ldots, \lambda_r \), respectively, then we have the relations

\[ f_1(\hat{\lambda}_1, \ldots, \hat{\lambda}_r; t_1, \ldots, t_m) = 0 \quad 1 = 1, \ldots, r. \quad (40) \]

By expanding the function in (40) in a Taylor's series about \( \hat{\lambda} = \lambda \) and \( t = \tau \) and neglecting non-linear terms, we obtain

\[ \sum_{j=1}^{r} (\hat{\lambda}_j - \lambda_j) \frac{\partial f_1}{\partial \lambda_j} + \sum_{j=1}^{m} (t_j - \tau_j) \frac{\partial f_1}{\partial \tau_j} = 0, \quad (41) \]

due to the fact that

\[ f_1(\lambda_1, \ldots, \lambda_r; \tau_1, \ldots, \tau_m) = 0 \]
since the statistics are consistent estimates of $\gamma$. In matrix form, equation (41) can be written as

$$\left( \hat{\lambda} - \lambda \right) J(f, \lambda) = -(t - \gamma) J(f, \gamma). \quad (42)$$

Hence,

$$\left| C(\hat{\lambda}) \right| = \left| E(\hat{\lambda} - \lambda)'(\hat{\lambda} - \lambda) \right|
= \left| E J(f, \lambda)' - 1 J(f, \gamma)' (t - \gamma)' (t - \gamma) J(f, \gamma) J(f, \lambda)^{-1} \right|
= \frac{|J(f, \gamma)' C(t) J(f, \gamma)|}{|J(f, \lambda)|^2} \quad (43).$$

If the $\gamma$'s are expressible in terms of $\epsilon$'s as in section 2, the covariance matrix $C(t)$ is replaced by the expression in equation (37).

E. Some Results on Generalized Variances for Some Particular Transformations

Here, we will establish some results which will effectively reduce the task of evaluating the generalized variance of the statistics used in the estimation of parameters.

1. Lemma 1

If the statistics $t_1, \ldots, t_r$ and the statistics $s_1, \ldots, s_r$ bear a triangular relationship with each other and are such that

$$\frac{s_{t_i}}{s_{s_1}} = \frac{s_{t_i}}{s_{s_1}} = 1 \quad i = 1, \ldots, r \quad (44)$$

the large sample generalized variances of the $t$'s is the same as the large sample generalized variance of the $s$'s.
Proof: Because of the triangular relationship, t₁'s have the form
\[ t₁ = t₁(s₁, \ldots, s₁) \quad i = 1, \ldots, r. \] (45)

Hence
\[
\begin{vmatrix}
1 & 0 & 0 \\
\frac{J(t,s)}{s₁} & 1 & 0 \\
\vdots & \vdots & \vdots \\
\frac{J(t,s)}{s₁} & \frac{J(t,s)}{s₁} & \ldots & 1
\end{vmatrix}
= 1 . \] (46)

Thus from equation (37),
\[ |C(t)| = |C(s)| \]
and this proves the lemma. This result is useful when the statistics used are the t's and the s's have a relatively simple algebraic form for the covariance matrix.

In order to indicate the families of statistics for which the conditions in this lemma are satisfied, we first give a definition of \( \kappa \)-generating functions as follows:

a. Definition of \( \kappa \)-generating functions
A function \( f(z) \) is said to be a \( \kappa \)-generating function, generating the quantities \( 1, 2, \ldots \), if
\[ f(s) = b, \text{ not involving } \kappa \text{'s} \] (47)
2. **Lemma 2**

If \( g_1(z) \) is a \( k \)-generating function and \( g_2(z) \), a known function of \( z \), with \( g_2^{(1)}(b) = 1 \), then the quantities \( k_1, k_2, \ldots \), generated by the \( k \)-generating function \( h(z) = g_2 g_1(z) \) about \( z = a \) are triangularly related to the \( s \)'s and

\[
\frac{\partial}{\partial z} f(a) = \mu_1
\]

(48)

Proof: Since \( h(a) = g_2^{(1)}(a) = g_2^{(1)}(b) = 1 \), \( h(z) \) can be regarded as a \( k \)-generating function. If we set \( u = g_1(z) \), we can write

\[
h(z) = g_2(u)
\]

From here, it is easy to observe that the derivative of an arbitrary order \( r \) of \( h(z) \) will not involve any derivative of \( g_1(z) \) of order higher than \( r \). Hence \( k_r \) can be expressed in terms of \( \mu_1, \ldots, \mu_r \) only. Besides the coefficient of \( g_1^{(r)}(z) \) in \( h^{(r)}(z) \) is \( g_2^{(1)}(u) \). Hence at \( z = a \), the coefficient of \( \mu_r \) will be \( g_2^{(1)}(b) \) which is 1 by definition. The equations (47) and (48) are therefore satisfied and this establishes the lemma.

3. **Corollary 1**

Let \( s, t, u \) be three sets of statistics, satisfying the conditions that
(i) s and t are triangularly related and satisfy condition (44), and
(ii) t and u are triangularly related and satisfy condition (44).

Then s and u are triangularly related and satisfy condition (44).

The proof is very straightforward and hence is omitted here.

4. Corollary 2

The following sets of statistics have the same generalized variance:

(i) the first n sample central moments
(ii) the first n sample moments about the origin
(iii) the first n sample factorial moments
(iv) the first n sample cumulants, and
(v) the first n sample factorial cumulants.

Proof: Let $g(z)$ denote the probability generating function, $\mu'_1$, the $i$th moment about the origin, $\mu_1$, the $i$th central moment, $\kappa'_1$, the $i$th cumulant and $\kappa_{111}$ the $i$th factorial cumulant. Then by the definition of these statistics, the $\mu'$-generating function is given by $g(e^z)$, the $\mu$-generating function by $g(e^z) \exp(-z\mu'_1)$, the $\mu_{11}$-generating function by $g(1+z)$, the $\kappa$-generating function by $\log g(\exp(z))$ and the $\kappa_{11}$-generating function by $\log g(1+z)$. It is
apparent from lemma 2 that the $\mu'$'s, the $\kappa'$'s, the $k'$'s and the $k_{i,j}'$s are triangularly related to the $\kappa_{i,j}$'s and each of the sets along with the set of $\kappa_{i,j}$'s satisfies the condition (44). Hence from corollary 1, the sets are mutually triangularly related and any pair of these sets satisfies equation (44). Therefore, the result in this corollary follows from lemma 1.

It is to be noted that the $(i,j)^{th}$ element in the covariance matrix of $(\kappa_1, \kappa_2, \ldots, \kappa_p)$ is $\kappa_{i+j} - \kappa_i \kappa_j$ for all $i$ and $j$. The simplicity of this form is remarkable.

5. An extension of lemma 1

Let $\xi = (\xi_1^{(1)}, \xi_2^{(2)})$ and $\eta = (\eta_1^{(1)}, \eta_2^{(2)})$ be two vectors of $m$ statistics each. Suppose that $c = (c_1, c_2)$ and $e = (e_1, e_2)$ are consistent asymptotically normal estimates of $\mu$ and $\kappa$, respectively. Then, if $c_1 = (c_1, \ldots, c_s)$ and $e_1 = (e_1, \ldots, e_s)$ ($s \leq m$) are triangularly related and satisfy condition (44), then

$$|C(c)| = |C(e)| |J(\xi^{(2)}, \eta^{(2)})|.$$  (49)

Proof: Under the conditions mentioned, the relationship between the $c$'s and the $e$'s will be of the form

$$c_1 = c_1(e_1, \ldots, e_1) \quad i = 1, \ldots, s$$

$$= c_1(e_1, \ldots, e_m) \quad i = s + 1, \ldots, m$$  (50)

with
\[
\begin{align*}
\xi_1 &= c_1(\eta_1, \ldots, \eta_1) \quad i = 1, \ldots, s \\
&= c_1(\eta_1, \ldots, \eta_m) \quad i = s + 1, \ldots, m. \quad (51)
\end{align*}
\]

Therefore

\[
J(z', \eta) = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
\frac{c_2}{\eta_1} & 1 & 0 & 0 & 0 \\
\frac{c_3}{\eta_1} & \frac{c_3}{\eta_2} & 1 & 0 & 0 \\
\frac{c_{s+1}}{\eta_1} & \frac{c_{s+1}}{\eta_2} & \frac{c_{s+1}}{\eta_s} & 1 & 0 \\
\frac{c_m}{\eta_1} & \frac{c_m}{\eta_2} & \ldots & \frac{c_m}{\eta_m} & \frac{c_n}{\eta_{s+1}} & \ldots & 1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\frac{c_{s+1}}{\eta_{s+1}} & \ldots & \frac{c_m}{\eta_{s+1}} \\
\frac{c_{s+1}}{\eta_{s+1}} & \ldots & \frac{c_m}{\eta_m} \\
\ldots & \ldots & \ldots & \ldots & \ldots
\end{bmatrix}
\]

\[
= J(z', (2), (2)) \quad (52)
\]

If we substitute this in equation (34), the required result follows. This result is of special help when one is dealing with a combination of moments and frequencies wherein the
F. Relationship between Factorial Cumulants and Factorial Moments

In the sections on Finding the efficiencies of the methods of estimation, we will frequently be converting the first six factorial cumulants into moments about the origin. The formulae are given below by which factorial cumulants are converted into factorial moments and the factorial moments into moments about the origin.

1. Relationship between factorial cumulants and factorial moments

If \( h(u) \) is the factorial moment generating function, then it is known that the factorial cumulant generating function \( \psi(u) \) is given by

\[
\psi(u) = \log h(u).
\]

By using the notation in section D.4 above, we observe that the \( i \)th factorial cumulant \( \kappa_i \) can be obtained in terms of the factorial moments by differentiating the above equation \( i \) times and setting \( u = 0 \). The relations obtained this way are as follows:

\[
\begin{align*}
\kappa[1] &= \kappa[1] \\
\end{align*}
\]
\[ M_{[5]} = \kappa_{[5]} + 5 \mu_{[1]} \mu_{[4]} - 20 \mu_{[1]}^2 \mu_{[3]} + 10 \mu_{[2]} \mu_{[3]} \]
\[ - 30 \mu_{[1]} \mu_{[2]} + 36 \mu_{[1]}^3 \mu_{[2]} + 24 \mu_{[1]} \mu_{[2]}^2 \]
\[ - 24 \mu_{[1]}^5 \]
\[ \mu_{[6]} = \kappa_{[6]} + 6 \mu_{[1]} \mu_{[5]} + 15 \mu_{[2]} \mu_{[4]} - 30 \mu_{[1]}^2 \mu_{[4]} \]
\[ - 120 \mu_{[1]} \mu_{[2]} \mu_{[3]} + 120 \mu_{[1]}^3 \mu_{[3]} + 10 \mu_{[3]}^2 \]
\[ - 30 \mu_{[2]}^3 - 324 \mu_{[1]}^4 \mu_{[2]} + 270 \mu_{[1]}^2 \mu_{[2]}^2 \]
\[ + 120 \mu_{[1]}^5 \].

2. **Relationship between factorial moments and moments about origin**

By definition, the \( i \)th factorial moment \( \mu_{[1]} \) is given by \( \mu_{[1]} = \sum x(x - 1) \ldots (x - i + 1) \). The expression for \( \mu_{[1]} \) in terms of the moments about the origin are directly obtainable by expanding the product and taking the expectation term by term. The results obtained by this method, rearranged with a view to obtaining \( \mu_{[1]}'s \) from \( \mu_{[1]}'s \) are as follows:

\[ \mu_1 = \mu_{[1]} \]
\[ \mu_2 = \mu_{[2]} + \mu_{[1]} \]
\[ \mu_3 = \mu_{[3]} + 3\mu_{[2]} - 2\mu_{[1]} \]
\[ \mu_4 = \mu_{[4]} + 6\mu_{[3]} - 11\mu_{[2]} + 6\mu_{[1]} \]
\[ \mu_5 = \mu_{[5]} + 10\mu_{[4]} - 35\mu_{[3]} + 50\mu_{[2]} - 24\mu_{[1]} \]
\[ \mu_6 = \mu_{[6]} + 15\mu_{[5]} - 85\mu_{[4]} + 225\mu_{[3]} + 274\mu_{[2]} - 120\mu_{[1]} . \]
IV. EFFICIENCY OF CERTAIN METHODS OF ESTIMATION FOR THE PASCAL AND THE NEYMAN TYPE A DISTRIBUTIONS

A. Introductory Remarks

In this chapter, it is intended to study the efficiency of the estimators obtained by the method of the first two moments, of the first moment and the first frequency and of the first moment and the ratio of the first two frequencies* for the Pascal (or Negative Binomial) and the Neyman Type A distributions. The first sample moment has been selected in all of these methods of estimation since it is a maximum likelihood estimator of the population mean. As the first two expected frequencies are substantial for certain regions in the parameter space, it is expected that a combination of frequencies and sample mean will yield a better method of estimation than that based on the first two moments alone, at least in an appropriate region of the parameter space. The efficiency of the minimum chi square method wherein it is possible to combine the information on more than two statistics

*For the sake of brevity, we will refer to the method of using the sample estimates of \( \tau_1, \tau_2, \ldots \) as the method of \( \tau_1, \tau_2, \ldots \). Also since the frequency of a count is a known multiple of the corresponding observed proportion, use of the first frequency in estimation is equivalent to using the observed proportion. Hence, we shall use the word frequency instead of the "observed proportion" in the titles of the methods for brevity. However, we shall use the observed proportion in the formulae to avoid the appearance of the various powers of the sample size \( N \).
for these two parameter families will also be discussed. The principal aim here, as explained in the introduction, is to develop "simple" methods of estimation which do not lose much efficiency.

B. Efficiency for the Pascal Distribution*

1. Notation

For the probability generating function of the Pascal distribution, we shall use the form \((q - pz)^{-k}\) where \(p > 0\), \(q = 1 + p\) and \(k > 0\). The expression for the \(i^{th}\) factorial cumulant is

\[ i = kp^i (i - 1)! \quad i = 1, 2, \ldots \]  

(1)

The probabilities \(P_0\) and \(P_1\) of the first two counts are

\[ P_0 = q^{-k} \]  

(2)

and

\[ P_1 = kp P_0/q \]  

(3)

respectively.

2. Efficiency of the method of the first two moments relative to that of maximum likelihood

The formulae for the generalized variance of these estimates and for the information determinant, as given by Fisher 10 are

\[ 2q^3(k + 1)/(pN^2), \text{ and} \]  

(4)

*For the Pascal distribution, the equations for obtaining maximum likelihood estimates are relatively easy. The efficiency of the various alternative methods is discussed here, for the sake of illustration.
respectively. Hence the efficiency $E$ of this method is given by

$$E = \frac{1}{\left\{ \text{(information determinant)} \times \text{(generalized variance)} \right\}}$$

$$= \left\{ \frac{2(k + 1)}{\left(\frac{1}{y + p/q}\right)} \frac{(p/q)^y}{(y + 1)!/(y + k + 1)!} \right\}^{-1}$$

(where the dummy variable $x$ has been replaced by $y - p$).

Since formula (6) involves an infinite series, it was necessary for computational purposes to decide upon a reliable rule about the number of terms to be taken in summing the series. Our goal was to obtain the value of $E$ correct to three decimal places. In selecting the rule, two possibilities were considered. One was to take a fixed number $n$ of terms for all $k$ and $p$. Since it is clear from equation (6) that the series converges fast for small $p$ while the convergence is slow for large $p$, such a fixed $n$ will lead to unnecessarily high precision for smaller values of $p$ and very low accuracy for larger values. Consequently, this method was rejected in favor of the second alternative which was to keep adding terms till the sum reached a certain degree of precision. Let $T_n$ denote the $n^{th}$ term. The rule, which we used, was to stop adding further terms when a value of $n$ is reached for which
The constant $K$ is chosen in accordance with the desired degree of precision. It is apparent, $(6)$ is a series with positive terms. Also
\[
\frac{T_{n+1}}{T_n} = \frac{(n + 2)^p}{(n + 3)(n + k + 3)} \leq \frac{p}{a}.
\]
The sum of the terms left out is therefore
\[
\sum_{i=n+1}^{\infty} T_i \leq T_n \sum_{i=1}^{\infty} \left( \frac{p}{a} \right)^i = \frac{pT_n}{1 - \frac{p}{a}}.
\]
With the above rule, the error introduced is less than
\[
pK \sum_{i=1}^{n} T_i.
\]
Thus for moderate values of $p$ (i.e. $0 < p \leq 10$), if we use $10^{-u}$ for $K$, the relative error $\sum_{n+1}^{\infty} T_i / \sum_{1}^{n} T_i \leq pK$ here will be bounded by $0.001$. Consequently, the sum of the series and the efficiency will be correct to three significant places. As $E$ is less than 1, it follows that $E$ is correct to three decimal places. The values of $E$ are given for certain combinations of $k$ and $p$ in Table 1. It is of interest to note that for any fixed $p$, the efficiency approaches unity as $k \to \infty$.

Since the expression for the generalized variance for
Table 1. Efficiency of the method of moments for the Pascal distribution

<table>
<thead>
<tr>
<th>k</th>
<th>p</th>
<th>.2</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
<th>5.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>.909</td>
<td>.814</td>
<td>.713</td>
<td>.602</td>
<td>.473</td>
<td>.399</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.924</td>
<td>.845</td>
<td>.760</td>
<td>.667</td>
<td>.559</td>
<td>.499</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>.935</td>
<td>.867</td>
<td>.794</td>
<td>.715</td>
<td>.624</td>
<td>.576</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.943</td>
<td>.884</td>
<td>.821</td>
<td>.752</td>
<td>.675</td>
<td>.635</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>.940</td>
<td>.897</td>
<td>.841</td>
<td>.781</td>
<td>.714</td>
<td>.660</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>.955</td>
<td>.907</td>
<td>.858</td>
<td>.804</td>
<td>.745</td>
<td>.716</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>.968</td>
<td>.934</td>
<td>.899</td>
<td>.863</td>
<td>.824</td>
<td>.805</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>.981</td>
<td>.962</td>
<td>.942</td>
<td>.922</td>
<td>.901</td>
<td>.892</td>
<td></td>
</tr>
</tbody>
</table>

the method of moments is considerably simpler than that for the information determinant, for convenience we shall compute the efficiencies relative to the method of moments. If there is any interest in the efficiency of any of these estimators relative to maximum likelihood, one can easily obtain it by using the formula,

\[
\text{Efficiency relative to maximum likelihood} = \frac{(\text{efficiency relative to method of moments}) \times (\text{efficiency of the method of moments})}{(\text{efficiency of the method of moments})}. \tag{7}
\]

3. Efficiency of the method of the first moment and the first frequency

By substituting the formulae from section B.1 for the Pascal distribution into the appropriate general expressions
for the covariances of the various statistics of Chapter III, section B, we obtain

\[ V(\hat{\mathbf{x}}_1) = \frac{kp(1 + p)}{N} \]  \hspace{1cm} (8)

\[ \text{Cov}(\hat{\mathbf{x}}_1, \hat{P}_0) = -\frac{kp(q)^{-k}}{N} \]  \hspace{1cm} (9)

and

\[ V(\hat{P}_0) = q^{-k}(1 - q^{-k})/N \]. \hspace{1cm} (10)

Also, it can easily be shown by differentiating the formulae in section B.1 that

\[ \frac{\partial}{\partial k} \mathbf{K}[1] = p ; \frac{\partial}{\partial p} \mathbf{K}[1] = k \]

\[ \frac{\partial}{\partial k} P_0 = -q^{-k} \log(q) \quad \text{and} \quad \frac{\partial}{\partial p} P_0 = -kq^{-k-1} \]. \hspace{1cm} (10a)

On referring to section D.2 of Chapter III, it is to be noted that the vector \( (\mathbf{K}[1], P_0) \) corresponds to \( \tau \) and the vector \( (k, p) \) to \( \lambda \). For \( \theta \), we chose \( (\mathbf{K}[1], P_0) \). It is to be noted that

\[ \mathbf{C}(\hat{\theta}) = \frac{1}{N} \begin{vmatrix} kp & -kp q^{-k} \\ -kp q^{-k} & q^{-k}(1 - q^{-k}) \end{vmatrix} \]

\[ J(\tau, \theta) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \]

and

\[ J(\tau, \lambda) = \begin{vmatrix} p & k \\ -q^{-k} \log(q) & -kq^{-k-1} \end{vmatrix} \]. \hspace{1cm} (11)

Hence by using equation (38) of that section, we obtain for the generalized variance of this method, the expression
The efficiency $E'$ relative to the method of moments is given by

$$E' = \frac{\text{generalized variance of the method}}{\text{of the first two moments}} \cdot \frac{\text{generalized variance of the present method}}{\text{.}}$$

By referring to (4) and (11a) we obtain

$$E' = \frac{2k(k + 1)(p - q \log c)^2}{p^2 (q^k - 1 - kp/q)}.$$  \hspace{1cm} (12)

Since the expression for $E'$ is too complex to give us an adequate idea of the shape of the surface in $E'$, $k$, $p$ space, it was decided to draw cross-sections for fixed values of $E'$ on graph paper. We shall refer to such cross-sections as contours of fixed efficiency. Two of the methods that can be usefully employed in this connection are described below.

a. Graphical method for drawing the contours of $E'$

Here, equation (12) is rewritten in the form

$$ \frac{(q^k - 1 - kp/q)}{k} = \frac{2(k + 1)(p - q \log c)^2}{E' p^2}.$$  \hspace{1cm} (13)

Let

$$z = \frac{q^k - 1 - kp/q}{k}$$  \hspace{1cm} (14)

and also

$$z = \frac{2(k + 1)(p - q \log c)^2}{E' p^2}.$$  \hspace{1cm} (15)

For fixed values of $p$, a graph of $z$ versus $k$ is drawn using
equation (15) and the lines given by equation (14) are then drawn on the same graph paper for the same fixed \( p \) and for various desired values of \( E' \). Their intersections give the values of \( k \) corresponding to the fixed \( p \) and the various \( E' \). A typical graph for \( p = 1 \) and \( E' = 0.25, 0.5, 1 \) and \( 2 \) is given in Figure 1.

The principal advantage of the graphical method lies in the fact that it indicates the nature and shape of fixed efficiency contours in greater detail than do the numerical methods since the latter yield only co-ordinates of certain points on the contours but always leave open the possibility of radical changes or kinks in the shape of the curves at points at which the efficiency has not been calculated.

b. Numerical method To improve the contours drawn by the graphical method, it was decided to get a few points on the contour by numerical methods. For computations, the electronic computer, IBM 650 was used. Since our aim is only to draw graphs which do not need accuracy to any large number of decimal places, the following method was employed in computing.

Fix a value of \( k \), say at \( k_1 \), compute the efficiency at a suitably chosen value, \( p = p_0 \) (the graphical method aids in the choice of \( p_0 \)), increase the value of \( p \) successively by an amount \( \Delta p \) and compute the efficiency for that value till the efficiency becomes just equal to or less than the value
Figure 1. Graph of equations (14) and (15)
LEGEND
1. GRAPH OF EQUATION (14)
2. $E' = .25$
3. $E' = .5$
4. $E' = 1.0$

$P = 1$

$Z$ vs $K$
for which the contour is to be drawn. When this stage is reached, a new fixed value of $k$ is selected and the above procedure repeated for a new suitable initial value $p_0$ of $p$. The various values of $k$ were selected in order to get a relatively large number of points in the regions in which the curves show relatively great change in shape. The co-ordinates of the points on the contour were found by interpolating between the points $(k,p)$ for which the efficiency is close to the value of the required efficiency.

Fortran transit system of coding was used in programming. Since the formulae for the efficiency of the method in this section and for the one in the next section both involve a number of common terms and the shapes of the contours are similar, it was found economical to program for the two cases simultaneously. A flow chart for the two together appears in Figure 2.

c. Comments on the efficiency contours in Figure 3
(i) All contours are asymptotic to $p$ axis as $p \to \infty$.
This could be established by examining the expression in (12) for fixed $E'$ and showing that $k \to 0$ as $p \to \infty$, but the details are too cumbersome to be included here.
(ii) All contours for which $E' < 1$ are asymptotic to the $k$ axis as $k \to \infty$. This could be shown by an argument similar to that in (i).
Figure 2. Flow-chart for computing the relative efficiency $E_1$ of the method of the first moment and the first frequency and that of $E_2$ of the method of the first moment and the ratio of the first two frequencies relative to the method of moments.
Figure 3. Contours for the relative efficiency of the method of the first moment and the first frequency relative to that of the first two moments for the Pascal distribution
LEGEND

1. $E' = 0.25$
2. $E' = 0.8$
3. $E' = 1.0$
4. $E' = 1.5$
5. $E' = 2.0$
6. $E' = 3.0$
(iii) All contours for which $E' \geq 1$ intersect the $p$ axis once and once only, for $p > 0$. This property is suggested by the plotted contours in Figure 3 for $E' \geq 1$, but an analytical proof based on (10) is too tedious to be included here.

(iv) The method of the first moment and the first frequency can be almost twice as efficient as the method of moments. This can be seen from (12) by setting $k = 0$ and determining the maximum $E' = \frac{1}{p}$ which occurs at $p = \infty$.

4. **Efficiency of the method of the first moment and the value of the first two frequencies**

By substituting the general formulae in section B.1 corresponding to the Pascal distribution in the appropriate general expressions in section B.2 of Chapter III for the covariances of the various statistics, we obtain

$$\text{Cov}(\hat{\mu}_1, \hat{P}_1/P_0) = kp/Nq \quad (16)$$

and

$$V(\hat{P}_1/P_0) = q^{k-2} kp(1 + kp) \quad (17)$$

Also, we have by differentiating the equations in section B.1,

$$\frac{\partial}{\partial k} (P_1/P_0) = \frac{p}{q} \text{ and } \frac{\partial}{\partial p} (P_1/P_0) = k/q^2 \quad (18)$$

$V(\hat{\mu}_1)$ and the derivatives of $\kappa_1$ are given in section B.3.
On referring to section D.2 of Chapter III, we note that the vectors \( (\kappa_{\text{ll}}, P_{1}/P_{0}) \) and \((k,p)\) correspond to \( \gamma \) and \( \lambda \), respectively. For \( \theta \), we use the vector \( (\kappa_{\text{ll}}', P_{1}/P_{0}) \). Clearly,

\[
C(\hat{\theta}) = \frac{1}{N} \begin{vmatrix} \kappa_{2}' - \kappa_{1}'^{2} & kp/q \\ kp/q & \frac{P_{1}(P_{0} + P_{1})}{P_{0}^3} \end{vmatrix}
\]

\[
J(\gamma, \lambda) = \begin{vmatrix} p & k \\ p/q & k/q^{2} \end{vmatrix}
\]

\[
J(\gamma, \Theta) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}
\]

(19)

The generalized variance, as given by equation (38) of that section can then be simplified to

\[
\frac{q^{2}}{p_{N}^{2}} \left\{ q^{k+1}(q + kp) - 1 \right\} . \tag{20}
\]

By using the expression for the generalized variance of the method of moments, given in equation (4), we get the efficiency \( E' \) of the present method relative to the method of moments as

\[
E' = \frac{\text{generalized variance of the method of the first two moments}}{\text{generalized variance of the present method}} = \frac{2pq(k + 1)}{\left\{ q^{k+1}(q + kp) - 1 \right\}} . \tag{20a}
\]
To draw contours of $E'$, we again use the graphical and the numerical methods. The details are described below.

a. **Graphical method to draw the contours**

Here, we rearrange the equation (20) as

$$q^k(p/q + k) = 1/(pq) + 2(k + 1)/E' \quad (21)$$

Set

$$z = q^k(p/q + k) \quad (22)$$

and

$$z = 1/(pq) + 2(k + 1)/E' \quad (23)$$

By assuming a fixed value for $p$, a graph of $z$ versus $k$ is drawn using equation (22). Straight lines given by (23) are then drawn for various values of $E'$. The intersections yield the co-ordinates of the points on the corresponding constant efficiency contour. A typical graph for $p = 1$ and different values for $E'$ is given in Figure 4.

b. **Numerical method**

The method for this case is analogous to that for the case in section 3. The combined flow chart for this case and the case in section 3 is given on page . The efficiency contours are given in Figure 5.

c. **Comments on the contours**

A few interesting properties of these contours are given below.

(i) The method of the first moment and ratio of the first two frequencies can be almost twice as efficient as the method of two moments. This can be shown as follows:
Figure 4. Graphs of equations (22) and (23)
LEGEND
1. GRAPH OF EQUATION (22)
2. \( E' = 2 \)
3. \( E' = 5 \)
4. \( E' = 1.0 \)
5. \( E' = 1.2 \)
6. \( E' = 1.33 \)
Figure 5. Contours corresponding to the relative efficiency of the method of first moment and the ratio of the first two frequencies relative to that of two moments for the Pascal distribution
LEGEND
1. \( E' = 0.2 \)
2. \( E' = 0.5 \)
3. \( E' = 1.0 \)
4. \( E' = 1.05 \)
5. \( E' = 1.10 \)
6. \( E' = 1.20 \)
From (20a), the expression for $E'$ is

$$E' = 2q(p(k + 1)/(a^{k+1}(a + kp) - 1).$$

It is obvious that for a given $p$, $E'$ is a monotonically decreasing function of $k$. Hence the maximum value of $E'$ occurs on the axis $k = 0$. At $k = 0$,

$$E' = 2pa(q^2 - 1)$$

$$= 2 - 2/(1 + p)$$

which has a maximum value of 2 for large $p$.

(ii) The contours for which $E' < 1$ are asymptotic to both the $k$ and $p$ axes. This could be established analytically on the basis of the expression (20a).

(iii) It can be seen that the contours for which $E' > 1$ are asymptotic to the $p$ axis as $p \to \infty$ and intersect the $p$ axis once and once only for $p > 0$. This curve corresponding to $E = 1$ passes through the origin.

(iv) The region in which the method of moments is not superior to the method of first moment and the ratio of the first two frequencies covers such a small region (cf. Figure 5) of the $(u,p)$ space as compared to the method of the first moment and the first frequency (cf. Figure 3) that its utility may be very limited in practice. More will be said about this in Chapter VII on reliability of statistics.
C. Efficiency for the Neyman Type A Distribution

1. Preliminary notation

For the probability generating function of the Neyman Type A distribution, we will use the form

\[ \exp \left\{ \lambda_1 \left[ \exp \left( \lambda_2 (z - 1) \right) - 1 \right] \right\} \]

where \( \lambda_1 > 0 \) and \( \lambda_2 > 0 \). The expression for the \( i \)th factorial cumulant is

\[ \kappa_i = \lambda_1 \lambda_2^i. \]  \hfill (36)

The probabilities \( P_0 \) and \( P_1 \) of the first two counts are

\[ P_0 = \exp \left\{ \lambda_1 \left[ \exp (\lambda_2) - 1 \right] \right\} \]  \hfill (37)

and

\[ P_1 = \lambda_1 \lambda_2 P_0 \exp (-\lambda_2). \]  \hfill (38)

respectively.

2. Efficiency of the method of the first two moments

a. To find the generalized variance for the method of two moments

From the probability generating function, we obtain the \( \kappa' \)-generating function as

\[ \exp \left\{ \lambda_1 \left[ \exp \left( \lambda_2 (\exp(t) - 1) \right) - 1 \right] \right\}. \]

Since the mean of the distribution is \( \lambda_1 \lambda_2 \), the expression for the \( \kappa \)-generating function can be obtained by multiplying the \( \kappa' \)-generating function by \( \exp(-\lambda_1 \lambda_2 t) \). A straightforward expansion of the \( \kappa \)-generating function at the point
\( t = 0 \) yields

\[
\begin{align*}
\mu_1 &= \lambda_1 \lambda_2 \\
\mu_2 &= \lambda_1 \lambda_2 (1 + \lambda_2) \\
\mu_3 &= \lambda_1 \lambda_2 \left( 1 + 3 \lambda_2 + \lambda_2^2 \right) \\
\mu_4 &= \lambda_1 \lambda_2 (1 + \lambda_2) \left( 1 + 6 \lambda_2 + 3 \lambda_1 \lambda_2 (1 + \lambda_2) \right).
\end{align*}
\]

Hence by using the formula (11) of Chapter III for the generalized variance of the sample moments \( (\hat{\mu}_1, \hat{\mu}_2) \) we obtain

\[
\left| \text{cov}(\hat{\mu}_1, \hat{\mu}_2) \right| = \lambda_1 \lambda_2 \left( 1 + (1 + \lambda_2)^2 + \lambda_1 (1 + \lambda_2)^2 \right)^2 / N^2.
\]

Now, from the form of \( \kappa_{[1]} \) given in (26), it is apparent that

\[
\frac{\partial}{\partial \lambda_1} \kappa_{[1]} = \lambda_2, \quad \frac{\partial}{\partial \lambda_2} \kappa_{[1]} = \lambda_1
\]

and

\[
\frac{\partial}{\partial \lambda_1} \kappa_{[2]} = \lambda_2^2 \quad \text{and} \quad \frac{\partial}{\partial \lambda_2} \kappa_{[2]} = 2 \lambda_1 \lambda_2.
\]

By referring to section D.2 of Chapter III, we note that the vectors \( (\kappa_{[1]}, \kappa_{[2]}) \) and \( (\lambda_1, \lambda_2) \) correspond to \( \gamma \) and \( \lambda \) respectively and hence

\[
\begin{vmatrix}
\lambda_2 & \lambda_1 \\
\lambda_2^2 & 2 \lambda_1 \lambda_2
\end{vmatrix}
\]

For \( \mathcal{G} \), we choose \( \gamma \) itself so that

\[
\left| J(\gamma, \mathcal{G}) \right| = 1.
\]

From corollary 2 of Chapter III, section E, we observe that
\[ C(\hat{\theta}) = C(\hat{\mu}_1, \hat{\mu}_2) \]
\[ = C(\hat{\mu}_1', \hat{\mu}_2') \]

which is given in equation (29). By using equation (38) of section D.2, Chapter III, we obtain the generalized variance of the estimates for the present method in the form
\[ \left\{ 1 + (1 + \lambda_2)^2 + 2 \lambda_1 (1 + \lambda_2)^3 \right\}/(N^2 \lambda_2). \]  

(b) To evaluate the information determinant

Shenton [19] gives the formula for the information determinant as
\[ \left\{ (1 + \lambda_2) - \lambda_1 \lambda_2^2 (\lambda_1 + \lambda_2) \right\}/(\lambda_1 \lambda_2^3) \]  

where
\[ \varphi = E\{(r + 1)^2 \frac{P_{r+1}}{P_r}\}. \]

Douglas [6] has shown that
\[ \frac{P_{r+1}}{P_r} = \lambda_2 \frac{\kappa_{r+1}}{\kappa_r (r + 1)} \]

where \( \kappa_r \) is the rth moment about the origin for the Poisson distribution with mean \( \lambda_1 \exp(-\lambda_2) = \lambda \), say. Hence we can easily deduce that
\[ \frac{P_{r+1}}{P_r} = \frac{P_0 \lambda_2^{r+1} \kappa_{r+1}}{(r + 1)!} \]

with \( P_0 = \exp\left\{ \lambda_1 [\exp(-\lambda_2) - 1] \right\} \). \( \varphi \) can therefore be reduced to
\[ \varphi = \sum_{r=0}^{\infty} (r + 1)^2 \frac{P_{r+1}^2}{P_r^2}. \]
Thus, in evaluating \( \Phi \), we must (i) evaluate the various \( \mu_r^' \) and (ii) decide on the number of terms to be used in the summation of the series in (33a).

1. To evaluate \( \mu_r^' \) We have, \( \lambda = \lambda_1 \exp(-\lambda_2) \).

So, \( \mu_r^' \) is the \( r \)th moment in the Poisson distribution with probability generating function \( \exp\left\{ \lambda (z - 1) \right\} \). Therefore, the \( ' \)-generating function is \( \exp\left\{ \lambda \left[ \exp(t) - 1 \right] \right\} \). Let it be denoted by \( m(t) \). Then we have

\[
m'(t) = \lambda \ m(t) \exp(t) \quad (34)
\]

and

\[
m^{(r+1)}(t) = \lambda \sum_{i=0}^{r} \binom{r}{i} \ m^{(i)}(t) \exp(t), \quad x > 0 .
\]

By setting \( t = 0 \), we obtain

\[
\mu_1^' = \lambda \quad (35)
\]

and

\[
\mu_{r+1}^' = \lambda \sum_{i=0}^{r} \binom{r}{i} \mu_i^', \quad x > 0 .
\]

This recurrence formula was used to obtain the various moments needed in evaluating the series in (33a).
ii. The number of terms used

Our aim again was to obtain the final answer correct to three decimal places. Upon trying a few values for \( \lambda \), it became apparent that the number of terms needed when \( \lambda \) is small is considerably smaller than when \( \lambda \) is large. If we were to fix \( n \), the number of terms to be taken in summing the series, there would be a great waste of time in evaluating the series when the value of \( \lambda \) is small and the accuracy of the results will be relatively small for large \( \lambda \). Hence, as in section B.2, it was decided to take the smallest number of \( n \) of terms for which

\[
\frac{T_n}{\sum_{i=1}^{n} T_i} \leq 10^{-4}.
\]  

(36)

It was hoped that the terms would decrease fast enough following the \( n^{th} \) term so that the error due to omitting them would be less than a small multiple (say \( \leq 10 \)) \( T_n \). In that case the relative error

\[
\sum_{n+1}^{\infty} \frac{T_i}{\sum_{1}^{n} T_i} \leq 10^{-3},
\]

and the sum will be correct to three figures; if the significant figures do not cancel out, the resulting computed efficiency will be correct to three figures as well. Since the efficiency is bounded by 1, this implies that the computed efficiency is correct at least to three decimal places. A
rationale to show that the series does converge fast and that the computed efficiency is accurate to three decimal places in most of the cases is described below:

We have from (33a)

\[ T_r = \frac{\kappa^{r+1} \lambda^r}{r!} \]

Therefore, if the series for \( \phi \) converges at a point \((\lambda, \lambda_2)\), it is necessary that

\[ \lim_{r \to \infty} \frac{T_r}{T_{r-1}} = \lim_{r \to \infty} \frac{\kappa^{r-1} \lambda_2^{r+1}}{r! \kappa^r} = \lambda_2 \]

\[ \leq 1, \]

at that point.

We first note that \( \phi \) is a series with positive terms. So, if it does not converge for a value of \((\lambda, \lambda_2)\), it diverges to \(+\infty\) for that value. It is clear from equation (33), that such a divergence would make the information determinant infinite. Since information determinant is the reciprocal of the generalized variance of the maximum likelihood estimates, an infinite value of the information determinant is impossible. Therefore the series for \( \phi \) converges for all values of \((\lambda, \lambda_2)\).

Next, we note from (36a) that \( \lim_{r \to \infty} \frac{T_r}{T_{r-1}} \) is a linear function of \( \lambda_2 \). Hence for the inequality (36) to be satisfied for all \( \lambda_2 \), it is necessary that the coefficient of \( \lambda_2 \)
be zero and hence that
\[ \lim_{r \rightarrow \infty} \frac{T_r}{T_{r-1}} = 0. \]

This implies that the series converges very rapidly for \( r \) larger than a certain number.

The question now is, does this rapid convergence start immediately after \( r \) attains the value \( n \) satisfying (36)? As a theoretical answer is not feasible we resorted to numerical checking. Instead of the constant \( 10^{-4} \) in the rule for deciding \( n \), the number of terms to be used in summing the series, use was made of the constant \( 10^{-7} \) and the information determinant recalculated for a few extreme values of the vector \( (\lambda_1, \lambda_2) \). This decreasing of the constant increased \( n \) but the new computed value of the information determinant differed from the original computed value by a maximum of 2 in the third place - thus giving evidence that most of the answers are correct to three significant places with an error of about 2 in the third position.

When the number of terms necessary for the inequality (36) to be satisfied was so large that the values of some of the functions involved in the computation of the later terms exceeded the limits of the machine, the series was calculated by using all the terms that the machine could compute and the efficiency calculated using this value of the series. Since such values are not likely to be correct to three places, they
are marked with an asterisk.

Since the terms in the series for $\Phi$ are all positive, the above values which may not be correct to three decimal places may still serve as upper bounds for the efficiency.

The efficiency of the method of moments relative to maximum likelihood obtained by using the information determinant computed as above is given for certain values of $(\lambda_1, \lambda_2)$ in Table 2.

Table 2. Efficiency of the method of moments relative to maximum likelihood for the Neyman Type A distribution

<table>
<thead>
<tr>
<th></th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.955</td>
<td>.870</td>
<td>.797</td>
<td>.659</td>
<td>.572</td>
<td>.516</td>
<td>.404</td>
</tr>
<tr>
<td>.3</td>
<td>.939</td>
<td>.827</td>
<td>.733</td>
<td>.666</td>
<td>.565</td>
<td>.402</td>
<td>.270</td>
</tr>
<tr>
<td>.5</td>
<td>.929</td>
<td>.904</td>
<td>.703</td>
<td>.529</td>
<td>.428</td>
<td>.365</td>
<td>.238</td>
</tr>
<tr>
<td>1.0</td>
<td>.901</td>
<td>.788</td>
<td>.685</td>
<td>.517</td>
<td>.421</td>
<td>.363</td>
<td>.239</td>
</tr>
<tr>
<td>1.5</td>
<td>.928</td>
<td>.796</td>
<td>.698</td>
<td>.545</td>
<td>.458</td>
<td>.404</td>
<td>.30</td>
</tr>
<tr>
<td>2.0</td>
<td>.924</td>
<td>.805</td>
<td>.718</td>
<td>.587</td>
<td>.510</td>
<td>.463</td>
<td>--</td>
</tr>
</tbody>
</table>

In this connection, it is of interest to observe that Shenton [19] has also worked on this problem of evaluating the efficiency of the method of moments. The difficulty, again, lies in computing the value of $\Phi$. He expresses $\Phi$ as an infinite series in which the various terms are polynomials,
orthogonal to each other and uses the first four terms to evaluate $\phi$. A table of the upper bounds obtained by using this value of $\phi$ appears on page 453 of Shenton [19]. That this method is much less efficient than the method given above, especially for large values of $(\lambda_1, \lambda_2)$ can be illustrated by noting that for $(\lambda_1, \lambda_2) = (\phi, \phi)$, the upper bound by Shenton's method is .707 while the one by our method is .463.

Note also that as $\lambda_2$ becomes large, $E^1$ appears to approach zero whatever be the fixed value of $\lambda_1$. Regarding the behaviour of $E^1$ for a fixed $\lambda_2$ and large $\lambda_1$, we prove the following result.

a. Lemma 1. For a fixed $\lambda_2$, the efficiency approaches unity as $\lambda_1$ approaches infinity.

Proof: The moment generating function (m.g.f.) of the Neyman Type A random variable is given by

$$\exp \left\{ \lambda_1 \left[ \exp \left[ \lambda_2 (\exp(t) - 1) \right] - 1 \right] \right\}. \quad (40)$$

Hence the m.g.f. of the random variable

$$Y = (x - \lambda_1 \lambda_2) / \sqrt{\lambda_1 \lambda_2 (1 + \lambda_2)} \quad (41)$$

is given by

$$\exp \left\{ - \sqrt{\lambda_1 \lambda_2 / (1 + \lambda_2)} \right\} \exp \left\{ \lambda_1 \left[ \exp \left[ \lambda_2 \left( \exp(t/\sqrt{\lambda_1 \lambda_2 (1 + \lambda_2)} - 1) \right) \right] - 1 \right] \right\}, \quad (42)$$

which can be approximated for large $\lambda_1$ to
\[
\exp\left\{-t \sqrt{\frac{\lambda_1 \lambda_2}{(1 + \lambda_2)}} \right\} \exp\left\{\lambda_1 \left[\exp\left[\lambda_2 \left(\frac{t}{\sqrt{\frac{\lambda_1 \lambda_2}{(1 + \lambda_2)}} + \frac{1}{2} t^2/(\lambda_1 \lambda_2(1 + \lambda_2))\right]\right] - 1\right]\right\}
\]

which can be further approximated to
\[
\exp\left\{t^2/2\right\}.
\]

This corresponds to the m.g.f. of a standard normal distribution. Hence, for fixed \( \lambda_2 \) and large \( \lambda_1 \), the Neyman Type A random variable has a normal distribution with mean \( \lambda_1 \lambda_2 \) and variance \( \lambda_1 \lambda_2(1 + \lambda_2) \). It is known that for the normal distribution, the method of moments is most efficient. Hence it follows that for large \( \lambda_1 \), the method of moments is most efficient for estimating the parameters of a Neyman Type A distribution.

3. Efficiency of the method of the first moment and the first frequency

By referring to section D.2 of Chapter III, we note that the vectors \( ( \kappa, \mu_2, P_0 ) \) and \( ( \lambda_1, \lambda_2 ) \) here respectively correspond to \( \gamma \) and \( \lambda \) of that section. For \( \Theta \), we choose \( ( \kappa_1, P_0 ) \). It can be easily shown through equations (26) and (27) that in the present case
\[ J(\sigma, \lambda) = \lambda \beta P_0(e^{-\lambda \beta} - 1) \]
\[ P_0(e^{-\lambda \beta} - 1) - \lambda_1 P_0 e^{-\lambda \beta} \]
\[ J(\gamma, \delta) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

By virtue of Chapter III, section B, we obtain

\[ C(\hat{\phi}) = \frac{1}{N} \begin{bmatrix} \lambda_1 \lambda_2 (1 + \lambda_2) & -\lambda_1 \lambda_2 P_0 \\ -\lambda_1 \lambda_2 P_0 & P_0(1 - P_0) \end{bmatrix} \]

Hence, by using equation (38) of Chapter III, we obtain the generalized variance of the estimates of the present method in the form

\[ N^2 \lambda_2 \left\{ (1 + \lambda_2) \exp \left[ \lambda_1 (1 - \exp(-\lambda_2)) \right] - 1 - \lambda_2 - \lambda_1 \lambda_2 \right\} \]

By dividing the generalized variance for the method of moments as given in equation (32) by formula (44a), we set the efficiency of this method relative to the method of moments as

\[ E' = \frac{\lambda_1 \left\{ (1 + \lambda_2) \exp(-\lambda_2) - 1 \right\}^2 \left\{ (1 + \lambda_2) \left[ \lambda_1 (1 + \lambda_2)^2 \lambda_2 + \lambda_2 \right] \right\}}{\lambda_2 \left\{ (1 + \lambda_2) \exp\left[ \lambda_1 (1 - \exp(-\lambda_2)) \right] - 1 - \lambda_2 - \lambda_1 \lambda_2 \right\}} \]
Since the method of drawing contours here is analogous to that in the case of the Pascal, no elaboration of the method will be given here. The contours are drawn in Figure 6.

a. Comments on the contours

Some interesting properties of the contours are given below.

(i) All the contours of relative efficiency less than or equal to 1 have the $\lambda_1$ axis as one of their asymptotes.

(ii) For $\lambda_2 \to \infty$, we have

$$E' = 2 \frac{\lambda_2^2}{\exp(\lambda_1) - 1 - \lambda_1}.$$  \hspace{1cm} (47)

Therefore, every contour has an asymptote parallel to the $\lambda_2$ axis. By taking the limit as $\lambda_1 \to 0$ in (47), we get $E' = 4$. This result and the nature of the contours indicate that this method can be almost four times as efficient as the method of moments.

(iii) The efficiency of the method of the first moment and first frequency is relatively high compared to the method of moments for small $\lambda_1$. More about this will be said in the section on estimation by minimum chi square.

(iv) For $\lambda_2 \to 0$, we have

$$\lim_{\lambda_2 \to 0} E' = 1$$

for every $\lambda_1$. This shows that very close to the
Figure 6. Efficiency of the method of first moment and first frequency relative to the method of moments for the Neyman Type A distribution
LEGEND
1. $E' = 0.5$
2. $E' = 0.8$
3. $E' = 1.0$
4. $E' = 1.5$
5. $E' = 2.0$
6. $E' = 3.0$

METHOD OF MOMENTS BETTER
\( \lambda_1 \) axis, the method of moments is just as efficient as the method of the first moment and first frequency.

4. **Efficiency of the method of the first moment and the ratio of the first two frequencies**

From the equations in section C.1, we have

\[
P_1/P_0 = \lambda_1 \lambda_2 \exp(-\lambda_2).
\]

By referring to section D.2 of Chapter III, we note that the vector \((\kappa_1, P_1/P_0)\) corresponds to \(\gamma\) and the vector \((\lambda_1, \lambda_2)\) to \(\lambda\). For \(\gamma\), we select \((\kappa_1, P_1/P_0)\). Hence it can be easily shown through equations (26) and (27) that in the present case,

\[
J(\tau, \lambda) = \begin{vmatrix}
\lambda_2 & \lambda_1 \\
\lambda_2 \exp(-\lambda_2) & \lambda_1 (1 - \lambda_2) \exp(-\lambda_2)
\end{vmatrix}
\]

\[
J(\gamma, \theta) = \begin{vmatrix}
1 & 0 \\
0 & 1
\end{vmatrix}
\]

by virtue of Chapter III, section B, we obtain

\[
C(\theta) = \begin{vmatrix}
\lambda_1 \lambda_2 (1 - \lambda_2) & \lambda_1 \lambda_2 \exp(-\lambda_2) \\
\lambda_1 \lambda_2 \exp(-\lambda_2) & \frac{P_1(P_0 + P_1)}{P_0^3}
\end{vmatrix}
\]

(51)
The use of formula (38) of that section yields for the generalized variance, the expression

\[
\frac{(1 + \lambda \varphi)^2 \exp(\lambda \varphi) + \lambda \int \lambda \varphi}{N^2 \lambda \varphi} \left\{ \lambda [1 - \exp(-\lambda \varphi)] \right\} - 1.
\]

By dividing the expression for the generalized variance of the method of moments given in equation (39) by formula (58), we get the efficiency \(E'\) of the present method relative to the method of moment as

\[
E' = \frac{\lambda \varphi \left\{ 1 + (1 + \lambda \varphi)^2 + \lambda^2 (1 + \lambda \varphi)^3 \right\}}{(1 + \lambda \varphi)(\exp(\lambda \varphi) + \lambda \int \lambda \varphi)/\exp(\lambda \varphi)} - 1.
\]

Because of the unusual shapes that the contours of \(E'\) take in this case, an elaboration of the method of drawing them is given below.

a. A graphical method of drawing contours of \(E'\)

Equation (55) is first rewritten in the form

\[
(\exp(\lambda \varphi) + \lambda \int \lambda \varphi)/\exp(\lambda \varphi) = 1/(1 + \lambda \varphi) + \lambda \int \lambda \varphi [1 + (1 + \lambda \varphi)^2 + 2 \lambda^2 (1 + \lambda \varphi)^3 / (E'(1 + \lambda \varphi))].
\]

Now set

\[
z = (\exp(\lambda \varphi) + \lambda \int \lambda \varphi)/\exp(\lambda \varphi)
\]

and

\[
z = 1/(1 + \lambda \varphi) + \lambda \int \lambda \varphi [1 + (1 + \lambda \varphi)^2 + 2 \lambda^2 (1 + \lambda \varphi)^3 / (E'(1 + \lambda \varphi))].
\]
Fix $\lambda_2$ equal to a constant and draw the graph of $z$ versus $\lambda_1$ using (58). Families of straight lines given by equation (58) are also drawn for various values of $E'$. A typical graph is shown in Figure 7. It is to be noted that there is an upper limit on the value of $E'$ for which the lines cut the curve at least once.

b. Comments on the contours in Figure 8

Some of the distinguishing properties of the contours are mentioned below.

(i) Contours in the neighborhood of the point $(\lambda_1, \lambda_2) = (1,2)$ are bounded. The contour where this method has highest efficiency relative to that of the method of moments degenerates to a point with non-zero co-ordinates.

(ii) The contours which are not bounded on the right hand side are asymptotic to the $\lambda_1$ axis.

(iii) No contour is asymptotic to the $\lambda_2$ axis.

(iv) On comparing Figure 8 with Figure 6, we note that the method of first moment and first frequency is better than the present method of first moment and ratio of the first two frequencies at all points where the latter is superior to the method of moments. Hence, on these grounds, the method of the present section is not recommendable in practice.
Figure 7. Graphs of equations (27) and (28)
LEGEND
1. GRAPH OF EQUATION (57)
2. $E' = 0.25$
3. $E' = 0.50$
4. $E' = 0.8$
5. $E' = 0.9$
6. $E' = 0.98$
Figure 8. Contours for the relative efficiency of the method of first moment and ratio of first two frequencies relative to the method of moments for the Neyman Type A distribution.
4. The minimum chi square method of estimation

a. Preliminary remarks  In the above sections, we considered the estimation of parameters by methods in which the number of statistics used is equal to the number of the unknown parameters in the distribution. In this present section, we shall consider the case of estimation when the number \( m \) of statistics exceeds the number \( r \) of the unknown parameters. In particular, we shall deal with the Pascal and the Neyman Type A distributions when \( r = 0 \) and estimate the parameters by using the following sets at three statistics \( (m = 3) \):

(i) the first three factorial cumulants

(ii) the first two factorial cumulants and the logarithm of the proportion of zeros

(iii) the first two factorial cumulants and the ratio of the first two frequencies.

b. Development of the minimum chi square method  The following review of the minimum chi square method of generating B.A.N. (Best Asymptotically Normal) estimates is from Ferguson 9, p. 1047.

Let \( X_1, X_2, \ldots, X_n \) be a sequence of independent identically distributed \( s \)-dimensional random vectors whose distribution depends upon a parameter \( \Theta \) belonging to an open subset \( Q \) of \( k \)-dimensional Euclidean space \( \mathbb{R}_k \) with \( k \leq s \). Let \( P(\Theta) = E(X|\Theta) \) be the \( s \)-dimensional vector of the expectations of the vector \( X_n \), and let \( \Sigma(\Theta) = \text{var}(X|\Theta) = E[(X - P(\Theta))(X - P(\Theta))^\prime] \) be the \( s \times s \) covariance matrix which is assumed to be finite and nonsingular for
each \( \theta \in \Theta \). Furthermore, it is assumed that \( P(\theta) \) is a one-to-one bicontinuous map from \( \Theta \) to a subset of \( s \)-dimensional Euclidean space with continuous partial derivatives of the second order. Let \( Z_n \) be the \( s \)-dimensional random vector defined by

\[
nZ_n = \sum_{j=1}^{n} X_j.
\]

The quadratic form

\[
n\left[ Z_n - P(\theta) \right]' \sum(\theta)^{-1} \left[ Z_n - P(\theta) \right]
\]

will be designated by the name of \( \chi^2 \). The value \( \hat{\theta}(Z_n) \) of \( \theta \) which minimizes this quadratic form will be called the minimum \( \chi^2 \) estimate of \( \theta \). As an example take the multinomial case where there are \( n \) independent trials each capable of producing any of \( s + 1 \) possible outcomes. Let the probability on each successive trial be \( p_i(\theta) \) of producing the \( i \)th outcome. Let \( z_i \) denote the proportion of the trials which result in the \( i \)th outcome. Then

\[
2 = n \sum_{i=1}^{s+1} \frac{(z_i - p_i(\theta))^2}{p_i(\theta)}
\]

is the familiar Pearson \( \chi^2 \).

The modification, which is of particular interest to us here is the transformed chi square. For a brief description we again quote Ferguson [9, p. 1048] since his notation is relatively easy to follow:

Let \( g(x) \) be any function from \( \mathbb{R}_s \) to \( \mathbb{R}_s \) with continuous first partial derivatives

\[
g(x) = \begin{bmatrix} g_1(x_1, \ldots, x_s) \\ \vdots \\ g_s(x_1, \ldots, x_s) \end{bmatrix}
\]

(2.5)
Let the $s \times s$ matrix of first partial derivatives be denoted by

$$g(x) = \begin{bmatrix}
\frac{\partial}{\partial x_1} g_1 & \cdots & \frac{\partial}{\partial x_1} g_s \\
\vdots & \ddots & \vdots \\
\frac{\partial}{\partial x_s} g_1 & \cdots & \frac{\partial}{\partial x_s} g_s 
\end{bmatrix}$$

(2.6)

We shall call the quadratic form

$$n \left[ g(Z_n) - g(P(\Theta)) \right]' \left[ g(P(\Theta)) \sum(\Theta) g(P(\Theta)) \right]^{-1} \left[ g(Z_n) - g(P(\Theta)) \right]$$

(2.7)

the transformed $\chi^2$. More generally, . . . [if we] replace the matrix of the quadratic form (2.7) by an estimate $g_n(Z_n, \Theta)$, we get

$$Q_n(\Theta) = n \left[ g(Z_n) - g(P(\Theta)) \right]' M_n(Z_n, \Theta) \left[ g(Z_n) - g(P(\Theta)) \right]$$

(2.8)

We assume that $M_n(Z_n, \Theta) \rightarrow \left[ g(P(\Theta)) \sum(\Theta) g(P(\Theta)) \right]^{-1}$ in probability . . . In addition, one needs regularity conditions on $g$, namely that $\Theta$ is a one-to-one bicontinuous map from a neighborhood of $P(\Theta)$ into $R_s$, with continuous partial derivatives of the second order and that the matrix $g(P(\Theta))$ is nonsingular for each $\Theta \in \Theta$. Then the minimum transformed $\chi^2$ estimates, that is the value $\Theta_n(Z_n)$ of $\Theta$ minimizing (2.7), will be a B.A.N. estimate of $\Theta$.

In the present discussion, we will take the vector of certain statistics for the "random vector" mentioned by Ferguson above and use a suitable transform so that the resulting equations to be solved are not cumbersome. It is expected that the increase in the complexity due to the
inclusion of more statistics to estimate the parameters will be offset by an increase in the efficiency of the estimates.

c. **Formula for the generalized variance of the minimum chi square estimates**

Let $t = (t_1, t_2, \ldots, t_n)$ be the statistics used in the minimum chi square method, and suppose that $t$ is a consistent estimator of $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$. Here, the $\gamma_i$'s are functions of the $r$-parameters $(\lambda_1, \lambda_2, \ldots, \lambda_r)$ to be estimated. Let $C(t)$ denote the covariance matrix of $t$ and $\hat{C}(t)$, a matrix of consistent estimates $(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_m)$, say, of the elements $(\gamma_1, \gamma_2, \ldots, \gamma_m)$, say, of $C(t)$. Then the quantity that is to be minimized through this method is given by

$$Q = (t - \gamma) \hat{C}(t)^{-1} (t - \gamma)'$$

By taking the partial derivatives of $Q$ with respect to $\lambda_1$, we obtain the equations for estimating the parameters $\lambda$ as

$$\frac{\partial Q}{\partial \lambda_1} \hat{C}(t)^{-1} (t - \gamma)' = 0 \quad i = 1, \ldots, r$$

Denote the estimators so obtained by $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_r)$.

Since $\hat{\lambda}$ is a solution of (60), we have

$$f_1(\lambda, \hat{\gamma}, t) = \left\{ \frac{\partial Q}{\partial \lambda_1} \hat{C}(t)^{-1} (t - \gamma)' \right\}_{\lambda = \hat{\lambda}}$$

$$= \phi_1(\hat{\lambda}) \phi_2(\hat{\gamma}) \phi_3(t, \hat{\lambda})$$

$$= 0 \quad i = 1, 2, \ldots, r$$
where

\[ \Phi_1(\lambda) = \left\{ \lambda \right\} \]

and

\[ \Phi_2(\hat{\xi}_j) = O(t)^{-1} \]

and

\[ \Phi_3(t, \lambda) = (t - \tau) \]

at \[ \lambda = \lambda \]. (608)

Due to the consistency of \[ \hat{\lambda}, \hat{\xi}_j, t \], it is clear that

\[ f_1(\lambda, \xi_j, \tau) = \Phi_1(\lambda) \Phi_2(\hat{\xi}_j) \Phi_3(\tau, \lambda) = 0 \]

\[ i = 1, \ldots, r \]. (61a)

Hence, in order to obtain the large sample covariance matrix of the estimates \[ \hat{\lambda} \], we expand \[ f_1 \] around the true values \[ \lambda, \xi_j, \lambda \] and neglect the second and higher order terms. Thus from (61), we have

\[ f_1(\lambda, \xi_j, t) = \Phi_1(\lambda) \Phi_2(\xi_j) \Phi_3(t, \lambda) \]

\[ = \left[ \Phi_1(\lambda) + \sum_{i=1}^{r} \left( \frac{\lambda_i - \lambda_1}{\lambda} \right) \frac{\phi_1(\lambda_i)}{\phi_1(\lambda_1)} \right] \]

\[ \Phi_2(\xi_j) + \sum_{i=1}^{m} \left( \frac{\xi_j - \xi_1}{\xi_1} \right) \frac{\phi_2(\xi_j)}{\phi_2(\xi_1)} \]

\[ \Phi_3(\tau, \lambda) + \sum_{i=1}^{n} \left( \frac{\tau - \tau_1}{\tau_1} \right) \frac{\phi_3(\tau, \lambda)}{\phi_3(\tau_1, \lambda)} \]

\[ + \sum_{i=1}^{r} \left( \frac{\lambda_i - \lambda_1}{\lambda} \right) \frac{\phi_3(\lambda)}{\phi_3(\lambda_1)} \]

\[ \lambda = \lambda \]

\[ \tau = \tau \]

\[ \lambda = \lambda \]

\[ \lambda = \lambda \]
\[ = \varphi_1(\lambda) \varphi_2(\hat{z}_1) \varphi_3(\tau, \lambda) \]
\[ + \sum_{i=1}^{r} \left( \hat{\lambda}_1 - \lambda_1 \right) \left\{ \frac{2}{\partial \hat{\lambda}_1} \varphi_1(\hat{\lambda}_1) \right\} \]
\[ \left. \hat{\lambda} = \lambda \right. \]
\[ \varphi_2(\hat{z}_1) \varphi_3(\tau, \lambda) \]
\[ + \sum_{i=1}^{n} \left( \hat{z}_1 - \xi_1 \right) \varphi_1(\lambda) \varphi_3(\hat{z}_1) \left\{ \frac{2}{\partial \hat{z}_1} \varphi_3(\hat{z}_1) \right\} \]
\[ \hat{z}_1 = \hat{z} \]
\[ \varphi_3(\tau, \lambda) \]
\[ + \sum_{i=1}^{n} \left( t_1 - \tau_i \right) \varphi_1(\lambda) \varphi_3(t, \hat{\lambda}) \left\{ \frac{2}{\partial t_1} \varphi_3(t, \hat{\lambda}) \right\} \]
\[ \left. t = \tau \right. \]
\[ \hat{\lambda} = \lambda \]
\[ + \sum_{i=1}^{r} \left( \hat{\lambda}_1 - \lambda_1 \right) \varphi_1(\lambda) \varphi_2(\hat{z}_1) \left\{ \frac{2}{\partial \hat{\lambda}_1} \varphi_2(\hat{z}_1) \right\} \]
\[ \left. t = \tau \right. \]
\[ \hat{\lambda} = \lambda \]
\[ = 0 \quad \text{(61b)} \]

Note that
\[ \left\{ \varphi_3(t, \hat{\lambda}) \right\} \quad t = \tau \quad \hat{\lambda} = \lambda \quad \quad \text{= \tau - \gamma = 0} \quad \text{(61c)} \]

Hence by (61a) and (61c), we can simplify (61b) to
\[
\sum_{i=1}^{n} (t_i - \tau_i) \varphi_1(\lambda) \varphi_2(\xi) \left\{ \frac{\partial}{\partial t_i} \varphi_3(t, \lambda) \right\} = 0 \quad (i = 1, \ldots, r).
\]

Now,
\[
\sum_{i=1}^{n} (t_i - \tau_i) \left\{ \frac{\partial}{\partial t_i} \varphi_3(t, \lambda) \right\} = \sum_{i=1}^{n} (t_i - \tau_i) \frac{\partial}{\partial t_i} (t - \tau) = \sum_{i=1}^{n} (t_i - \tau_i)(0,0,\ldots,0,1,0,\ldots,0) = t - \tau
\]

and
\[
\frac{\partial}{\partial \lambda_1} \varphi_3(t, \lambda) = \frac{\partial}{\partial \lambda_1} (t - \tau) = -\frac{\partial^2 \tau}{\partial \lambda_1}.
\]

Therefore, equation (61d) can be further simplified to
\[
\varphi_1(\lambda) \varphi_2(\xi) (t - \tau) - \sum_{i=1}^{r} (\lambda_1 - \lambda_1) \varphi_1(\lambda) = \varphi_2(\xi) \frac{\partial \tau}{\partial \lambda_1} = 0 \quad (i = 1, \ldots, r).
\]
By substituting for \( \sigma_1(\lambda) \) and \( \sigma_2(\lambda) \) in terms of \( J(\tau, \lambda) \) and \( C(t) \) by using (60a), we get
\[
\frac{\partial \sigma_1}{\partial \lambda} C(t)^{-1} (t - \tau) - \sum_{i=1}^{r} (\lambda_i - \lambda) \frac{\partial \sigma_1}{\partial \lambda_i} C(t)^{-1} = 0
\]
\[i = 1, \ldots, r\]
\[i.e.
\]
\[
\frac{\partial \sigma_1}{\partial \lambda} C(t)^{-1} (t - \tau) = \sum_{i=1}^{r} (\lambda_i - \lambda) \frac{\partial \sigma_1}{\partial \lambda_i} C(t)^{-1} J(\tau, \lambda)(\lambda_i - \lambda)^t
\]
\[i = 1, \ldots, r\]
\[i.e.
\]
\[
J(\tau, \lambda)C(t)^{-1}(t - \tau)^t = J(\tau, \lambda)C(t)^{-1}J(\tau, \lambda)^t(\lambda_i - \lambda)^t
\]
\[i.e.
\]
\[
(\lambda_i - \lambda)^t = J(\tau, \lambda)C(t)^{-1}J(\tau, \lambda)^t
\]
\[\text{(64)}\]
Therefore, the covariance matrix \( C(\hat{\lambda} - \lambda) \) of \( \hat{\lambda} \) is given by
\[
C(\hat{\lambda}) = E(\hat{\lambda} - \lambda)^t(\hat{\lambda} - \lambda)
\]
\[
= E J(\tau, \lambda)C(t)^{-1}J(\tau, \lambda)^t C(t)^{-1} J(\tau, \lambda)^t
\]
\[
(t - \tau)^t(t - \tau)C(t)^{-1}J(\tau, \lambda)^t J(\tau, \lambda)^t
\]
\[\text{(62)}\]
\[
\begin{align*}
&= \mathbf{J}(\tau, \lambda) \mathbf{C}(t)^{-1} \mathbf{J}(\tau, \lambda)' \mathbf{J}(\tau, \lambda) \mathbf{C}(t)^{-1} \mathbf{J}(\tau, \lambda)')^{-1} \\
&= \mathbf{J}(\tau, \lambda) \mathbf{C}(t)^{-1} \mathbf{J}(\tau, \lambda)' \\
\end{align*}
\]

thus

\[
\mathbf{C}(\hat{\lambda}) = \mathbf{J}(\tau, \lambda) \mathbf{C}(t)^{-1} \mathbf{J}(\tau, \lambda)'^{-1}.
\]

(65)

Suppose the functions \(\tau_1, \tau_2, \ldots, \tau_n\) are expressible in terms of another set of \(n\) functions \(\varphi_1, \varphi_2, \ldots, \varphi_n\) of \(\lambda_1, \ldots, \lambda_n\). Let \((\hat{\varphi}_1, \hat{\varphi}_2, \ldots, \hat{\varphi}_n) = \hat{\varphi}\) be a consistent estimator of \(\varphi\) for which the covariance matrix \(\mathbf{C}(\hat{\varphi})\) has a simple form. Then by substituting for \(\mathbf{C}(t)\) in terms of \(\mathbf{C}(\hat{\varphi})\) using equation (37) of Chapter III, we obtain

\[
\mathbf{C}(\hat{\lambda}) = \left[ \mathbf{J}(\tau, \lambda) \{ \mathbf{J}(\tau, \varphi) \mathbf{C}(\hat{\varphi}) \mathbf{J}(\tau, \varphi)' \}^{-1} \mathbf{J}(\tau, \lambda)' \right]^{-1}.
\]

(65a)

For computing the asymptotic efficiency of the estimators \(\hat{\lambda}\) obtained through the minimum chi square procedure described above for various statistics \(t = (t_1, \ldots, t_n)\), a program was set up on the IBM 650 by which the elements of the matrices \(\mathbf{J}(\tau, \lambda), \mathbf{J}(\tau, \varphi)\) and \(\mathbf{C}(\hat{\varphi})\) could be read into the machine and the generalized variance obtained by the application of equation (65a). The generalized variance was used further in conjunction with the information determinant obtained in the above sections to calculate the efficiency of the estimators obtained by these methods. The problem involved in computing the efficiency, therefore, reduces to that of obtaining the elements of the matrices just mentioned.
This will be described below for the various cases individually.

d. Estimators and their efficiency for the Pascal distribution

1. Use of the first three factorial cumulants

Here, we have

\[
\tau = (\kappa_{[1]}, \kappa_{[2]}, \kappa_{[3]}) = (kp, kp^2, kp^3) \quad (65b)
\]

\[
\lambda = (k, p) \quad (65c)
\]

and

\[
\tau = (\hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \hat{\kappa}_{[3]}) \quad (65d)
\]

For \( \theta \), we choose,

\[
\theta = (\mu_1', \mu_2', \mu_3') \quad (65e)
\]

As mentioned in section B.1 of Chapter III, we take

\[
\hat{\theta} = (\hat{\kappa}_1', \hat{\kappa}_2', \hat{\kappa}_3') \quad (66)
\]

It is easy to see from the expressions for the factorial cumulants given in section B.1 above, that

\[
J(\tau, \lambda) = \begin{vmatrix}
    p & p^2 & kp^3 \\
    p^2 & kp^3 & \end{vmatrix} \quad (66)
\]

To obtain \( J(\tau, \lambda) \), we first obtain the expressions for the first three factorial cumulants in terms of the moments about the origin by using equations (53) and (54) of Chapter III as:

\[
\kappa_{[1]} = \mu_1' \quad (66a)
\]

\[
\kappa_{[2]} = \mu_2' - \mu_1' - \mu_1' \quad (66b)
\]
\[ \kappa_3 = \kappa_3^1 - 3 \kappa_0^1 - 3 \kappa_2^1 \kappa_1^1 + 3 \kappa_1^2 + 2 \kappa_1^1 + 3 \kappa_1^3. \] (66c)

From this, we get

\[ J(\gamma, \delta) = \begin{pmatrix}
1 & 0 & 0 \\
-1 - 3 \kappa_1^1 & 1 & 0 \\
2 - 3 \kappa_2^1 + 6 \kappa_1^1 + 6 \kappa_1^2 & -3 - 3 \kappa_1^1 & 1
\end{pmatrix}. \] (67)

To obtain the values of \( \kappa_1^1 \)'s, we first use the values of \( \kappa_1 \) given in equation (1) of this chapter in conjunction with equations (53) of Chapter III to compute factorial moments and then use these factorial moments in conjunction with equations (54) of Chapter III to obtain \( \kappa_1^1 \). The covariance matrix \( C(\hat{\theta}) \) can be obtained from the results of equation (11), Chapter III, on the covariance of the moments. However, we give below the formula for \( C(\hat{\theta}) \) here for the sake of completeness:

\[ C(\hat{\theta}) = \begin{vmatrix}
\kappa_2^1 - \kappa_1^2 & \kappa_3^1 - \kappa_1 \kappa_0^1 & \kappa_4^1 - \kappa_1 \kappa_3^1 \\
\kappa_3^1 - \kappa_1 \kappa_2^1 & \kappa_4^1 - \kappa_2^2 & \kappa_5^1 - \kappa_2 \kappa_3^1 \\
\kappa_4^1 - \kappa_1 \kappa_3^1 & \kappa_5^1 - \kappa_3 \kappa_3^1 & \kappa_5^1 - \kappa_3^2
\end{vmatrix}. \] (67a)

The equations for estimating the parameters \( \kappa \) and \( \sigma \) can be obtained by substituting in equations (60), the expressions for \( \gamma \), \( t \) and \( \lambda \) as given in (65b), (65c) and (65d) in the
form

\((p, p^2, \sigma_p^3) \hat{C}(t)^{-1} (kp - \hat{k}_{11}, kp^2 - \hat{k}_{22}, \sigma kp^3 - \hat{k}_{33}) = 0\)  \hspace{1cm} (68)

and

\((k, 2kp, 6kp^2) \hat{C}(t)^{-1} (kp - \hat{k}_{11}, kp^2 - \hat{k}_{22}, \sigma kp^3 - \hat{k}_{33}) = 0\) \hspace{1cm} (69)

These can be written in the form

\(k = \frac{(1, p, \sigma p^2) \hat{C}(t)^{-1} (\hat{k}_{11}, \hat{k}_{22}, \hat{k}_{33})}{p(1, p, \sigma p^2) \hat{C}(t)^{-1} (1, p, \sigma p^2)}\) \hspace{1cm} (70)

and

\(k = \frac{(1, 2p, 6p^2) \hat{C}(t)^{-1} (\hat{k}_{11}, \hat{k}_{22}, \hat{k}_{33})}{p(1, 2p, 6p^2) \hat{C}(t)^{-1} (1, p, 2p^2)}\) \hspace{1cm} (71)

respectively. For solving (70) and (71), one may evaluate the values of \(k\) for some extreme values of \(p\), graph the values of \(k\) corresponding to these values of \(p\) and choose the next trial value of \(p\) on the basis of the relative positions of the values \(k\). For refining the estimates of \(k\) and \(p\) obtained by this trial and error method, one may graph \(k\) against \(p\) from the two equations (70) and (71) in the interval of \(p\) in which a solution is expected, join the points by smooth curves and take their intersection as the solution.

As an alternative method of solving equations (70) and (71), we may eliminate \(k\) and solve for \(p\) iteratively.

The generalized variance of the estimators of \(k, p\) is
obtained by substituting in equation (65a) the expressions for $J(\gamma, \alpha)$, $J(\gamma, \beta)$ and $C(\theta)$ given above. The efficiency obtained by using this generalized variance and the information determinant of Table 1 is given in Table 3 for certain values of $k$ and $p$. A few interesting properties of this method are given below:

(i) The efficiency of this method is higher than that of the method of the first two moments (cf. Table 1) at every value of $(k,p)$ in Table 3.

(ii) For a given $k$, the efficiency approaches zero as $p \to \infty$ and for a given $p$, efficiency approaches 1 as $k \to \infty$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$p$</th>
<th>0.2</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
<th>5.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>.990</td>
<td>.959</td>
<td>.903</td>
<td>.816</td>
<td>.682</td>
<td>.593</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.992</td>
<td>.970</td>
<td>.922</td>
<td>.864</td>
<td>.765</td>
<td>.700</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>.994</td>
<td>.977</td>
<td>.945</td>
<td>.896</td>
<td>.820</td>
<td>.775</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.995</td>
<td>.981</td>
<td>.957</td>
<td>.919</td>
<td>.862</td>
<td>.826</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>.996</td>
<td>.985</td>
<td>.965</td>
<td>.935</td>
<td>.890</td>
<td>.853</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>.997</td>
<td>.988</td>
<td>.971</td>
<td>.947</td>
<td>.911</td>
<td>.889</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>.998</td>
<td>.993</td>
<td>.985</td>
<td>.972</td>
<td>.954</td>
<td>.944</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>.999</td>
<td>.997</td>
<td>.994</td>
<td>.990</td>
<td>.984</td>
<td>.981</td>
<td></td>
</tr>
</tbody>
</table>
11. Use of the first two factorial cumulants and the logarithm of the proportion of zeros

Here we have,

$$\gamma = (\kappa_{[1]}, \kappa_{[2]}, \log P_0) = (k_p, k_p^2, - k \log \sigma) \quad (72a)$$

$$\lambda = (k, p) \quad (72b)$$

and

$$t = (\hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \log \hat{P}_0). \quad (72c)$$

For \( \Theta \), we choose

$$\Theta = (\kappa_1', \kappa_2, \log P_0). \quad (72d)$$

As mentioned in section B.1 of Chapter III, we take

$$\Theta = (\kappa_1', \kappa_2, \log P_0). \quad (72e)$$

It is easy to observe from (72a) and (72b) that

$$J(\gamma, \lambda) = \begin{bmatrix} 0 & p^2 & - \log \sigma \\ k & k_p & - k/(1 + p) \end{bmatrix} \quad (73)$$

Note that from the relationships (66a) and (66b), we get

$$J(\Theta, \Theta) = \begin{bmatrix} \frac{\delta \kappa_{[1]}}{\delta \kappa_1} & \frac{\delta \kappa_{[2]}}{\delta \kappa_1} & \frac{\delta \log P_0}{\delta \kappa_1} \\ \frac{\delta \kappa_{[1]}}{\delta \kappa_2} & \frac{\delta \kappa_{[2]}}{\delta \kappa_2} & \frac{\delta \log P_0}{\delta \kappa_2} \\ \frac{\delta \log P_0}{\delta \log P_0} & \frac{\delta \log P_0}{\delta \log P_0} & \frac{\delta \log P_0}{\delta \log P_0} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ -1 - \kappa_2' & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (73a)$$
$C(\hat{\theta})$ obtained by substituting the expressions given in section B, Chapter III, for the covariance between the various statistics is

$$C(\hat{\theta}) = \frac{1}{N} \begin{bmatrix}
\mu_1 - \mu_1^2 & \kappa_3 - \kappa_1 \kappa_2 & - \kappa_1 \\
\kappa_3 - \kappa_1 \mu_2 & \kappa_4 - \kappa_2^2 & - \kappa_2 \\
- \kappa_1 & - \kappa_2 & P_0(1 - P_0)
\end{bmatrix}.$$ (73b)

The values for the $\kappa_i$'s are obtained as in section 1 above.

By substituting for $\gamma$, $t$ and $\lambda$ from equations (79), (72a) and (72b) in (60), we get the equations for estimating the parameters $k$ and $p$ in the form

$$(p, p^2 - \log(1 + p)) \hat{C}(t)^{-1} \left( kp - \hat{\kappa}_{[1]}, kp^2 - \hat{\kappa}_{[2]}, -k \log(1 + p) - \log \hat{P}_0 \right) = 0 \quad (74)$$

and

$$(k, 2kp, -k/(1 + p)) \hat{C}(t)^{-1} \left( kp - \hat{\kappa}_{[1]}, kp^2 - \hat{\kappa}_{[2]}, -k \log(1 + p) - \log \hat{P}_0 \right) = 0 , \quad (75)$$

which can be rewritten in the form

$$k = \frac{(p, p^2, - \log(1 + p)) \hat{C}(t)^{-1} \left( \hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \log \hat{P}_0 \right)}{(p, p^2, - \log(1 + p)) \hat{C}(t)^{-1} \left( p, p^2, - \log(1 + p) \right)} \quad (76)$$

and

$$k = \frac{(1, 2p, -1/(1 + p)) \hat{C}(t)^{-1} \left( \hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \log \hat{P}_0 \right)}{(p, p^2, - \log(1 + p)) \hat{C}(t)^{-1} \left( p, p^2, - \log(1 + p) \right)} \quad (77)$$

A method similar to that in section 1 above may be
employed for obtaining the estimates.

Generalized variance of the estimators of \( k \) and \( \theta \) was obtained by substituting in (65), the expressions for \( J(\gamma, \lambda) \), \( J(\gamma, \theta) \) and \( C(\theta) \) given above. The efficiency, evaluated by using this generalized variance and the information determinant in Table 1, is given in Table 4. The following interesting properties of this method are worth mentioning:

(i) The present method is superior to the method of using the first three factorial cumulants for small \( k \) (cf. Table 3).

Table 4. Efficiency of the minimum chi square method for the Pascal distribution when the characteristics used are the first two factorial moments and the logarithm of the zero frequency

<table>
<thead>
<tr>
<th>( k )</th>
<th>0.2</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
<th>5.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>.996</td>
<td>.998</td>
<td>.995</td>
<td>.987</td>
<td>.966</td>
<td>.440</td>
</tr>
<tr>
<td>1.0</td>
<td>.999</td>
<td>.995</td>
<td>.985</td>
<td>.963</td>
<td>.905</td>
<td>.837</td>
</tr>
<tr>
<td>1.5</td>
<td>.998</td>
<td>.991</td>
<td>.975</td>
<td>.939</td>
<td>.850</td>
<td>.761</td>
</tr>
<tr>
<td>2.0</td>
<td>.997</td>
<td>.992</td>
<td>.967</td>
<td>.919</td>
<td>.814</td>
<td>.726</td>
</tr>
<tr>
<td>2.5</td>
<td>.997</td>
<td>.985</td>
<td>.959</td>
<td>.904</td>
<td>.796</td>
<td>.722</td>
</tr>
<tr>
<td>3.0</td>
<td>.996</td>
<td>.983</td>
<td>.953</td>
<td>.893</td>
<td>.791</td>
<td>.714</td>
</tr>
<tr>
<td>5.0</td>
<td>.995</td>
<td>.976</td>
<td>.940</td>
<td>.885</td>
<td>.827</td>
<td>.805</td>
</tr>
<tr>
<td>10.0</td>
<td>.993</td>
<td>.973</td>
<td>.946</td>
<td>.901</td>
<td>.891</td>
<td></td>
</tr>
</tbody>
</table>
(ii) The efficiency of this method does not decrease as rapidly with \( p \) as in the case of the method of using the first three factorial cumulants (cf. Table 3).

(iii) For a given \( p \), the efficiency seems to approach unity as \( k \) becomes large but does so much slower than in the case of the method using the first three factorial cumulants.

### Use of the first two factorial cumulants and the ratio of the first two frequencies

Here we have,

\[
\tau = (\kappa_{[1]}, \kappa_{[2]}, P_1/P_0) = (kp, kp^2, kp/\alpha) \quad (77a)
\]

\[
\lambda = (k, p) \quad (77b)
\]

and

\[
t = (\hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \hat{P}_1/\hat{P}_0) \quad (77c)
\]

For \( \sigma \), we take

\[
\Theta = (\kappa_{[1]}^1, \kappa_{[2]}^1, P_1/P_0) \quad (77d)
\]

and as mentioned in section B.1 of Chapter III, we take

\[
\hat{\sigma} = (\hat{\kappa}_{[1]}^1, \hat{\kappa}_{[2]}^1, \hat{P}_1/\hat{P}_0) \quad (77e)
\]

It is easy to observe from (77a) and (77b) that

\[
J(\tau, \lambda) = \begin{bmatrix} p & p^2 & p/\alpha \\ k & kp & k/\alpha^2 \end{bmatrix} \quad (78a)
\]

By arguments similar to those in section ii, we get

\[
J(\gamma, \sigma) = \begin{bmatrix} 1 & 0 & 0 \\ -1 - \alpha \kappa_{[1]}^1 & \alpha \kappa_{[1]}^1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (78b)
\]
C(\theta), obtained by using the expressions given in section B, Chapter III, for the covariance between the various statistics is

$$C(\theta) = \frac{1}{N} \begin{bmatrix} \kappa_2^1 - \kappa_1^1 & \kappa_3^1 - \kappa_1^1 \kappa_0^1 & P_1/P_0 \\ \kappa_3^1 - \kappa_1^1 \kappa_2^1 & \kappa_4^1 - \kappa_2^1 & P_1/P_0 \\ P_1/P_0 & P_1/P_0 & P_1(P_0 + P_1)/P_0^3 \end{bmatrix}$$

(79)

where the values of the various \(\kappa^1\)'s are obtained as in section i and \(P_0, P_1\) from equations (\theta) and (3).

By substituting in (50), the expressions for \(\gamma, t\) and \(\lambda\) given in (78), (78a) and (79), we get the equations for obtaining the estimates of \(k\) and \(p\) as

$$(p, p^2, p/q) \hat{C}(t)^{-1} (kp - \hat{\kappa}_{[1]}^1, kp^2 - \hat{\kappa}_{[2]}^2, kp/q - \hat{P}_1/\hat{P}_0) = 0$$

(80)

and

$$(k, kp, k/q^2) \hat{C}(t)^{-1} (kp - \hat{\kappa}_{[1]}^1, kp^2 - \hat{\kappa}_{[2]}^2, kp/q - \hat{P}_1/\hat{P}_0) = 0$$

(81)

For simplicity, these equations may be rewritten as

$$k = \frac{(1, p, 1/q) \hat{C}(t)^{-1} (\hat{\kappa}_{[1]}^1, \hat{\kappa}_{[2]}^2, \hat{P}_1/\hat{P}_0)}{p(1, p, 1/q) \hat{C}(t)^{-1} (1, p, 1/q)}$$

(82)

and

$$k = \frac{(1, p, 1/q^2) \hat{C}(t)^{-1} (\hat{\kappa}_{[1]}^1, \hat{\kappa}_{[2]}^2, \hat{P}_1/\hat{P}_0)}{p(1, p, 1/q^2) \hat{C}(t)^{-1} (1, p, 1/q)}$$

(82a)
Solving of these equations follows along lines similar to those in section 1.

Generalized variance of the estimators of \( k \) and \( p \) obtained by this method was computed by substituting in equation (65a), the expressions for \( J(\tau,\rho) \), \( J(\gamma,\delta) \) and \( C(\hat{\delta}) \) given above. The efficiency, obtained by using this method and the information determinant of Table 1, is given in Table 5. It is to be noted that the efficiency of this method is considerably smaller than that of the minimum chi square method using the first two factorial cumulants and the logarithm of the zero count throughout the tabulated region in the parameter space. Since equations (80) and (81) do not

<table>
<thead>
<tr>
<th>( k )</th>
<th>( p )</th>
<th>0.2</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
<th>5.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>.996</td>
<td>.982</td>
<td>.947</td>
<td>.873</td>
<td>.760</td>
<td>.527</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.995</td>
<td>.975</td>
<td>.929</td>
<td>.839</td>
<td>.678</td>
<td>.571</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>.994</td>
<td>.969</td>
<td>.916</td>
<td>.822</td>
<td>.679</td>
<td>.601</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.993</td>
<td>.965</td>
<td>.908</td>
<td>.817</td>
<td>.699</td>
<td>.643</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>.992</td>
<td>.962</td>
<td>.904</td>
<td>.821</td>
<td>.725</td>
<td>.683</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>.991</td>
<td>.957</td>
<td>.915</td>
<td>.818</td>
<td>.750</td>
<td>.717</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>.989</td>
<td>.957</td>
<td>.915</td>
<td>.818</td>
<td>.804</td>
<td>.805</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>.989</td>
<td>.965</td>
<td>.943</td>
<td>.922</td>
<td>.901</td>
<td>.892</td>
<td></td>
</tr>
</tbody>
</table>
appear to be in any way simpler than equations (76) and (77), this method may not be recommended for practical use.

e. Estimates and their efficiencies for the Neyman Type A distribution

1. Use of the first three factorial cumulants

Here we have

$$\gamma = (\kappa_{11}, \kappa_{12}, \kappa_{33}) = (\lambda_1 \lambda_2, \lambda_1 \lambda_2^2, \lambda_1 \lambda_2^3) \quad (82a)$$

$$\lambda = (\lambda_1, \lambda_2) \quad (82b)$$

and

$$t = (\hat{\kappa}_{11}, \hat{\kappa}_{12}, \hat{\kappa}_{33}) \quad (82c)$$

For $$\Theta$$, we choose

$$\Theta = (\mu_1^1, \mu_2^1, \mu_3^1) \quad (82a)$$

As in section d.i, we again have

$$\hat{\Theta} = (\hat{\mu}_1^1, \hat{\mu}_2^1, \hat{\mu}_3^1) \quad (82e)$$

From (82a) and (82b), we can easily deduce that

$$J(\gamma, \lambda) = \begin{bmatrix} \lambda_2 & \lambda_2^2 & \lambda_2^3 \\ \lambda_1 & 2 \lambda_1 \lambda_2 & 3 \lambda_1 \lambda_2^2 \end{bmatrix} \quad (83)$$

$$J(\gamma, \Theta)$$ and $$C(\hat{\Theta})$$ have obviously the same general expressions as in (67) and (67a), excepting that we now substitute for the $$\mu^i$$'s, the formulae corresponding to the Neyman Type A distribution. Along the same lines as in section d.i, above, the various moments $$\mu_1^i$$ can be obtained by using formulae (26) of this chapter for the factorial cumulants and interrelations (53) and (54) of Chapter III between factorial
cumulants and factorial moments and between the factorial moments and the moments about the origin. By substituting for $\gamma$, $t$ and $\lambda$ from equations (82a), (82c) and (82b) in equations (60), we obtain the equations for estimating the parameters $\lambda_1$ and $\lambda_2$ as

$$\begin{align*}
(\lambda_2, \lambda_2^2, \lambda_2^3) \hat{C}(t)^{-1} (\lambda_1 \lambda_2 - \hat{\kappa}_{[1]}, \lambda_1 \lambda_2^2 - \hat{\kappa}_{[2]}, \lambda_1 \lambda_2^3 - \hat{\kappa}_{[3]}) = 0
\end{align*}$$

(84)

and

$$\begin{align*}
(\lambda_1, 2 \lambda_1, \lambda_2, 3 \lambda_1 \lambda_2^2) \hat{C}(t)^{-1} (\lambda_1 \hat{\kappa}_{[1]}, \lambda_1 \lambda_2 \hat{\kappa}_{[2]}, \lambda_1 \lambda_2^2 \hat{\kappa}_{[3]}) = 0
\end{align*}$$

(84a)

which can be rewritten in the form

$$\begin{align*}
\lambda_1 = \frac{(1, \lambda_2, \lambda_2^2) \hat{C}(t)^{-1} (\hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \hat{\kappa}_{[3]})'}{\lambda_2 (1, \lambda_2, \lambda_2^2) \hat{C}(t)^{-1} (1, \lambda_2, \lambda_2^2)}
\end{align*}$$

(85)

and

$$\begin{align*}
\lambda_2 = \frac{(1, 2 \lambda_2, 3 \lambda_2^2) \hat{C}(t)^{-1} (\hat{\kappa}_{[1]}, \hat{\kappa}_{[2]}, \hat{\kappa}_{[3]})'}{\lambda_2 (1, 2 \lambda_2, \lambda_2^2) \hat{C}(t)^{-1} (1, \lambda_2, \lambda_2^2)}
\end{align*}$$

(86)

For solving, one may use a method similar to the one in section d.i. The generalized variance of the estimators of $\lambda_1$ and $\lambda_2$ obtained by this method was computed by substituting in (65a), the expressions for $J(\gamma, \lambda)$, $J(\gamma, \phi)$, and $C(\delta)$ obtainable from the above discussion. The efficiency, evaluated by using this generalized variance and the
information determinant in Table 2 is given in Table 6. Some interesting properties of the efficiency function are:

(i) For a given $\lambda_2$, the efficiency approaches unity for $\lambda_1 \to \infty$.

(ii) For a given $\lambda_1$, the efficiency approaches zero as $\lambda_2 \to \infty$.

Table 6. Efficiency of the minimum chi square method for the Neyman Type A distribution when the characteristics used are the first three factorial cumulants

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td></td>
<td>.998</td>
<td>.987</td>
<td>.968</td>
<td>.920</td>
<td>.841</td>
<td>.792</td>
<td>.704</td>
</tr>
<tr>
<td>.3</td>
<td></td>
<td>.967</td>
<td>.975</td>
<td>.939</td>
<td>.889</td>
<td>.729</td>
<td>.659</td>
<td>.457</td>
</tr>
<tr>
<td>.5</td>
<td></td>
<td>.995</td>
<td>.965</td>
<td>.916</td>
<td>.769</td>
<td>.641</td>
<td>.548</td>
<td>.329</td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td>.993</td>
<td>.946</td>
<td>.866</td>
<td>.659</td>
<td>.507</td>
<td>.491</td>
<td>.357</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>.992</td>
<td>.928</td>
<td>.821</td>
<td>.605</td>
<td>.490</td>
<td>.424</td>
<td>.307</td>
</tr>
<tr>
<td>2.0</td>
<td></td>
<td>.990</td>
<td>.908</td>
<td>.791</td>
<td>.614</td>
<td>.523</td>
<td>.472</td>
<td>--</td>
</tr>
</tbody>
</table>

11. Use of the first two factorial cumulants and the logarithm of the first frequency Here, the vectors ($\kappa_{(1)}$, $\kappa_{(2)}$, log $P_0$) and ($\lambda_1$, $\lambda_2$) correspond to the vectors $\gamma$ and $\lambda$ respectively. It can be easily seen that
Expression for \( J(\tau, \varphi) \) and \( C(\hat{\delta}) \) are obviously the same as in formulae (73) and (74), excepting that we now substitute for \( \lambda' \)'s and the \( P' \)'s, the values corresponding to the Neyman Type A distribution. The equations for estimating the parameters are obtained by substituting the expressions corresponding to the Neyman Type A for the functions involved in equations (60) as

\[
(\lambda_2, \frac{\lambda_2}{2}, \exp(-\lambda_2) - 1) \hat{G}(t)^{-1} \left( \lambda_1 \lambda_2 - \hat{K}_{11}, \lambda_1 \lambda_2^2 \right) - \hat{K}_{21}, \lambda_1(\exp(-\lambda_2) - 1) - \log \hat{P}_0 = 0 \quad (88)
\]

and

\[
(\lambda_1, \frac{\lambda_1}{2}, \exp(-\lambda_1) - 1) \hat{G}(t)^{-1} \left( \lambda_1 \lambda_2 - \hat{K}_{11}, \lambda_1 \lambda_2^2 \right) - \hat{K}_{21}, \lambda_1(\exp(-\lambda_2) - 1) - \log \hat{P}_0 = 0 \quad (89)
\]

On rearrangement, these equations can be written in the form

\[
\lambda_1 = \frac{(\lambda_2, \frac{\lambda_2}{2}, \exp(-\lambda_2) - 1) \hat{G}(t)^{-1} \left( \hat{K}_{11}, \hat{K}_{12}, \log \hat{P}_0 \right)}{(\lambda_2, \frac{\lambda_2}{2}, \exp(-\lambda_2) - 1) \hat{G}(t)^{-1} \left( \lambda_2, \frac{\lambda_2}{2}, \exp(-\lambda_2) - 1 \right)}
\]

and

\[
\lambda_1 = \frac{(1, 2 \lambda_2, - \exp(-\lambda_2)) \hat{G}(t)^{-1} \left( \hat{K}_{11}, \hat{K}_{12}, \log \hat{P}_0 \right)}{(1, 2 \lambda_2, - \exp(-\lambda_2)) \hat{G}(t)^{-1} \left( \lambda_2, \frac{\lambda_2}{2}, \exp(-\lambda_2) \right)}.
\]
They can be solved along the same lines as in section d.i.

Generalized variance of the estimators $\lambda_1$ and $\lambda_2$ is computed by substituting in equation (65a), the expressions for $J(\gamma, 1)$, $J(\gamma, 0)$ and $C(\theta)$ developed above. Efficiency, evaluated by using this generalized variance and the information in Table 2 is given in Table 7 for certain values of $(\lambda_1, \lambda_2)$.

Table 7. Efficiency of the minimum chi square method of estimation for the Neyman Type A distribution when the characteristics used are the first two moments and the logarithm of the zero frequency

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.999</td>
<td>.999</td>
<td>.999</td>
<td>.992</td>
</tr>
<tr>
<td>.3</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.999</td>
<td>.997</td>
<td>.995</td>
<td>.957</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1.000</td>
<td>.999</td>
<td>.999</td>
<td>.996</td>
<td>.999</td>
<td>.986</td>
<td>.911</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
<td>.999</td>
<td>.999</td>
<td>.995</td>
<td>.987</td>
<td>.974</td>
<td>.836</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>.999</td>
<td>.999</td>
<td>.999</td>
<td>.998</td>
<td>.989</td>
<td>.978</td>
<td>.873</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.999</td>
<td>1.000</td>
<td>.999</td>
<td>1.000</td>
<td>.988</td>
<td>.990</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

Some of the interesting properties of this method are:

(i) This method has efficiency, almost as high as 98\% throughout the region $0 \leq \lambda_1 \leq 0$, $0 \leq \lambda_2 \leq 2$.

(ii) For a given $\lambda_1$, efficiency decreases with $\lambda_2$ but the rate at which it decreases is remarkably small.
(iii) For a fixed $\lambda_2$, the efficiency is high for very small $\lambda_1$, decreases at the beginning with $\lambda_1$, but again starts increasing for large $\lambda_1$ and seems to approach unity as $\lambda_1 \to \infty$.

It is to be noted that in most of the samples that the author has come across, to which a Neyman Type A has given reasonable fit, the values of $\lambda_1$ and $\lambda_2$ have been in the region in which the efficiency is tabulated. The high efficiency therefore indicates that this method may have great practical ability in fitting Neyman Type A to sample data.

iii. Use of the first two factorial cumulants and the ratio of the first two frequencies

Here we have

$$\tau = (\kappa_1, \kappa_2, P_1/P_0) = (\lambda_1 \lambda_2, \lambda_1 \lambda_2^2, \lambda_1 \exp(-\lambda_2))$$

$$\lambda = (\lambda_1, \lambda_2)$$

and

$$t = (\hat{\kappa}_1, \hat{\kappa}_2, \hat{P}_1/\hat{P}_0) .$$

For $\Theta$, we choose

$$\Theta = (\mu_1, \mu_2, P_1/P_0)$$

and consequently, for $\hat{\Theta}$, we choose

$$\hat{\Theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{P}_1/\hat{P}_0) .$$

From (89a) and (89b), it is clear that
\[ J(\tau, \lambda) = \begin{bmatrix} \lambda_2 & \lambda_2 & \exp(-\lambda_2) \\ \lambda_1 & \rho \lambda_1 \lambda_2 & -\lambda_1 \exp(-\lambda_2) \end{bmatrix} \]

The expressions for \( J(\tau, \theta) \) and \( C(\theta) \) are the same as those in formulae (73) and (79) excepting that for the \( \mu \)'s and the \( P \)'s, we now substitute the formulae corresponding to the Neyman Type A distribution. The equations for estimating the parameters \( \lambda_1 \) and \( \lambda_2 \) can be obtained from equations (60) as

\[
\begin{align*}
\lambda_1, 2 \lambda_1 \lambda_2 - \lambda_1 \exp(-\lambda_2) \hat{C}(t)^{-1} (\lambda_1 \lambda_2 - \hat{k}_1, \lambda_1 \lambda_2 \\
- \hat{k}_2, \lambda_1 \exp(-\lambda_2) - \hat{P}_1/\hat{P}_0) = 0 \tag{90}
\end{align*}
\]
and

\[
\begin{align*}
\lambda_2, \hat{k}_1, \lambda_2, \exp(-\lambda_2) \hat{C}(t)^{-1} (\hat{k}_1, \hat{k}_2, \hat{P}_1/\hat{P}_0) = 0 \tag{91}
\end{align*}
\]

which can be written on rearrangement as

\[
\begin{align*}
\lambda_1 &= \frac{(\lambda_2, \lambda_2, \exp(-\lambda_2) \hat{C}(t)^{-1} (\hat{k}_1, \hat{k}_2, \hat{P}_1/\hat{P}_0)}{(\lambda_2, \lambda_2, \exp(-\lambda_2) \hat{C}(t)^{-1} (\hat{k}_1, \hat{k}_2, \hat{P}_1/\hat{P}_0} \\
\lambda_2 &= \frac{(1, 2 \lambda_2, -\exp(-\lambda_2) \hat{C}(t)^{-1} (\hat{k}_1, \hat{k}_2, \hat{P}_1/\hat{P}_0)}{(1, 2 \lambda_2, -\exp(-\lambda_2) \hat{C}(t)^{-1} (\hat{k}_1, \hat{k}_2, \hat{P}_1/\hat{P}_0)} \\
\end{align*}
\]

They can be solved along the same lines as in d.i.

Generalized variance of the estimates of \( \lambda_1 \) and \( \lambda_2 \) is
computed by substituting in equation (65), the expressions for \( J(\gamma, \lambda) \), \( I(\gamma, \phi) \) and \( C(\phi) \), obtainable from the discussion given above. The efficiency, evaluated by using this generalized variance and the information determinant of Table 2 is given in Table 8 for certain values of \( \lambda_1 \) and \( \lambda_2 \). It is to be noted that the efficiency of this method is again smaller than that of the method which uses the first two factorial cumulants and the first proportional frequency.

Table 8. Efficiency of the minimum chi-square method for the Neyman Type A distribution when the statistics used are the first two factorial cumulants and the ratio of the first two frequencies

<table>
<thead>
<tr>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.00</td>
<td>.998</td>
<td>.994</td>
<td>.974</td>
<td>.938</td>
<td>.888</td>
<td>.537</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>1.00</td>
<td>.998</td>
<td>.993</td>
<td>.970</td>
<td>.927</td>
<td>.807</td>
<td>.435</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.999</td>
<td>.998</td>
<td>.993</td>
<td>.970</td>
<td>.930</td>
<td>.870</td>
<td>.420</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.00</td>
<td>.997</td>
<td>.991</td>
<td>.968</td>
<td>.934</td>
<td>.986</td>
<td>.448</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>1.00</td>
<td>.998</td>
<td>.987</td>
<td>.960</td>
<td>.930</td>
<td>.992</td>
<td>.532*</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>1.00</td>
<td>.994</td>
<td>.987</td>
<td>.947</td>
<td>.937</td>
<td>.899</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

Since the equations for this method are not much simpler than those for the latter method, there appears to be no merit in using this method.

It will be shown in Chapter VII, that according to e
certain definition of the reliability of statistics, the method using a ratio is poorer than the one using one of the frequencies. In view of this, the present result looks quite interesting.
V. EFFICIENCY OF ESTIMATORS FOR CERTAIN THREE PARAMETER FAMILIES

A. Introductory Remarks

Certain compound and generalized distributions have been studied by Katti [15] for their shape and flexibility. A family which has been found to be particularly useful is the "Poisson generalized Pascal" distribution, denoted by Poisson v Pascal. It is obtained by generalizing (see, for example, Katti [15]) the Poisson distribution through the Pascal distribution. In terms of the well-known example of the egg masses and the larvae, the model underlying this distribution is equivalent to assuming that the distribution of egg masses is Poisson and that the distribution of the surviving larvae within an egg mass is a Pascal. It has been shown that this distribution includes the Neyman Type A and the Negative Binomial distributions as limiting cases and that for finite non-zero values of the parameters (as they occur in the standard form to be given in the next section), its shape, as judged by skewness, kurtosis and the ratio of the first two frequencies, lies in between those of the Neyman Type A and the Pascal distributions. Since the latter distributions have been extensively used in the literature and are found to represent a number of biological populations fairly well, it is believed that the Poisson v Pascal distribution which forms a sort of a bridge between them can be
used with advantage in practical situations.

When the mortality rate within an egg mass is high or when the insects are capable of travelling large distances, a better approximation for the distribution is the Poisson v Binomial distribution obtained by generalizing the Poisson distribution through the Binomial. This is essentially equivalent to saying that (i) eggs are distributed as a Poisson and (ii) the probability that a larva hatched in an egg mass survives and is found in a plot spread randomly on the field is a Binomial.

In this chapter, it is intended to evaluate for the Poisson v Pascal and the Pascal v Poisson distributions, the asymptotic efficiency of the following methods of estimation: (i) the method of the first three moments, (ii) the method of the first two moments and the first frequency and (iii) the method of the first two moments and the ratio of the first two frequencies. We will show in section D.1 that when one is dealing with large samples, the problem of estimating the three parameters of the Poisson v Binomial reduces essentially to that of estimating only two parameters and discuss the asymptotic efficiency of estimating these two parameters by using (i) the first two moments, (ii) the first moment and the first frequency and (iii) the first moment and the ratio of the first two frequencies.
B. Asymptotic Efficiency for the Poisson v Pascal Distribution

1. Preliminary remarks

The formula
\[ g(z) = \exp \left\{ \lambda \left[ (q - pz)^{-k} - 1 \right] \right\} \quad k > 0, \ p > 0, \ q = 1 + p \] (1) will be taken as the standard form for the probability generating function of this distribution. In order to compute efficiency, the information determinant will be evaluated at the outset and the generalized variances will be calculated for the different methods individually in the subsequent sections. Efficiency is then obtained by using the formula:

\[ \text{Efficiency} = \frac{1}{(\text{Generalized variance} \times \text{Information})}. \] (2)

2. Information determinant

It is easy to show (see, for example, Shenton [12]) that the information determinant for an arbitrary integral valued distribution can be written in the form
where \( P_r \) is the probability of obtaining a count \( r \). The problems involved in evaluating the information determinant, therefore, are (i) evaluating the derivatives of \( P_r \) and (ii) determining the number of terms to be used in calculating the infinite series. These are dealt with in the next two sections.

a. Derivatives of \( P_r \)

1. \( \frac{\partial P_r}{\partial \lambda} \)

We have

\[
g(z) = \exp \left\{ \lambda \left[ (a - pz)^{-k} - 1 \right] \right\}.
\] (4)

Hence

\[
\frac{\partial}{\partial \lambda} g(z) = g(z) \left\{ (a - pz)^{-k} - 1 \right\}.
\] (5)

Since

\[
\frac{\partial}{\partial z} g(z) = \lambda k p g(z) (a - pz)^{-k-1}
\] (6)

we can rewrite (5) in the form
\[ \frac{\partial}{\partial z} g(z) = -g(z) + \frac{c}{\lambda kp} \frac{\partial g(z)}{\partial z} - \frac{1}{\lambda k} \frac{\partial}{\partial z} g(z). \]  

On differentiating (7) \( r \) times, we obtain

\[ \frac{\partial^r}{\partial z^r} g(z) = -\frac{\partial^r}{\partial z^r} g(z) + \frac{c}{\lambda kp} \frac{\partial^{r+1}}{\partial z^{r+1}} g(z) \]

\[ - \frac{1}{\lambda k} \left\{ z \frac{\partial^{r+1}}{\partial z^{r+1}} g(z) + r \frac{\partial^r}{\partial z^r} g(z) \right\}. \]  

Take limit \( z \to 0 \) in (8). From the form of \( g(z) \), it is apparent that limits and derivatives are interchangeable, i.e., we can write

\[ \lim_{z \to 0} \frac{\partial^r}{\partial z^r} g(z) = \frac{\partial}{\partial \lambda} \lim_{z \to 0} \frac{\partial^r}{\partial z^r} g(z). \]  

Hence, we get on simplification

\[ \frac{\partial}{\partial r} P_r = -P_r + \frac{c}{\lambda kp} (r + 1) P_{r+1} - \frac{r P_r}{\lambda k}, \quad r = 0, 1, \ldots. \]  

11. \( \frac{\partial P_r}{\partial p} \) First, note that

\[ \frac{\partial}{\partial p} g(z) = \lambda k g(z) (a - pz)^{-k-1} (z - 1) \]

\[ = \frac{1}{P} \left\{ z \frac{\partial g(z)}{\partial z} - \frac{\partial g(z)}{\partial z} \right\}. \]  

and that

\[ \frac{\partial^r}{\partial z^r} \frac{\partial}{\partial p} g(z) = \frac{1}{P} \left\{ (z - 1) \frac{\partial^{r+1}}{\partial z^{r+1}} g(z) + r \frac{\partial^r}{\partial z^r} g(z) \right\}. \]  

By taking limit as \( z \to 0 \), and using arguments similar to those in (i), we obtain
\[
\frac{\partial}{\partial k} p_r = \frac{1}{p_r} \left\{ - (r + 1) p_{r+1} + r p_r \right\} \quad r = 0, 1, \ldots . \tag{13}
\]

iii. \( \frac{\partial p_r}{\partial k} \) Differentiation of \( p(z) \) with respect to \( k \) yields

\[
\frac{\partial}{\partial k} p(z) = -\left\{ \log(p - p z) \right\} \frac{a - \frac{dz}{k}}{k} \frac{\partial}{\partial z} p(z) , \tag{14}
\]

differentiation of \( p(z) \) with respect to \( k \) yields

\[
\frac{\partial}{\partial k} \log(p) = -\sum_{i=1}^{\infty} \left( \frac{p}{q} \right)^i \frac{1}{i} \log(p) + \frac{a - \frac{dz}{k}}{k} \frac{\partial}{\partial z} p(z) ,
\]

which can be rewritten in the form

\[
\frac{\partial}{\partial k} g(z) = -\left\{ \log(p) - \sum_{i=1}^{\infty} \left( \frac{p}{q} \right)^i \frac{1}{i} \right\} \frac{a - \frac{dz}{k}}{k} \frac{\partial}{\partial z} p(z) .
\tag{15}
\]

By equating the coefficients of \( z^r \) on the two sides we obtain

\[
\frac{\partial}{\partial k} p_r = \frac{1}{k p} \sum_{i=1}^{r} \left( \frac{p}{q} \right)^i \frac{1}{i} \left\{ a(r - i + 1) p_{r+i+1} - p p_{r+i} \right\} - \log(p) \frac{1}{k p} \left\{ a(r + 1) p_{r+1} - p r p_r \right\} \quad r = 0, 1, \ldots .
\tag{16}
\]

For the sake of reference we give here the recurrence formulae which may be used in obtaining the probabilities involved in the various stages of computing the information determinant:

\[
P_0 = \exp \left\{ \lambda \left[ a^{k} - 1 \right] \right\}
\tag{16a}
\]

\[
P_{r+1} = \frac{k a^{-k}}{r + 1} \sum_{i=0}^{r} \frac{(k+1)(k+2)\ldots(k+r-1)}{(r-1)!} \left( \frac{p}{q} \right)^{r+1} p_i .
\tag{16b}
\]

For derivation, refer to section V.C. of Kastli [15].
b. **Number of terms**  
As in Chapter IV, we reject the idea of fixing the number of terms to be used in calculating the infinite series and seek for a rule by which the value of the determinant can be obtained correct to three significant digits. A rule that looks reasonable is the following:

Let \( \lambda, p, \) and \( k \) be referred to as the first, second and the third parameter respectively. Denote by \( T_n(i,k) \), the \( n \)th term in the series involving the product of the derivatives of \( P_n \) with respect to the \( i \)th parameter and the \( k \)th parameter. Let

\[
R = \sum_{i,k} \frac{T_n(i,k)^2}{\left( \sum_{j=0}^{n} T_j(i,k) \right)^2} \quad (17)
\]

Starting from \( n = 0 \), calculate the value of each of the series involved in the determinant successively by taking the first \( n \) terms, calculate \( R \) and when \( R \) becomes less than a pre-assigned constant \( K \), take the value of \( \sum_{j=0}^{n} T_j(i,k) \) as the value of the corresponding series and go on to compute the determinant.

It is apparent from (17) that \( R \leq K \) implies that

\[
\frac{T_n(i,k)^2}{\left( \sum_{j=1}^{n} T_j(i,k) \right)^2} \leq k \quad \text{all } (i,k).
\]

Hence
We take for \( k \), the value \( 10^{-8} \). As in Chapter IV, section C.2.b, we observe that if the terms in all the series start decreasing at a fast rate after the \( n \)th term so that the remainder term in each one of them is less than a multiple \( m \) of \( T_n(i,k) \), then the relative error

\[
\left| \frac{T_n(i,k)}{\sum_{j=1}^{n} T_j(i,k)} \right| < \sqrt{k} \quad \text{all} \ (i,k)
\]

will be bounded by \( m\sqrt{k} \). If \( m < 10 \) and if the significant figures do not cancel out, the computed value of the information determinant will be correct to three decimal places and consequently, the efficiency calculated by using this information determinant will be correct to three significant figures. Since efficiency is less than 1, this implies that the computed value of the efficiency is correct to three decimal places.

It was found in the course of computing that the number of terms needed according to this rule was very large when the parameters were large and the time taken by the IBM 650 to compute the efficiency rose steeply. The cost of computing forced us to put a ceiling of 20 over the number of terms used. The efficiencies obtained by using the computed in-
formation determinants when the series are calculated using 20 terms will be marked with an asterisk. Obviously, the figures therein are not necessarily correct to three figures. It is hoped that they will serve at least as close upper bounds. When the upper bounds so calculated have slightly exceeded 1.0, they have been replaced by 1. When these upper bounds are much higher than 1, the corresponding cells in the efficiency tables have been left blank. A flow chart used for these computations is given in Figure 9.

3. Generalized variance and efficiency for the method of three moments

The first six factorial cumulants for this distribution (cf. Katti 15) are given by

\[ \kappa_{11} = \lambda p^1 k(k + 1) \ldots (k + i - 1) \quad i = 1, 2, \ldots, 6. \]

By referring to section D.9 of Chapter III, we first note that the vectors \( (\kappa_1, \kappa_2, \kappa_{[3]}) \) and \( (\lambda, p, k) \) to \( \gamma \) and \( \lambda \) respectively. It is clear that

\[
J(\gamma, \lambda) = \begin{pmatrix}
\kappa_{[1]} / \lambda & \kappa_{[2]} / \lambda & \kappa_{[3]} / \lambda \\
\kappa_{[1]} / P & 2 \kappa_{[2]} / P & 3 \kappa_{[3]} / P \\
\lambda P & \lambda p^2 (2k+1) & \lambda p^3 (3k^2+6k+2)
\end{pmatrix}
\]

From corollary 2 of section E of Chapter III, we note that the generalized variance of \( (\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_{[3]}) \) is the same as
Figure 9. Flow-chart for obtaining the information determinant for three parameter families.
that of \((\hat{\lambda}_1^1, \hat{\lambda}_2^1, \hat{\lambda}_3^1)\). An expression for the covariance matrix of \((\hat{\lambda}_1^1, \hat{\lambda}_2^1, \hat{\lambda}_3^1)\) is given in equation (67), Chapter IV. To evaluate this generalized variance, the first six factorial cumulants are first calculated using formula (18) and then the first six moments calculated using the equations in section III.F. The value of generalized variance of the estimates of the parameters obtained by this method was computed by substituting for \(J(\gamma, \lambda), J(\gamma, \beta)\) and \(C(\beta)\) in equation (34) of section D, Chapter III.

The efficiency of this method is given in Tables 9, 10, 11, 12 and 13. It should be noted that the efficiency is high in the very close neighborhood of the origin in the parameter space and the efficiency falls off rapidly with \(p\) and \(k\) increasing. The efficiency appears to be relatively less affected by an increase in \(\lambda\). However, it can be shown

<table>
<thead>
<tr>
<th>(p)</th>
<th>(k)</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.843</td>
<td>0.824</td>
<td>0.824</td>
<td>0.815</td>
<td>0.762</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.592</td>
<td>0.589</td>
<td>0.583</td>
<td>0.549</td>
<td>0.479</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.454</td>
<td>0.441</td>
<td>0.434</td>
<td>0.400</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.260</td>
<td>0.256</td>
<td>0.248</td>
<td>0.226</td>
<td>0.185</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>0.130</td>
<td>0.129</td>
<td>0.129</td>
<td>0.128</td>
<td>0.134</td>
<td></td>
</tr>
</tbody>
</table>
Table 10. Efficiency of the method of the first three moments for the Poisson v Pascal distribution at $\lambda = 0.5(1)$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.906</td>
<td>.812</td>
<td>.828</td>
<td>.779</td>
<td>.674</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.593</td>
<td>.587</td>
<td>.561</td>
<td>.499</td>
<td>.369</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.460</td>
<td>.440</td>
<td>.416</td>
<td>.348</td>
<td>.224</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.265</td>
<td>.257</td>
<td>.239</td>
<td>.185</td>
<td>.104</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.131</td>
<td>.131</td>
<td>.126</td>
<td>.911*</td>
<td>.993*</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Efficiency of the method of the first three moments for the Poisson v Pascal distribution at $\lambda = 1.0(1)$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.815</td>
<td>.835</td>
<td>.821</td>
<td>.751</td>
<td>.634</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.598</td>
<td>.591</td>
<td>.554</td>
<td>.463</td>
<td>.289</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.463</td>
<td>.443</td>
<td>.409</td>
<td>.312</td>
<td>.155</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.270</td>
<td>.256</td>
<td>.232</td>
<td>.154</td>
<td>.057</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.882*</td>
<td>.896*</td>
<td>.914*</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
### Table 12. Efficiency of the method of the first three moments for the Poisson vs Pascal at $\lambda = 5.0$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$k$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.1</td>
<td>.940*</td>
<td>.934*</td>
<td>.783*</td>
<td>.583*</td>
<td>.219*</td>
</tr>
<tr>
<td>.3</td>
<td>.1</td>
<td>.627*</td>
<td>.566*</td>
<td>.411*</td>
<td>.115*</td>
<td>.019*</td>
</tr>
<tr>
<td>.5</td>
<td>.1</td>
<td>.488*</td>
<td>.386*</td>
<td>.202*</td>
<td>.040*</td>
<td>.007*</td>
</tr>
<tr>
<td>1.0</td>
<td>.1</td>
<td>.296*</td>
<td>.175*</td>
<td>.064*</td>
<td>.023*</td>
<td>.003*</td>
</tr>
<tr>
<td>2.0</td>
<td>.1</td>
<td>.905*</td>
<td>.986*</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

### Table 13. Efficiency of the method of the first three moments for the Poisson vs Pascal distribution at $\lambda = 10.0$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$k$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.1</td>
<td>.906*</td>
<td>.764*</td>
<td>.725*</td>
<td>.034*</td>
<td>.043*</td>
</tr>
<tr>
<td>.3</td>
<td>.1</td>
<td>.645*</td>
<td>.340*</td>
<td>.121*</td>
<td>.019*</td>
<td>.080*</td>
</tr>
<tr>
<td>.5</td>
<td>.1</td>
<td>.491*</td>
<td>.148*</td>
<td>.041*</td>
<td>.018*</td>
<td>.026*</td>
</tr>
<tr>
<td>1.0</td>
<td>.1</td>
<td>.284*</td>
<td>.045*</td>
<td>.015*</td>
<td>.013*</td>
<td>.012*</td>
</tr>
<tr>
<td>2.0</td>
<td>.1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
analytically that the efficiency approaches zero as \( \lambda \to \infty \).

4. **Generalized variance and efficiency of the method of the first two moments and the zero frequency**

First, it is to be observed that

\[
P_0 = \exp \left\{ \lambda \left[ \sigma^{-k} - 1 \right] \right\} \quad (92)
\]

\[
\frac{\partial P_0}{\partial \lambda} = \sigma^{-k} P_0 \quad (93)
\]

\[
\frac{\partial P_0}{\partial \sigma} = -\lambda k \sigma^{-k-1} P_0 \quad (94)
\]

and

\[
\frac{\partial P_0}{\partial \sigma} = -\lambda \log(\sigma) \sigma^{-k} P_0 \quad (95)
\]

On referring to section D of Chapter III, we note that the vectors \( (\kappa_{11}, \kappa_{12}, P_0) \) and \( (\lambda, k, \sigma) \) correspond to vectors \( \gamma \) and \( \lambda \). The value of the Jacobian \( J(\gamma, \lambda) \) can then be obtained from these and other results in section 3 above in the form

\[
J(\gamma, \lambda) = \begin{bmatrix}
k_{11}/\lambda & k_{12}/\lambda & \sigma^{-k} P_0 \\
k_{11} P & k_{12} \gamma/P & -\lambda k \sigma^{-k-1} P_0 \\
\lambda \sigma & \lambda \sigma^{2}(2k+1) & -\lambda \log(\sigma) \sigma^{-k} P_0
\end{bmatrix}
\]

For \( \theta \), we choose \( (\mu_1, \mu_2, P_0) \). From section D.5 of Chapter III, it is clear that \( |C(t)| = |C(\hat{\theta})| \). An expression for \( C(\hat{\theta}) \) is given in equation (73) of Chapter IV.
The generalized variance of the estimates can then be obtained from formula (34) in section III.D. The efficiency is given in Tables 14, 15, 16, 17 and 18. It is to be noted that the efficiency of this method is high and is considerably higher than that of the method of the three moments for even sufficiently large values of the parameters. However, as any of the parameters approaches infinity, the efficiency does approach zero.

5. Generalized variance and efficiency for the ratio of the first two frequencies

It can be easily shown that

\[ \frac{P_1}{P_0} = \lambda \frac{q^{q-k-1}}{q-k-1} \]  

\[ \frac{P_1}{P_0} = k \lambda \frac{q^{q-k-1}}{q-k-1} \]  

\[ \frac{P_1}{P_0} = \lambda k(q-k-1 - (k+1)p\alpha^{-k}) \]  

and

\[ \frac{P_1}{P_0} = \lambda p q^{-k-1} - \lambda k p(\log \sigma) q^{-k} \]  

The vectors \((\gamma_1, \gamma_2, P_1/P_0)\) and \((\lambda, \lambda, \sigma)\) correspond to the vectors \(\gamma\) and \(\lambda\) respectively. The value of the Jacobian \(J(\gamma, \lambda)\) obtained from these and other results of section 3, is
Table 14. Efficiency of the method of the first two moments and the first frequency for the Poisson v Pascal distribution at $\lambda = 0.1$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.1</td>
<td>.995</td>
<td>.980</td>
<td>.988</td>
<td>.987</td>
<td>.980</td>
</tr>
<tr>
<td>.3</td>
<td>.932</td>
<td>.942</td>
<td>.950</td>
<td>.943</td>
<td>.937</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.908</td>
<td>.902</td>
<td>.911</td>
<td>.912</td>
<td>.899</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.829</td>
<td>.836</td>
<td>.840</td>
<td>.843</td>
<td>.833</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.748</td>
<td>.765</td>
<td>.787</td>
<td>.841</td>
<td>.786</td>
<td></td>
</tr>
</tbody>
</table>

Table 15. Efficiency of the method of the first two moments and the first frequency for the Poisson v Pascal distribution at $\lambda = 0.5$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.934</td>
<td>.948</td>
<td>.938</td>
<td>.923</td>
<td>.979</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.918</td>
<td>.911</td>
<td>.905</td>
<td>.890</td>
<td>.829</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.838</td>
<td>.850</td>
<td>.850</td>
<td>.820</td>
<td>.739</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.752</td>
<td>.788</td>
<td>.812</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
Table 16. Efficiency of the method of the first two moments and the first frequency for the Poisson v Pascal distribution at $\lambda = 1.0$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000</td>
<td>.995</td>
<td>.975</td>
<td>.966</td>
<td>.947</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.940</td>
<td>.961</td>
<td>.947</td>
<td>.931</td>
<td>.973</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.922</td>
<td>.925</td>
<td>.918</td>
<td>.890</td>
<td>.918</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.849</td>
<td>.859</td>
<td>.860</td>
<td>.822</td>
<td>.797</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.882*</td>
<td>.878*</td>
<td>.879*</td>
<td>.871*</td>
<td>.831*</td>
<td></td>
</tr>
</tbody>
</table>

Table 17. Efficiency of the method of the first two moments and the first frequency for the Poisson v Pascal distribution at $\lambda = 5.0$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>--</td>
<td>--</td>
<td>.998*</td>
<td>.983*</td>
<td>.983*</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.976*</td>
<td>.994*</td>
<td>1.000*</td>
<td>1.000*</td>
<td>.985*</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.958*</td>
<td>.979*</td>
<td>.999*</td>
<td>.994*</td>
<td>.966*</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.906*</td>
<td>.962*</td>
<td>.979*</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.893*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
Table 18. Efficiency of the method of the first two moments and the first frequency for the Poisson v Pascal distribution at \( \lambda = 10.0 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( k )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000*</td>
<td>.951*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>1.000*</td>
<td>.999*</td>
<td>--</td>
<td>.9852*</td>
<td>.9856*</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.996</td>
<td>.951*</td>
<td>.901*</td>
<td>.889*</td>
<td>.968*</td>
<td></td>
</tr>
</tbody>
</table>

We choose for \( \hat{\theta} \), the vector \((k_1^1, k_2^1, \lambda P_1/P_0)\). As in section 4 above, we observe that \( C(t) = C(\hat{\theta}) \). An expression for \( C(\hat{\theta}) \) appears in equation (79) of Chapter IV. The computing of the generalized variance of the estimates, therefore, follows from formula (34) of Chapter III.

Efficiency of the estimates is given for certain values of \((\lambda, p, k)\) in Tables 19, 20, 21, 22 and 23. It is to be noted that the efficiency of this method is higher than that of the method of the first two moments and the first frequency in the region \( 0 < \lambda < 10, 0 < p < 2 \) and \( 0 < k < 2 \) in which the efficiency is tabulated. It can be shown analytically...
Table 19. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Poisson v Pascal at $\lambda = 0.1$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.989</td>
<td>.997</td>
<td>1.004</td>
<td>.994</td>
<td>.985</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.991</td>
<td>.980</td>
<td>.986</td>
<td>.979</td>
<td>.942</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.945</td>
<td>.942</td>
<td>.936</td>
<td>.914</td>
<td>.838</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.877</td>
<td>.885</td>
<td>.890</td>
<td>.911</td>
<td>.993</td>
<td></td>
</tr>
</tbody>
</table>

Table 20. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Poisson v Pascal distribution at $\lambda = 0.5$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.001</td>
<td>.990</td>
<td>1.000*</td>
<td>1.000</td>
<td>.991</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.989</td>
<td>1.000</td>
<td>.991</td>
<td>.984</td>
<td>.960</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.999</td>
<td>.985</td>
<td>.977</td>
<td>.967</td>
<td>.921</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.949</td>
<td>.948</td>
<td>.939</td>
<td>.899*</td>
<td>.801*</td>
<td></td>
</tr>
</tbody>
</table>
Table 21. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Poisson v Pascal distribution at $\lambda = 1.0$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$k$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td></td>
<td>.984</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.998</td>
</tr>
<tr>
<td>.3</td>
<td>.994</td>
<td>1.000</td>
<td>.997</td>
<td>.990</td>
<td>.957</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1.000</td>
<td>.994</td>
<td>.984</td>
<td>.963</td>
<td>.913</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.955</td>
<td>.943</td>
<td>.934</td>
<td>.888</td>
<td>.798</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.882*</td>
<td>.896*</td>
<td>.914*</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

Table 22. Efficiency of the method of first two moments and the ratio of the first two frequencies for the Poisson v Pascal distribution at $\lambda = 5.0$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$k$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td></td>
<td>--</td>
<td>--</td>
<td>.994*</td>
<td>.959*</td>
<td>.918*</td>
</tr>
<tr>
<td>.3</td>
<td>1.000*</td>
<td>.993*</td>
<td>.990*</td>
<td>.992*</td>
<td>.766*</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1.000*</td>
<td>.961*</td>
<td>.915*</td>
<td>.806*</td>
<td>.650*</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.952*</td>
<td>.891*</td>
<td>.798*</td>
<td>.845*</td>
<td>1.000*</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.905*</td>
<td>.986*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
Table 23. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Poisson v Pascal distribution at $\lambda = 10.0$

<table>
<thead>
<tr>
<th>p</th>
<th>k</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td></td>
<td>1.000*</td>
<td>.932*</td>
<td>1.000*</td>
<td>.907*</td>
<td>.769*</td>
</tr>
<tr>
<td>.3</td>
<td></td>
<td>1.000*</td>
<td>.916*</td>
<td>.844*</td>
<td>.575*</td>
<td>.464*</td>
</tr>
<tr>
<td>.5</td>
<td></td>
<td>.986*</td>
<td>.843*</td>
<td>.681*</td>
<td>.432*</td>
<td>--</td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td>.928*</td>
<td>.696*</td>
<td>.603*</td>
<td>.999*</td>
<td>--</td>
</tr>
<tr>
<td>2.0</td>
<td></td>
<td>.901*</td>
<td>--</td>
<td>--</td>
<td>.991*</td>
<td>--</td>
</tr>
</tbody>
</table>

that this method is less efficient than the method of the first two moments and the first frequency for large values of $\lambda$, $p$ and $k$. The proof is too cumbersome to be included here. However, in view of the fact that the equations for obtaining the estimates by this method are considerably simpler than those of the method of the first two moments and the first frequency (cf. Ketti [15]), this method looks very promising in fitting this distribution to sample data when the true values of the parameters are not too large.

C. Asymptotic Efficiency for the Pascal v Poisson Distribution

1. Preliminary remarks

The formula $\left(q - p \exp\left[\lambda(z - 1)\right]\right)^{-k}$, $k > 0$, $\lambda > 0$, $p > 0$, $q = 1 + p$ will be taken as the standard form for the
probability generating function of this distribution. As in section B, we will again calculate the information determinant separately at the outset. In the subsequent sections, generalized variances will be evaluated for the various methods of estimation and the efficiency obtained by using formula (2) of section B. Since the form of the efficiency functions for this distribution are very similar to the corresponding ones for the Poisson v Pascal distribution, the tables of efficiency will be restricted to much smaller ranges of the values of the parameters.

2. Information determinant

As in section B.2, the principal problems here again are (i) to evaluate the derivatives of the probabilities and (ii) to determine the number of terms to be used in summing the various infinite series involved. As for the latter, we will use the rule mentioned in section B.2 for determining the number of terms. Formulae for obtaining the derivatives of the probabilities are given below.

a. Derivatives of the probabilities

\[ \frac{\partial P_r}{\partial \lambda} \]

Let

\[ f(z) = \left\{ q - p \exp[\lambda (z - 1)] \right\}^{-k}. \]  

(31)

Then

\[ \frac{\partial}{\partial \lambda} g(z) = kp(z - 1) \left\{ q - p \exp[\lambda (z - 1)] \right\}^{-k} \exp \{\lambda (z - 1)\}. \]  

(32)
By differentiating equation (1) with respect to \( z \), we obtain
\[
\frac{\partial g(z)}{\partial z} = kp\left\{ \frac{1}{\lambda} \left[ \frac{z}{\lambda} \frac{\partial}{\partial z} - \frac{1}{\lambda} \right] \right\} - k^{1}\text{exp}\left\{ \frac{1}{\lambda} \left( z - 1 \right) \right\} .
\] (33)

Hence equation (3) can be simplified to
\[
\frac{\partial g(z)}{\partial z} = \frac{1}{\lambda} \left\{ \frac{z}{\lambda} \frac{\partial}{\partial z} - \frac{1}{\lambda} \right\} .
\] (34)

On equating the coefficient of \( z^r \) on the two sides, we get
\[
\frac{\partial P_r}{\partial \lambda} = \frac{1}{\lambda} \left\{ rP_r - (r + 1)P_{r+1} \right\}
\] (35)

\[\text{i.} \quad \frac{\partial P_r}{\partial p} \]

On differentiating \( g(z) \) with respect to \( p \), we get
\[
\frac{\partial g(z)}{\partial p} = -k\left\{ \frac{1}{\lambda} \text{exp}\left\{ \frac{1}{\lambda} \left( z - 1 \right) \right\} \right\} - k^{1}\text{exp}\left\{ \frac{1}{\lambda} \left( z - 1 \right) \right\}
\]
which can be simplified by using relation (33) above to
\[
\frac{\partial r(z)}{\partial p} = -\lambda^2(z) + \frac{1}{\lambda} \frac{\partial}{\partial z} g(z)
\] (36)

where
\[
\lambda(z) = \left\{ \frac{1}{\lambda} \text{exp}\left\{ \frac{1}{\lambda} \left( z - 1 \right) \right\} \right\} - k^{1},
\] (37)
is a Pascal v Poisson distribution with \((k+1)\) for the exponent. Denote the probability of a count \( r \) in this distribution by \( P_{r} \). Then by collecting the coefficients of \( z^r \) on the two sides of equation (6), we obtain
\[
\frac{\partial P_r}{\partial p} = -kP_r + \frac{1}{\lambda} (r + 1)P_{r+1} .
\] (38)

\[\text{iii.} \quad \frac{\partial P_r}{\partial k} \]

On differentiating \( g(z) \) with respect to \( k \), we get
\[
\frac{p g(z)}{g_k} = -\{q - p \exp[\lambda (z - 1)]\}^{-k-1} \log \{q - p \exp[\lambda (z - 1)]\}
\]
\[
= \log \{1 + g^n(z)\}, \text{ say (39)}
\]
where
\[
g^n(z) = 1 - \log \{q - p \exp[\lambda (z - 1)]\}
\]
is a logarithmic Poisson distribution. Denote by \(P^g_r\), the probability of count \(r\) in this distribution. Then, by collecting the coefficient of \(z^r\) on the two sides in (10), we obtain
\[
\frac{2}{\lambda k} P_r = -P_r + \sum_{i=0}^r P_i P^g_{r-1} . \tag{41}
\]
For the sake of completeness, the recurrence formulae for obtaining \(P_r\) and \(P^g_r\) are given below. For their derivation, refer to sections V.C.4 and V.C.5 of Katti 15.
\[
P_0 = \{q - p \exp(-\lambda)\}^{-k} \tag{42}
\]
\[
P_{r+1} = \frac{p \exp(-\lambda)}{q - p \exp(-\lambda)} \frac{1}{r + 1} \left\{ \sum_{i=0}^{r-1} (i + 1) p_{i+1} \frac{\lambda^{(r-1)}}{(r-1)!} \right\} + k \sum_{i=0}^{r-1} p_i \frac{\lambda^{(r-1)}}{(r-1)!} \right\} \quad r = 0, 1, \ldots \tag{43}
\]
\[
P^{g}_{0} = 1 - \log(q - p \exp(-\lambda)) \tag{44}
\]
and
\[
P^{g}_{r+1} = \frac{p}{\lambda} \frac{\lambda^{r+1}}{(r+1)!} + \frac{1}{r + 1} \sum_{i=0}^{r-1} (i+1)p_{i+1} \frac{\lambda^{r-1}}{(r-1)!} . \tag{45}
\]
3. **Generalized variance and efficiency**
   **for the method of the first three moments**

From section IV.D.8 of Katti [15], we have the formulae for the first three factorial cumulants as

\[ \kappa_{1} = \lambda kp \]
\[ \kappa_{2} = \lambda^{2} kp(l + p) \]

and

\[ \kappa_{3} = \lambda^{3} kp(l + p)(1 + kp) \] \hspace{1cm} (46)

Again, we have the vectors \((\kappa_{11}, \kappa_{12}, \kappa_{31})\) and \((\lambda, \rho, k)\) corresponding to \(\gamma\) and \(\beta\) respectively. We derive by simple algebra

\[ j(\gamma, \beta) = \frac{\kappa_{11}}{\lambda} \quad \frac{\kappa_{21}}{\lambda} \quad \frac{3 \kappa_{31}}{\lambda} \]
\[ \frac{\kappa_{11}}{k} \quad \frac{\kappa_{21}}{k} \quad \frac{\kappa_{31}}{k} \]
\[ \frac{\kappa_{11}}{F} \quad \lambda^{2} k(l + p) \quad \lambda^{3} k(l + 2p + 3p^2) \]

As in B.3, we observe that the generalized variance of \((\hat{\kappa}_{11}, \hat{\kappa}_{12}, \hat{\kappa}_{31})\) is the same as that of \((\hat{\lambda}, \hat{\rho}, \hat{k})\) and is given by formula (26) of that section. In order to calculate the first six moments involved therein, we give below a recurrence formula to evaluate the first six factorial cumulants from which the first six moments about the origin can be obtained by using the formulae in section F of Chapter III. The computation of the generalized variance of the estimates of the parameters of this method follows from formula (34) of Chapter III.
Efficiency of this method is given in Tables 24, 25 and 26. It is to be noted that the fall in the efficiency is rapid for increasing \( \lambda \) and \( p \) while the fall is considerably slow for increasing \( k \).

Table 24. Efficiency of the method of the first three moments for the Pascal v Poisson distribution at \( p = .1 \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.1</td>
<td>1.015</td>
<td>1.008</td>
<td>1.027</td>
<td>.974</td>
<td>.510</td>
</tr>
<tr>
<td>.3</td>
<td>.3</td>
<td>.984</td>
<td>.986</td>
<td>.988</td>
<td>.896</td>
<td>.463</td>
</tr>
<tr>
<td>.5</td>
<td>.5</td>
<td>.998</td>
<td>.987</td>
<td>.950</td>
<td>.834</td>
<td>.428</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1.023</td>
<td>1.023</td>
<td>1.081</td>
<td>.709</td>
<td>.363</td>
</tr>
</tbody>
</table>

Table 25. Efficiency of the method of the first three moments for the Pascal v Poisson distribution at \( p = .3 \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.1</td>
<td>.957</td>
<td>.974</td>
<td>.978</td>
<td>.680</td>
<td>.195</td>
</tr>
<tr>
<td>.3</td>
<td>.3</td>
<td>.941</td>
<td>.916</td>
<td>.865</td>
<td>.557</td>
<td>.164</td>
</tr>
<tr>
<td>.5</td>
<td>.5</td>
<td>.923</td>
<td>.851</td>
<td>.787</td>
<td>.472</td>
<td>.147</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>.909</td>
<td>.771</td>
<td>.657</td>
<td>.378</td>
<td>.193</td>
</tr>
</tbody>
</table>
Table 26. Efficiency of the method of the first three moments for the Pascal v Poisson distribution at $p = .5$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\lambda$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.928</td>
<td>.949</td>
<td>.941</td>
<td>.458</td>
<td>.109</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.908</td>
<td>.853</td>
<td>.767</td>
<td>.353</td>
<td>.087</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.890</td>
<td>.788</td>
<td>.666</td>
<td>.298</td>
<td>.076</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.861</td>
<td>.681</td>
<td>.599</td>
<td>.087</td>
<td></td>
<td>.065</td>
</tr>
</tbody>
</table>

2. Recurrence formula for obtaining factorial cumulants

Since the probability generating function is given by

$$f(z) = \frac{z - \alpha \exp\{\lambda (z - 1)\}^{-k}}{1 - \alpha \exp\{\lambda (z - 1)\}}$$  \hspace{1cm} (48)

the factorial cumulant generating function $(\psi)$ is given by

$$\psi'(u) = -k \log\{q - p \exp\{\lambda u\}\}.$$  \hspace{1cm} (49)

On differentiating (19) with respect to $u$, we get

$$\psi'(u) = \frac{kp \lambda \exp\{\lambda u\}}{q - p \exp\{\lambda u\}},$$  \hspace{1cm} (50)

which can be rewritten in the form

$$\psi'(u)\left\{q - p \exp\{\lambda u\}\right\} = kp \lambda \exp\{\lambda u\}.$$  \hspace{1cm} (51)

Successive differentiation $r$ times yields

$$\psi^{(r+1)}(u)\left\{q - p \exp\{\lambda u\}\right\} = \sum_{1=0}^{r-1} \binom{r}{1} \psi^{(r-1)}(u) \exp\{\lambda u\}$$

$$= kp \lambda^{r+1}.$$  \hspace{1cm} (52)

On setting $u = 0$ in (50) and (52) and observing that the
116

\( i \) th factorial cumulant \( \kappa_{(i)} \) is given by

\[
\kappa_{(i)} = \left( \frac{d}{du} \Phi(u) \right)_{u=0}.
\]

We get

\[ \kappa^2_{(1)} = kp \lambda \]

and

\[ \kappa_{(r+1)} = kp \lambda^{r+1} + \sum_{i=0}^{r-1} \binom{r}{i} \lambda^r \lambda^{-i-1} \kappa_{(i+1)} \]

\( r = 1, 2, \ldots \). (53)

4. Generalized variance and efficiency
   for the method of the first two moments and the first frequency

We first note that

\[ P_0 = \left\{ q - p \exp(-\lambda) \right\}^{-k} \] (54)

\[ \frac{\partial P_0}{\partial \lambda} = -kp \left\{ q - p \exp(-\lambda) \right\}^{-k-1} \] (55)

\[ \frac{\partial P_0}{\partial p} = k \exp(-\lambda) \left\{ q - p \exp(-\lambda) \right\}^{-k-1} \] (56)

and

\[ \frac{\partial P_0}{\partial k} = -P_0 \log \left\{ q - p \exp(-\lambda) \right\}. \] (57)

The value of the Jacobian \( J(\gamma, \lambda) \) can then be obtained from these and other results obtained in section 3.

From the result of section III.E.5, the generalized variance of the first two factorial cumulants and the first frequency is the same as that of the first two moments about the origin and the first frequency. An expression for the
covariance matrix is given in formula (734) of Chapter IV. The computing of the generalized variance then follows from formula (11) of section III.D.

The efficiency of the estimates is given in Tables 27, 28 and 29. The efficiency of this method for small \( \nu \) is remarkably high. For \( p = .1 \), the efficiency exceeds 97% for

### Table 27. Efficiency of the method of the first two moments and the first frequency for the Pascal \( v \) Poisson distribution at \( p = 0.1 \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.1</td>
<td>.956</td>
<td>1.000</td>
<td>1.000</td>
<td>.999</td>
<td>.967</td>
</tr>
<tr>
<td>.3</td>
<td>.944</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.975</td>
</tr>
<tr>
<td>.5</td>
<td>.963</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.980</td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.998</td>
<td>.974</td>
<td></td>
</tr>
</tbody>
</table>

### Table 28. Efficiency of the method of the first two moments and the first frequency for the Pascal \( v \) Poisson distribution at \( p = 0.3 \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.985</td>
<td>1.000</td>
<td>1.000</td>
<td>.992</td>
<td>.959</td>
<td>.902</td>
</tr>
<tr>
<td>.3</td>
<td>.988</td>
<td>1.000</td>
<td>.980</td>
<td>.952</td>
<td>.882</td>
<td>.882</td>
</tr>
<tr>
<td>.5</td>
<td>.993</td>
<td>.983</td>
<td>.976</td>
<td>.937</td>
<td>.880</td>
<td>.880</td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
<td>.977</td>
<td>.956</td>
<td>.924</td>
<td>.855</td>
<td></td>
</tr>
</tbody>
</table>
the whole of the region of tabulation, namely \( 0 < \lambda < 2 \) and \( 0 < k < 1 \). While this method retains its superiority even for \( p = .5 \), the drop in the efficiency from \( p = .1 \) to \( p = .5 \) is considerable for a given \( p \); the change in the efficiency is gradual with increasing \( k \) as well as \( \lambda \).

5. **Generalized variance and efficiency for the method of the first two moments and the ratio of the first two frequencies**

We first observe that

\[
P_1/P_0 = \frac{\lambda^k \rho}{\alpha \exp(\lambda) - \rho}
\]  

\[
\frac{\partial}{\partial \lambda} \left( P_1/P_0 \right) = - \left( P_1/P_0 \right) \left\{ \frac{\rho}{\alpha \exp(\lambda) - \rho} \right\}
\]  

\[
\frac{\partial}{\partial k} \left( P_1/P_0 \right) = \left( P_1/P_0 \right)/k
\]  

\[
\frac{\partial}{\partial \rho} \left( P_1/P_0 \right) = \left( P_1/P_0 \right) \left\{ \frac{1}{\rho} - \frac{\exp(\lambda)}{\alpha \exp(\lambda) - \rho} \right\}
\]
Since the evaluating of the Jacobian $J(\gamma, \lambda)$ and the generalized variance of $\hat{\kappa}_1, \hat{\kappa}_2, \hat{P}_1/\hat{P}_0$ follows along lines very similar to those in section 4, no further elaboration of it need be made here. The efficiency is given in Tables 30, 31 and 32. It is to be noted that, in contrast with the Poisson v Pascal distribution, the efficiency of this method

**Table 30. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Pascal v Poisson distribution at $p = .1$**

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\lambda$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000</td>
<td>.970</td>
<td>.943</td>
<td>.867</td>
<td>.659</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.990</td>
<td>.977</td>
<td>.955</td>
<td>.885</td>
<td>.679</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1.000</td>
<td>.983</td>
<td>.964</td>
<td>.900</td>
<td>.696</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
<td>.994</td>
<td>.981</td>
<td>.915</td>
<td>.722</td>
<td></td>
</tr>
</tbody>
</table>

**Table 31. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Pascal v Poisson distribution at $p = .3$**

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\lambda$</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000</td>
<td>1.000</td>
<td>.984</td>
<td>.923</td>
<td>.724</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>1.000</td>
<td>1.000</td>
<td>.989</td>
<td>.935</td>
<td>.734</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>1.000</td>
<td>.999</td>
<td>.994</td>
<td>.931</td>
<td>.747</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
<td>1.000</td>
<td>.988</td>
<td>.945</td>
<td>.764</td>
<td></td>
</tr>
</tbody>
</table>
Table 32. Efficiency of the method of the first two moments and the ratio of the first two frequencies for the Pascal v Poisson distribution at \( p = .5 \)

<table>
<thead>
<tr>
<th>k</th>
<th>( \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.000</td>
</tr>
<tr>
<td>.3</td>
<td>1.000</td>
</tr>
<tr>
<td>.5</td>
<td>1.000</td>
</tr>
<tr>
<td>1.0</td>
<td>1.000</td>
</tr>
</tbody>
</table>

falls rapidly with increasing \( k \), \( p \) and \( k \) and is much less efficient than the method of the first two moments and the first frequency for even moderate values of the parameters.

D. Asymptotic Efficiency for the Poisson v Binomial Distribution

1. Preliminary remarks

The formula \( \left\{ q + p \exp\left[ \lambda (z - 1) \right]\right\}^n, \lambda > 0, 0 \leq p \leq 1, u = 1 - p \) and \( n \), a positive integer, will be taken as the standard form for the probability generating function of this distribution. Since \( n \) takes on discrete values, a consistent estimate of \( n \) will equal the true value with probability 1, asymptotically \( (\text{cf. Hammersley [14]} \). Hence, the problem of estimation in large samples essentially consists of estimating \( p \) and \( \lambda \). In this perspective, we will consider \( n \) as known and evaluate the asymptotic efficiency of the following
methods of estimating $p$ and $\lambda$: (i) method of the first two moments, (ii) method of the first moment and the first frequency and (iii) method of the first moment and the ratio of the first two frequencies.

2. Information determinant

It is to be noted that the probability generating function of the Poisson v Binomial distribution can be obtained from that of the Poisson v Pascal by changing $p$ to $-p$ and $n$ to $-n$. Therefore, in order to obtain the various probabilities and derivatives of the probabilities, we can use the formulae for the Poisson v Pascal, given in section B. Since the only parameters being estimated are $p$ and $\lambda$, the information determinant is given by

$$
\frac{1}{N} \left[ \sum_{r=0}^{\infty} \frac{\left( \frac{\lambda P_r}{\lambda} \right)^2}{P_r} + \sum_{r=0}^{\infty} \frac{\lambda P_r}{\lambda} \frac{P_r}{P_r} + \sum_{r=0}^{\infty} \frac{\left( \frac{\lambda P_r}{\lambda} \right)^2}{P_r} \right].
$$

3. Generalized variance and efficiency of the method of moments

By referring to section D of Chapter III, we note that the vectors $\{K_{[1]}, K_{[2]}\}$ and $(\lambda, p)$ correspond to $\tau$ and $\lambda$. 
respectively. For \( \theta \), we choose \((\mu_1, \mu_2)\). As in section 2, we obtain the elements of the matrices \(J(\gamma, \lambda)\), \(J(\gamma, \phi)\) and \(C(\hat{\theta})\) by using the formulae for the corresponding elements in the case of the Poisson v Pascal distribution and substituting the vector \((\lambda, -p, -n)\) for \((\lambda, p, k)\). The generalized variance of the estimates is then obtained from formula (34) of Chapter III.

The evaluation of the efficiency follows from its definition and the information determinant, calculated as in section 2. The efficiency is given in Tables 33, 34 and 35 for certain values of the vector \((\lambda, p, n)\). The efficiency is easily seen to be high for small values of the parameters and small for large values.

<table>
<thead>
<tr>
<th>p ( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.928</td>
<td>.865</td>
<td>.843</td>
<td>.840</td>
<td>.870</td>
</tr>
<tr>
<td>.3</td>
<td>.730</td>
<td>.569</td>
<td>.525</td>
<td>.533</td>
<td>.635</td>
</tr>
<tr>
<td>.5</td>
<td>.494</td>
<td>.307</td>
<td>.264</td>
<td>.267</td>
<td>.392</td>
</tr>
</tbody>
</table>
Table 34. Efficiency of the method of moments for the Poisson v Binomial distribution for \( n = 3 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.896</td>
<td>.823</td>
<td>.793</td>
<td>.779</td>
<td>.810</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.658</td>
<td>.501</td>
<td>.452</td>
<td>.446</td>
<td>.542</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.426</td>
<td>.268</td>
<td>.231</td>
<td>.231</td>
<td>.333</td>
<td></td>
</tr>
</tbody>
</table>

Table 35. Efficiency of the method of moments for the Poisson v Binomial distribution for \( n = 5 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.816</td>
<td>.726</td>
<td>.688</td>
<td>.671</td>
<td>.715</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.527</td>
<td>.379</td>
<td>.337</td>
<td>.332</td>
<td>.435</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.345</td>
<td>.210</td>
<td>.178</td>
<td>.176</td>
<td>.277</td>
<td></td>
</tr>
</tbody>
</table>

4. Generalized variance efficiency for the method of the first moment and the first frequency

Here, the vectors \(( \theta_1, P_0)\) and \(( \lambda, p)\) correspond to \( \gamma \) and \( \lambda \) respectively. We choose for \( \theta \), the vector \(( \lambda_1, P_0)\). By following the lines in section 3 above, the elements of the matrices \( J( \gamma, \lambda) \) and \( C( \hat{\theta}) \) by using the formulae corresponding to the Poisson v Pascal distribution and substituting the vector \(( \lambda, p, k)\) by \(( \lambda, -p, -n)\). Generalized variance of the estimates is then obtained by using
formule (34) of Chapter III. Efficiency, evaluated by using this generalized variance and the information determinant of section 2 is given in Tables 36, 37 and 38 for certain values of \((\lambda, p, n)\). Again, it is to be noted that the improvement in the efficiency over the method of moments is considerable. Efficiency, of course, seems to decrease to zero as the values of the parameters increase.

Table 36. Efficiency of the method of the first moment and the first frequency for the Poisson v Binomial distribution for \(n = 2\)

<table>
<thead>
<tr>
<th>(\lambda)</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.984</td>
<td>.974</td>
<td>.977</td>
<td>.991</td>
<td>.947</td>
</tr>
<tr>
<td>.3</td>
<td>.937</td>
<td>.988</td>
<td>.883</td>
<td>.923</td>
<td>.981</td>
</tr>
<tr>
<td>.5</td>
<td>.862</td>
<td>.740</td>
<td>.700</td>
<td>.717</td>
<td>.851</td>
</tr>
</tbody>
</table>

Table 37. Efficiency of the method of the first moment and the first frequency for the Poisson v Binomial distribution for \(n = 3\)

<table>
<thead>
<tr>
<th>(\lambda)</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.994</td>
<td>.986</td>
<td>.984</td>
<td>.944</td>
<td>.994</td>
</tr>
<tr>
<td>.3</td>
<td>.968</td>
<td>.930</td>
<td>.918</td>
<td>.935</td>
<td>.974</td>
</tr>
<tr>
<td>.5</td>
<td>.896</td>
<td>.798</td>
<td>.763</td>
<td>.765</td>
<td>.850</td>
</tr>
</tbody>
</table>
Table 38. Efficiency of the method of the first moment and the first frequency for the Poisson v Binomial distribution for n = 5

<table>
<thead>
<tr>
<th>p (\lambda)</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.995</td>
<td>.987</td>
<td>.985</td>
<td>.993</td>
<td>.989</td>
</tr>
<tr>
<td>.3</td>
<td>.969</td>
<td>.924</td>
<td>.905</td>
<td>.911</td>
<td>.950</td>
</tr>
<tr>
<td>.5</td>
<td>.889</td>
<td>.769</td>
<td>.716</td>
<td>.690</td>
<td>.793</td>
</tr>
</tbody>
</table>

5. Generalized variance and efficiency for the method of the first moment and the ratio of the first two frequencies

Here, it is to be noted that the vectors \((k_{11}, P_1/P_0)\) and \((\kappa, p)\) correspond to \(\tau\) and \(\lambda\) of section D, Chapter III. For vector \(\Theta\), we choose \((k_1^1, P_1/P_0)\). As mentioned in section 3, the elements of the matrices \(J(\tau, \lambda)\) and \(C(\Theta)\) can be obtained by using the formulae corresponding to the Poisson v Pascal distribution and substituting \((\lambda, -\nu, -n)\) for \((\kappa, \nu, k)\). Generalized variance is then calculated by using formula (34) of Chapter III. Efficiency, calculated by using this generalized variance and the information determinant of section 2 is given for certain values of \((\lambda, \nu, n)\) in Tables 39, 40 and 41.

From these tables it is apparent that this method is somewhat more efficient than the method of the first moment and the first frequency for very small values of the para-
Table 39. Efficiency of the method of the first moment and the ratio of the first two frequencies for the Poisson v Binomial distribution for \( n = 2 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.997</td>
<td>.997</td>
<td>.999</td>
<td>.993</td>
<td>.933</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.986</td>
<td>.978</td>
<td>.981</td>
<td>.980</td>
<td>.970</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.969</td>
<td>.933</td>
<td>.918</td>
<td>.910</td>
<td>.847</td>
<td></td>
</tr>
</tbody>
</table>

Table 40. Efficiency of the method of the first moment and the ratio of the first two frequencies for the Poisson v Binomial distribution for \( n = 3 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.996</td>
<td>.998</td>
<td>.995</td>
<td>.976</td>
<td>.888</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.968</td>
<td>.986</td>
<td>.988</td>
<td>.962</td>
<td>.806</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.877</td>
<td>.908</td>
<td>.928</td>
<td>.938</td>
<td>.834</td>
<td></td>
</tr>
</tbody>
</table>

Table 41. Efficiency of the method of the first moment and the ratio of the first two frequencies for the Poisson v Binomial distribution for \( n = 5 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \lambda )</th>
<th>.1</th>
<th>.3</th>
<th>.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.974</td>
<td>.979</td>
<td>.975</td>
<td>.943</td>
<td>.814</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.842</td>
<td>.880</td>
<td>.896</td>
<td>.885</td>
<td>.734</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.564</td>
<td>.615</td>
<td>.648</td>
<td>.693</td>
<td>.706</td>
<td></td>
</tr>
</tbody>
</table>
meters but is much poorer than the latter method for moderate as well as large values of the parameters.
VI. SELECTION OF AD HOC METHODS BY PRELIMINARY TESTS

A. Introductory Remarks

In Chapters IV and V, a number of methods were developed for estimating the parameters in a somewhat simple fashion for certain families of distributions. It was found that different methods are more efficient in different regions of the parameters. Hence, when one has a sample data at hand and has no previous knowledge about the region in which the true parameters lie, there arises a necessity for setting up a rule to choose a method for estimating the parameters. In this chapter, it is intended to discuss the choosing of the method by preliminary tests and indicate some of their properties. A full scale study is beyond the scope of the thesis. We will start off by defining a few terms that will be used in the thesis in this connection.

B. Definitions

1. Preliminary test

A preliminary test (to be abbreviated to P.T.) is a test of hypotheses, conducted with the intention of using the accepted hypothesis as the true hypothesis in future work.
2. P.T. criterion

A P.T. criterion is the set of statistics used in performing the preliminary test.

3. P.T. partition

A P.T. partition is the division of the total range $R$ of the preliminary test criterion into parts ($R_1, R_2, \ldots, R_n$) such that the $i^{th}$ hypothesis is accepted if the value of the preliminary criterion belongs to $R_i$.

4. P.T. result

A P.T. result is the set of all conclusions reached on the basis of the preliminary test.

5. P.T. departure function

A P.T. departure function is a real valued function of the P.T. result, which is taken as a measure of the departure of the preliminary test result from the true conclusions.

6. Efficiency of the P.T. relative to a null hypothesis

We define the efficiency of a P.T. as the ratio:

\[
\frac{\text{The value of the P.T. departure function if the null hypothesis was always accepted}}{\text{The value of the P.T. departure function for the partition (}}R_1, R_2, \ldots, R_n\text{)}\]
P.T. partition, locally optimal with respect to a given P.T. criterion and a given P.T. departure function for a given true hypothesis

A P.T. partition is said to be locally optimal with respect to a given P.T. criterion and a given P.T. departure function for a given true hypothesis if the value of the P.T. departure function for this partition is not larger than that of any other partition for the given true hypothesis.

C. A Preliminary Test for Selecting One of the Two Methods of Estimation

1. Notation and statement of the preliminary test

Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be the k-dimensional vector of the parameters of a distribution. Let $X_1, X_2$ be two methods of estimation which use the statistics $T_1 = (t_{11}, \ldots, t_{1k})$ and $T_2 = (t_{21}, \ldots, t_{2k})$ and yield the estimates $\hat{\lambda}^{(1)} = (\hat{\lambda}_1^{(1)}, \ldots, \hat{\lambda}_k^{(1)})$ and $\hat{\lambda}^{(2)} = (\hat{\lambda}_1^{(2)}, \ldots, \hat{\lambda}_k^{(2)})$ respectively. Assume that $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ are consistent estimates of $\lambda$. Assume further that $T_1$ and $T_2$ estimate $\gamma_1 = (\gamma_{11}, \ldots, \gamma_{1k})$ and $\gamma_2 = (\gamma_{21}, \ldots, \gamma_{2k})$ respectively. In the space of $\gamma_1$, draw the region in which the efficiency of $X_1$ relative to $X_2$ is greater than unity and denote it by $R$. Denote the complement of $R$ by $\bar{R}$. For the preliminary test, we will use the following procedure.

Plot the vector $T$ for the sample at hand in the space
of vector $\gamma_1$. Use $\xi_1$ for estimating the parameters if the $T$ belongs to $R$ and use $\xi_2$ is $T$ belongs to $\bar{R}$.

The estimates obtained with the help of this preliminary test will be referred to as the P.T. estimates and denoted by $\lambda(p)$. For the P.T. departure function, use will be made of the determinant of the mean square error of $\lambda(p)$.

2. A property of the above preliminary test

In the space of the parameter vector $\lambda$ assume that the region in which the efficiency of $\xi_1$ relative to $\xi_2$ is larger than unity has been drawn.

It is easy to see that under the assumptions inherent in the estimation of the parameters by these methods, the contours of fixed relative efficiency in the space of $T_1$ and $\lambda$ are topological maps of each other. Hence the regions in the two spaces in which efficiency of $\xi_1$ relative to $\xi_2$ is larger than unity are also topological maps of each other. In view of this, we can denote these regions and their complements in their respective regions by the common symbols $R$ and $\bar{R}$ respectively to avoid cumbersome notation.

Suppose that the population under consideration is such that the point $\lambda p$ corresponding to its parameter vector $\lambda$ and consequently the point $\gamma p$ corresponding to the vector $\gamma$ does not lie on the boundary of $\bar{R}$. It will now be shown that asymptotically, the P.T. estimates are the same as the
ones obtained by the method that is more efficient at $\lambda_p$.

First, assume that $\gamma_p$ belongs to $R$. It is apparent from the formation of $R$, that it is an open set. Hence, there exists a neighborhood $\delta$ of $\gamma_p$ which lies wholly in $R$. Since, by assumption, $T$ is a consistent estimator of $\gamma$, $T$ belongs to $\delta$ asymptotically with probability 1. Hence, with the present preliminary test, method $M_1$ is selected asymptotically with probability 1 and it follows therefore that the P.T. estimate $\hat{\lambda}^{(p)}$ is the same as that of $M_1$ which is more efficient at the point $\lambda_p$ in the parameter space.

If $\gamma_p$ does not belong to $R$, then $\gamma_p$ belongs to $\overline{R}$. Since $\gamma_p$ does not lie on the boundary, by assumption, this implies that there exists a neighborhood $\delta$ of $\gamma_p$ which completely belongs to $\overline{R}$ and as above, the consistency of $T$, show that $T$ lies in $\delta$ and consequently in $\overline{R}$ with probability 1, asymptotically. Hence $M_2$ is accepted almost always, implying that the P.T. estimates are, again, the same as those of $M_2$, a method more efficient than $M_1$ at the point $\lambda_p$.

D. Locally Optimal Critical Region for a Simple P.T. Departure Function

Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be the vector of the parameters of a population and $M_1, M_2$ the two methods of estimation, giving the consistent estimates $\hat{\lambda}^{(1)}, \hat{\lambda}^{(2)}$ respectively. As above, assume that there is a partition $(R, \overline{R})$ of the parameter space and the preliminary test consists of choosing
$M_1$ or $M_2$ for estimating $\lambda$ according as $\lambda^{(1)}$ belongs to $R$ or $\overline{R}$. Let the P.T. departure function be the mean square error of $\lambda^{(p)}$. We will derive here the equation of the boundary of $R$ when $R$ is locally optimal with respect to the above preliminary test and the P.T. departure function.

We first observe that the expressions for the mean square error is

$$E(\lambda^{(p)} - \lambda_1)^2 = \int_{R} \int_{\lambda^{(p)} = -\infty}^{+\infty} (\lambda^{(1)} - \lambda_1)^2 p(\lambda^{(1)}, \lambda^{(2)}) d\lambda^{(1)} d\lambda^{(2)}$$

$$= \int_{\lambda^{(p)} = -\infty}^{+\infty} (\lambda^{(1)} - \lambda_1)^2$$

where $p(\lambda^{(1)}, \lambda^{(2)})$ is the joint distribution of the vector $\lambda^{(1)}$ and the term $\lambda^{(2)}$ and $d\lambda^{(1)} = d\lambda^{(1)} d\lambda^{(1)} ... d\lambda^{(1)}$. Hence if we transfer an infinite small region $\Delta\lambda_1 \Delta\lambda_2 ... \Delta\lambda_k$ in the neighborhood of the point $\bar{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_k)$, the change in the mean square error will be given by

$$E(\lambda^{(1)} - \lambda_1)^2 = -\Delta\lambda_1 \Delta\lambda_2 ... \Delta\lambda_k \int_{\lambda^{(p)} = -\infty}^{+\infty} (\lambda^{(2)} - \lambda_1)^2 d\lambda^{(1)}$$

$$(\bar{\lambda}_1 - \lambda_1)^2 p(\bar{\lambda}, \lambda^{(2)}) d\lambda^{(2)}$$
By making the large sample approximation that the joint distribution of \((\hat{\lambda}_1^{(1)}, \hat{\lambda}_1^{(2)})\) is a multivariate normal, we have

\[
E \left[ (\hat{\lambda}_1^{(2)} - \lambda_1)^2 \mid \lambda_1^{(1)} = \lambda \right] = K + \left( \sum a_i \left( \hat{\lambda}_1 - \lambda_1 \right) \right)^2
\]

where \(K\) is the conditional variance of \(\hat{\lambda}_1^{(2)}\) which does not depend on \(\lambda\) and \(\sum a_i (\hat{\lambda}_1 - \lambda_1)\) is the bias due to estimating \(\lambda_1\) by \(\hat{\lambda}_1^{(2)}\) in the conditional distribution of \(\hat{\lambda}_1^{(2)}\).
with \( \lambda^{(1)}_1 = \overline{\lambda}_1 \), \( a_1 \)'s not involving \( \overline{\lambda}'s \).

For the partition \((R, \overline{R})\) to be optimal, \( \Delta \mathcal{E}(\lambda^{(\rho)}_1 - \lambda_1)^6 = 0 \) for the points on the boundary of \( R \). Hence the equation of the boundary of \( R \) is

\[
(\overline{\lambda}_1 - \lambda_1)^2 = K + \left( \sum a_1(\overline{\lambda}_1 - \lambda_1) \right)^2.
\]

Note that it is a hyperbolic cylinder with the planes

\[
(a_1 - 1)(\overline{\lambda}_1 - \lambda_1) + \sum_{i=2}^{k} (\overline{\lambda}_1 - \lambda_1) = 0
\]

and

\[
(a_1 + 1)(\overline{\lambda}_1 - \lambda_1) + \sum_{i=2}^{k} (\overline{\lambda}_1 - \lambda_1) = 0
\]

as the asymptotic planes.

For the particular case of a two parameter family, if the quadratic form of \( \overline{\lambda}^{(1)}, \overline{\lambda}^{(2)}, \overline{\lambda}_2 \) is of form

\[
\rho(\overline{\lambda}_1 - \lambda_1)^2 + b(\overline{\lambda}_1 - \lambda_1)^2 + c(\overline{\lambda}_2 - \lambda_2)^2 + 2h(\overline{\lambda}_1 - \lambda_1)(\overline{\lambda}_1 - \lambda_1) + 2g(\overline{\lambda}_1 - \lambda_1)(\overline{\lambda}_2 - \lambda_2) + 2f(\overline{\lambda}_1 - \lambda_1)(\overline{\lambda}_2 - \lambda_2) = 0,
\]

\[
\Xi(\overline{\lambda}_1 - \lambda_1 | \overline{\lambda}_1 - \lambda_1) = \frac{h}{b}(\overline{\lambda}_1 - \lambda_1) - \frac{g}{b}(\overline{\lambda}_2 - \lambda_2)
\]

and the equation of the boundary is
\[(\bar{\lambda}_1 - \lambda_1)^2 = k + \left(\frac{h}{b} (\bar{\lambda}_1 - \lambda_1) + \frac{g}{b} (\bar{\lambda}_2 - \lambda_2)\right)^2.\]

It is a hyperbola with \((\lambda_1, \lambda_2)\) as centre and

\[\left(\frac{h}{b} + 1\right)(\bar{\lambda}_1 - \lambda_1) + \frac{g}{b} (\bar{\lambda}_2 - \lambda_2) = 0\]

as the asymptotes.

E. General Remarks

If one has at hand more than two methods of estimation, say \(n\), then by following the lines in section C, one may divide the parameter space in \(n\) parts, \(R_1, R_2, \ldots, R_n\) such that for a point in \(R_i\), the \(i\)th method is more efficient than each one of the remaining \((n - 1)\) methods. For the sake of simplicity in performing the test, the regions may, again, be drawn in the space of the statistics used in obtaining the estimates that are being used as the P.T. criterion. The preliminary test will consist of selecting the method \(M_i\) if the point corresponding to the P.T. criterion belongs to \(R_i\). It is easy to see that such a preliminary test has the property that the resulting P.T. estimates are the same as the ones obtained by the most efficient method except when the parameter of the population lies on the boundary of one of the regions. This restriction may not be strong in general since the boundary of the regions occupy an area of Lebesque measure zero in the parameter space. It is to be noted that the above test uses the estimates obtained by a
single method. It shows that when the sample is not large, there is scope for obtaining more efficient tests.

In section D, we found that for a simple P.T. departure function, we can obtain a locally optimal critical region. It should be clear from the method used therein that such optimal critical regions will exist for small samples and for every P.T. departure function so long as it satisfies certain regularity conditions of continuity and differentiability. A study of these will be taken up in future research.

The author likes to stress that the preliminary tests are being offered here as a means of selecting a good method of estimation when the available methods are not very simple. If obtaining estimates is not cumbersome, the best way of getting good estimates is to combine the estimates or estimate them afresh by combining all the statistics that are used in various methods of estimation.
VII. A MEASURE OF RELIABILITY OF STATISTICS

A. Introduction

In Chapter VI, a discussion was given of using preliminary tests for selecting one of the methods of estimation. An alternative is to select the statistics which are themselves reliable - reliable in some sense of the term - so that they will automatically lead to reliable estimates. In this chapter, it is intended to use the reciprocal of the squared coefficient of variation as a measure of the reliability of the statistics and study some of its properties.

B. A Definition of Reliability

Let $\tau$ be a certain function of the parameters of an integral valued distribution and $t$, a consistent estimate of $\tau$. Let $t$ be such that:

(i) $t = 0$ for a sample of size zero

(ii) If an observation is added to the original sample, the change in the value of the statistic $t$ will not involve any observation of the original sample.

The reliability of such a statistic $t$ will be measured by the ratio

$$ R = \frac{\tau^2}{\overline{V(t)}}. $$
If there are two statistics $t_1$ and $t_2$, satisfying the conditions (i) and (ii), then the relative reliability of $t_1$ relative to $t_2$ is defined as the ratio

$$R_{t_2}(t_1) = \frac{R_1}{R_2} = \frac{\tau^2_1}{V(t_1)} \cdot \frac{V(t_2)}{\tau^2_2}.$$  

(4)

C. Reliability of Certain Statistics

1. Reliability of the $i^{th}$ observed count

Let $n$ be the sample size, $p_i$, the probability of the $i^{th}$ count and $n_i$ the observed value of count $i$. It is easy to show that

$$V(n_i) = np_i(1 - p_i).$$

Clearly, $n_i$ satisfied the restrictions (i) and (ii). Hence its reliability, by definition (3), is given by

$$R = \frac{(np_i)^2}{np_i(1 - p_i)} = \frac{np_i}{1 - p_i}.$$  

(5)

It is to be noted that $R$ is large when $p_i$ is large. This measure of reliability, therefore, recommends the use of the highest mode for estimation. In this connection, it is of interest to note that the first frequency yielded relatively more efficient estimates in the methods of Chapters IV and V when it was large in size.
2. **Reliability of the ratio of the i**\textsuperscript{th} **count to the j**\textsuperscript{th} **count for large samples**

Define the ratio of the i\textsuperscript{th} count to j\textsuperscript{th} count as

\[ r = \frac{P_i}{P_j} \]

where \( P_i, P_j \) have the usual meanings. As shown in Chapter III, the large sample variance of \( r \) is given by

\[ V(r) = \frac{1}{N} \frac{P_i(P_i + P_j)}{P_j^3} \]

For large samples, \( r \) can be expanded in powers of \( P_i - P_1 \) and \( P_j - P_j \) and approximated by neglecting the second and higher order terms in them, to

\[ r = \frac{P_i}{P_j} + \frac{P_i - P_1}{P_j} - \frac{P_i(P_j - P_j)}{P_j^2} \]

It is clear that in the approximate form (7), \( r \) satisfies conditions (1) and (11). Hence the reliability for large samples is

\[ R = \frac{(P_i/P_j)^2}{\frac{1}{N} \frac{P_i(P_i + P_j)}{P_j^3}} = \frac{NP_iP_j}{P_1 + P_j} \]

Set \( P_i + P_j = P \). Then \( R \) can be rewritten as

\[ R = \frac{NP_i(P - P_i)}{P} \]

This shows that \( R \) increases with \( P \) and that for a given \( P \), \( R \) is maximum when \( P_i = P/2 \). Geometrically, this can be interpreted as follows.
The ratio $r$ has high reliability when the line joining the $i$\textsuperscript{th} count to the $j$\textsuperscript{th} count on the graph of the distribution is parallel to the axis of the random variable and is far off from it.

In connection with this, it is of interest to note that when the ratio of the first two frequencies was used as a component in estimating the parameters of the Neyman Type A distribution in Chapter IV, the efficiency of this method relative to that of the moments was a maximum, for non-zero finite values of the parameters for which the ratio is close enough to unity and the values of $P_1$ and $P_0$ not small.

3. Reliability of the $i$\textsuperscript{th} count relative to the ratio $r = \hat{P}_1/\hat{P}_j$

From the definition of relative reliability we obtain

$$R_{\hat{P}_1/\hat{P}_j}(\hat{P}_1) = \frac{\text{Reliability of } \hat{P}_1}{\text{Reliability of } (\hat{P}_1/\hat{P}_j)}$$

$$= \frac{N P_1}{1 - P_1} \cdot \frac{P_1 + P_j}{N P_1 P_j}$$

$$= \frac{P_1 + P_j}{(1 - P_1)P_j} ,$$

$\forall 1$ for all $P_1, P_j$.

This count shows that using the $i$\textsuperscript{th} count is better than using the ratio $\hat{P}_1/\hat{P}_j$ for every $P_1$ and $P_j$. The considerably smaller range in which the method of the first moment and
first ratio is superior to the method of the first moment and first frequency in Chapter IV and the considerably smaller superiority that the method involving the ratio has over the method involving the first frequency tend to corroborate this result.

4. Reliability of the mean

Let \( \mu_1 \) be the mean and \( \hat{\mu}_1 \) the sample mean which is taken as the estimate of \( \mu_1 \). Then we have

\[ \text{Var}(\hat{\mu}_1) = \frac{\mu_2}{N}. \]

It is apparent that \( N\mu_1 \) satisfies the conditions (i) and (ii). Hence, the reliability of \( N\mu_1 \) is given by

\[ R = \frac{N\mu_1^2}{N^2 \mu_2/N} = \frac{N\mu_1^2}{\mu_2}. \]

For the Negative Binomial and the Neyman Type A distribution, it is apparent that \( \mu_1 \) and \( \mu_2 \) are of the form \( kp \) and \( kp(1+p) \) respectively. Hence

\[ R = \frac{N(kp)^2}{kp(1+p)} = \frac{Nkp}{1+p}. \]

It is curious to note from this expression that the reliability of the mean increases with \( k \) and \( p \). The maximum likelihood method of estimation yields the sample mean as an estimate of \( \mu_1 \) for both of these distributions, indicating that
the mean has very high reliability for all values of the parameters.

D. Relation between Efficiency and Reliability

Consider the estimation of the parameter $\lambda$ of a uniparameter family. Let $t_1$, $t_2$ be the two statistics and let the equations for estimating the parameters in the first case be

$$t_1 = \gamma_1(\lambda), \quad t_2 = \gamma_2(\lambda)$$

respectively. Let $\hat{\lambda}_1$, $\hat{\lambda}_2$ be the two estimates. Then by using the method of statistical differentials, we have for large samples,

$$V(t_1) = \left(\frac{d \gamma_1}{d \lambda}\right)^2 V(\hat{\lambda}_1)$$

$$V(t_2) = \left(\frac{d \gamma_2}{d \lambda}\right)^2 V(\hat{\lambda}_2).$$

Hence, the relative efficiency of the first method, relative to the second is given by

$$t = \frac{V(\hat{\lambda}_2)}{V(\hat{\lambda}_1)} = \frac{\left\{\frac{d \gamma_2}{d \lambda}\right\}^2}{\left\{\frac{d \gamma_1}{d \lambda}\right\}^2 \frac{V(t_2)}{V(t_1)}}$$

Since the relative reliability of the statistic $t_1$, relative to $t_2$ is

$$R_{t_2}(t_1) = \frac{V(t_2)}{V(t_1)} \frac{\gamma_1^2}{\gamma_2^2}$$
we have

\[ E = R_{t_1}(t_1) \frac{t_1^2}{\left(\frac{d \log t_1}{d \lambda}\right)^2} \]

\[ = R_{t_1}(t_1) \frac{\left(\frac{d}{d \lambda} \log t_1\right)^2}{\left(\frac{d}{d \lambda} \log t_2\right)^2}. \]

This expression shows that when \( R_{t_1}(t_1) \) is small, \( E \) can be large and vice versa. It is of interest to observe that for a given \( E \), and \( \frac{d}{d \lambda} \log \gamma_2 \), \( R_{t_1}(t_1) \) is small if \( \frac{d}{d \lambda} \log \gamma_1 \) is large and large if \( \frac{d}{d \lambda} \log \gamma_1 \) is small.

E. General Remarks

The efficiency function, which is commonly used as a measure of the goodness of the estimates, uses the variation in the statistic about its true value and makes no use of the true value itself. Such a measure is not very realistic for small samples wherein the variance of the statistic can be comparable with the size of the statistic for the following reasons:

(1) The random variations may give the statistic such a large or small value that the resultant estimates may often not be acceptable.

(2) In the case of discrete distributions, the statistics can take on discrete values only while their expecta-
tions are many times continuous. Hence, in general, the observed statistic deviates from its expected value. This deviation may be referred to as the error due to discreteness. This error may introduce large errors in the estimate.

It is believed that if a statistic is large, the estimate will be less susceptible to these errors. Conditions (i) and (ii) have been imposed on the statistics while estimating their reliability so that the statistics available to the statisticians can be standardized to a form which brings out the error due to discreteness more clearly. It is apparent that there will be statistics which will not satisfy both of these conditions. The author intends to continue the study of these conditions in future research.
VIII. SOME DEVELOPMENTS ON CONTAGION

A. Introductory Remarks

Contagion is the phenomenon by which the occurrence of an event affects the occurrence of the next event. In this perspective, a state of least contagion is one in which the successive events are independent. A distribution developed on the basis of independence is the Poisson distribution. It is of interest to note that for the Poisson distribution, variance is equal to the mean. In a number of practical examples wherein contagion is present, variance has been larger than the mean. Feller [8] shows that for every contagious distribution which can be regarded as a "compound Poisson" distribution, variance is larger than the mean. Many of the compound and generalized distributions developed recently by Gurland [13], including the ones discussed in the earlier chapters have variance larger than or equal to the mean for all values of the parameters.

In section B, we will indicate a few situations that can give rise to distributions in which variance is smaller than the mean.

In section C, we will deal with the distribution of dental caries in children which is a case of true contagion and which has variance smaller than the mean. Of particular interest will be the formulation of a dynamic model for the
increase in the dental caries from year to year and setting up a null hypothesis for the distribution of caries at time $t_4$ with the help of a sample at a previous time $t_0$.

In section D, the effect of plot size and shape on the distribution of a sample in the presence of true contagion is briefly discussed and an indication is given as to how the distribution of the caries changes as various sets of teeth are taken as forming a plot.

B. Examples of Contagious Situations, Wherein Variance Can Be Smaller than the Mean

1. Increasing mean

Suppose that there is a large field in which a set of individuals have spread themselves according to a certain distribution. Suppose that each of these individuals, in the course of time, gives rise to $n$ individuals, of which one stays in the place of the parent and the rest spread themselves homogeneously over the field. It is apparent that the mean increases $n$ fold while the variance remains unaffected. Hence, for large enough $n$, mean will be larger than the variance.

2. Decreasing variance

Consider the distribution of accidents on individuals. If the individuals are highly susceptible to accidents until
they receive $n$ accidents and become resistant to accidents after they receive $n$ accidents, there occurs a pooling of the distribution of the number of accidents on an individual around the number $n$. By making the susceptibility to accidents before receiving $n$ of them large enough and by decreasing the susceptibility after receiving $n$ accidents, one can get a distribution with finite mean (close to $n$) and a very small variance.

3. **Finite size of the plot**

In some cases, the finiteness of the plot puts an upper limit on the number of events that can occur in it. For example, the limited number of teeth in a mouth restricts the number of carious teeth per mouth. In such cases, the mean may continue to increase with time while the corresponding variance starts decreasing after a certain lapse of time and decrease ultimately to zero.

4. **Competition and social instincts**

In the case of the distribution of animals, social instincts tend to keep them together and an upper limit over their density is set by the availability of living necessities like food. When the density at any place grows beyond this upper limit, there occurs an immigration of the animals to the sparsely populated region. Thus, if we assume that the
living conditions at all places are homogeneous, there will be a tendency for the animals to develop a homogeneous population in the course of time. Their distribution therefore has a finite mean and a relatively small variance in the long run.

C. Distribution of Dental Caries in Children

1. Statement of the problem

Attempts to obtain the statistical distribution of dental caries has been made by Grainger and Reid [11]. Their method consists of fitting the well-known distributions such as the Negative Binomial to sample data and selecting the one which gives a close fit. This method was highly unsuccessful in the case of the distribution of caries in the unilateral pitted surfaces, partly because the variance for these distributions is smaller than their mean while the variance of the theoretical distribution fitted is larger than their mean. While Grainger and Reid state in their paper that this failure is due to the heterogeneity of the susceptibility of these surfaces, the present author, following the lines of section B, believes that variance being smaller than the mean is characteristic of the highly contagious distributions with finite plot size. In the forthcoming sections, a model will be formulated for the distribution of caries and fitted to some of the data given by Grainger and Reid [11].
2. Model for the distribution of unilateral pitted surface cavitation count

Let us assume that the number of carious surfaces in the mouth of an individual at any age $t_1$ depends on:

(1) the number of carious surfaces that he gets by "pure chance", i.e. due to influences which cannot be accounted for by simple reasoning, and

(2) the number of carious surfaces that he develops due to the fact that he had some carious surfaces at a previous age.

Let $n$ be the number of pitted surfaces on the side of the mouth under consideration and $p(t_0)$, the probability that a surface without caries gets affected during the interval $(t_0, t_1)$ by pure chance. If $p(t_0)$ is small (and as we will find in the later calculations) and if $m$ of the surfaces have no cavity at age $t_0$, then the probability that a new carious surface appears at age $t_1$ is approximately given by $mp(t_0)$ and the probability that no new carious surface appears by $1 - mp(t_0)$. Further assume that the surfaces can be grouped into sets of two neighboring surfaces and that if a surface is carious, it, in essence, increases the probability of its neighbor becoming carious by an amount, say $s(t_0)$, if the neighbor is not already carious. Then if we denote by $P_r^{1(2), r-21}(t)$, the probability of an individual at age $t$ having $r$ carious surfaces, 21 of which belong to $i$ pairs of
neighbors and \( r - 2i \) of which are singles, i.e. they are such that none of them is a neighbor of any one of the carious surfaces, the above model leads to the general stochastic relationship

\[
p_{r}^{(2)}, r-2^{i}(t_{1}) = \sum_{j=0}^{\frac{1}{r-i}} \text{probability of having } r - j \text{ carious surfaces which consist of } r - 2i + j \text{ singles and } (i - j) \text{ pairs at time } t_{0} \text{ and developing } j \text{ carious surfaces by contagion during the interval } (t_{0}, t_{1}) + \sum_{j=0}^{1} \{ \text{probability of having } (r - j - 1) \text{ surfaces carious at time } t_{0}, \text{ which consist of } (i - j) \text{ pairs and } r - 2i + j - 1 \text{ singles and developing } j \text{ caries by contagion and } 1 \text{ by pure chance } \}
\]

\[
= \sum_{j=0}^{1} (r-2i+j)_{r}^{(1-j)(2)}, r-2^{i}+j(t_{0})s^{j}(t_{0})(1-s(t_{0}))^{r-2i}
\]

\[
[1 - (n - r + j)p(t_{0})] + \sum_{j=0}^{1} (r-2i+j-1)_{r-1}^{(1-j)(2)}, r-2i+j-1
\]

\[
(t_{0}) s^{j}(t_{0})(1-s(t_{0}))^{r-2i-1} (n - r + j + 1)p(t_{0}) .
\]

(1)

Here, we make the convention that a term will be set equal to zero if any superscript or subscript of \( p(t_{0}) \) is negative. The terms of (1) for certain small values of \( r \) are given below.
\[ P_0(2),0(t_1) = P_0(2),0(t_0)(1-np(t_0)) \]
\[ P_1(2),1(t_1) = P_0(2),1(t_0)np(t_0) + P_1(2),1(t_0) \]
\[ l-(n-1)p(t_0)(1-r(t_0)) \]
\[ P_2(2),1(t_1) = P_0(2),1(t_0)(n-1)p(t_0)(1-s(t_0)) \]
\[ + P_2(2),2(t_0)l-(n-\rho)p(t_0)(1-s(t_0)) \]
\[ P_2(2),0(t_1) = P_0(2),1(t_0)s(t_0) + P_1(2),0(t_0)(1-(n-\rho)p) \]
\[ P_3(2),3(t_1) = P_2(2),2(t_0)(1-s(t_0))(n-\rho)p \]
\[ + P_3(2),3(t_0)(1-s(t_0))^3(1-(n-3)p(t_0)) \]
\[ P_3(2),1(t_1) = 2P_2(2),2s(t_0)(1-s(t_0))(1-(n-\rho)p(t_0)) \]
\[ + P_1(2),0(n-\rho)p + P_3(2),1(t_0)(1-s(t_0)) \]
\[ (1-(n-3)p(t_0)) . \]

Obviously, the probability \( P_r(t_1) \) of obtaining in all \( r \) carious surfaces is given by
\[ \sum_{1=0}^{[r/2]} P_r(2),r-2(I(t_1)) \]
\[ \text{where } [r/2] \text{ stands for the largest integer less than or equal to } r/2 \text{ as usual.} \]
3. Application of the model

For illustrative purposes, this model is fitted to parts of Table VIII of Grainger and Reid [11]. In fitting such models, one is faced with two estimating problems:

(1) To calculate the expected frequencies at time \( t_1 \), we need to know the values of \( p^i_r(\tau), (r-\delta_1)(t_0) \) at a previous time \( t_0 \). In the data, the values corresponding to \( p^i_r(\tau), (r-\delta_1)(t) \) are generally so small and hence highly susceptible to random fluctuations that regarding the sample values as the true values for some \( t \) introduces inefficiency in the evaluation of the distribution for subsequent years. Hence, the distribution for the basal year \( t_0 \) needs to be estimated.

(2) The quantities \( p(t) \) and \( s(t) \) also need to be estimated.

As for (1), we first consider age 6 as the basal year, assume the sample distribution to represent the true distribution at that age, get the expressions for the first few frequencies at age 7 using equations (2) and equate these to the corresponding observed values at age 7. For \( n \), we choose the values from Table IX of Grainger and Reid [11].* Then we have

*In general, we propose that records be made of unilateral pitted surfaces of each individual and take their average as an estimate of \( n \). This will avoid (continued)
\[ P_0(\gamma) = \frac{30}{156} = \frac{31}{94} (1-5p(6)) \]

\[ P_1(\gamma) = \frac{31}{156} = \frac{31}{94} 5p(6) + \frac{25}{94} (1-4p(6))(1-s(6)) \]

\[ P_2(\gamma) = \frac{62}{156} = p^0(2),2(\gamma) + p^1(2),0(\gamma) \]

\[ = \frac{25}{94} (1-2(6))4p(6) + p^0(2),1(6)(1-s(6))p(1-3p(6)) \]

\[ + \frac{25}{94} s(6) + p^1(2),0(1-3p(6)) \]

\[ P_3(\gamma) = \frac{30}{156} = p^0(2),3(\gamma) + p^1(2),1(\gamma) \]

\[ = p^0(2),2(6)(1-s(6))^2 3p(6) + p^0(2),3(6)(1-s(6))^3 \]

\[ (1-2p(6)) + 2p^0(2),2(6)s(6)(1-s(6))(1-3p(6)) \]

\[ + p^0(2),1(6)(1-s(6))(1-2p(6)). \quad (3) \]

Upon using the fact that

\[ P_2(6) = p^0(2),2(6) + p^1(2),0(6) = \frac{25}{94} \quad (4) \]

and

\[ P_3(6) = p^0(2),3(6) + p^1(2),1(6) = \frac{10}{94}, \quad (5) \]

we can solve equations (3) to get estimates of \( p(6), s(6) \)

and \( P^i_r, r \geq 1(6) \) for \( i = 0, 1, \ldots, r/2 \) and \( r = 0, 1, 2, 3. \)

(Footnote continued from previous page) some of the loss of power that occurs when one bases his judgment about \( n \) on the sample distribution under consideration.
Since the sample value of $P_4^2(7)$ is very small ($= 3/156$) we set $P_4^2(2), 0 = 3/156$ and all other $P_4^2(2), t = 0$ equal to zero. $P_4^2(2), r = 0$ are of course zero for $r > 4$. Having thus gotten the detailed distribution of age 7, we now compute the expected frequencies at age 8. The only unknowns at this stage, namely $p(7)$ and $s(7)$, are estimated by using the first two frequencies at age 8, since no better methods are currently available. The expected frequencies are given in Table 42, column 3.

It is to be noted that the estimation of $P_4^2(2), r = 0$ is not based on the sample at age 8 and hence the chi-square value associated with the fit has degrees of freedom $5 - 2 - 1 = 2$, 5 being the number of classes, 2 being subtracted for having estimated $p(7)$ and $s(7)$ and 1 being subtracted to account for the fact that the total of the five expected probabilities add up to a known constant - unity in this case. Since our interest is to test the validity of the model concerning contagion, the error introduced by the estimation of $P_4^2(2), r = 0$ makes the calculated chi square an over estimate of the chi square that we would obtain if the distribution at age 7 were exactly specified. Hence, we can take the chi square obtained above as having Pearson's chi square distribution with 2 degrees of freedom without the fear of over appreciating our model with this criterion of goodness of fit.

Now, the expected frequencies at age 7 can be calculated
Table 42. Fitting of a stochastic model to the unilateral pitted surface cavitation count

<table>
<thead>
<tr>
<th>Count per child</th>
<th>Age 8 (years)</th>
<th>Age 9 (years)</th>
<th>Age 10 (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>20.00</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>24.00</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>67</td>
<td>66.00</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>32.80</td>
<td>40</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>3.58</td>
<td>1.66</td>
<td>.41</td>
</tr>
<tr>
<td>D.F.</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$\chi^2$ as in [11]</td>
<td>14.7</td>
<td>11.8</td>
<td>2.5</td>
</tr>
<tr>
<td>D.F. for this chi square</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$p(t)$</td>
<td>.061</td>
<td>.034</td>
<td>.031</td>
</tr>
<tr>
<td>$s(t)$</td>
<td>.45</td>
<td>.055</td>
<td>.016</td>
</tr>
</tbody>
</table>

From those at age 8 by using (?), but since the expected frequencies at age 8 themselves were based on the estimated and not the exact distribution at age 7, this procedure will lead to high inefficiency. Also, we cannot use a method analogous to the one above of taking help of the samples at ages 7 and 8 since firstly $p_4(8), q_4(8)$ are not estimable and secondly they are not too small to be neglected. Hence, we propose the
following method:

Take the observed proportion \( \hat{P}_r(8) \) of count \( r \) at age \( 8 \) and get the new estimate of \( \hat{P}_r(\phi), r-21(8) \) by using the formula

\[
\hat{P}_r(\phi), 4-21(8) = \frac{\hat{P}_r(\phi), r-21(8)}{\hat{P}_r(8)} \hat{P}_r(8)
\]

(6)

where \( \hat{P}_r(\phi), 4-21(8) \) in (6) is the value of \( \hat{P}_r(\phi), 4-21(8) \) as calculated above. The expected frequencies obtained by this procedure are given in Table 4, column 5.

Observe that the values of chi square for these fits are very much smaller than those given by Grainger and Reid (cf. line 3b, Table 4). The expected frequencies obtained by using the same procedure are given for age 10 in Table 3, column 7. The small chi square, while asserting the above statement, gives faith that the ad hoc method of estimation used in the process is not too efficient.

4. Conclusions regarding the contagion of caries

On the basis of the above fitting, we observe that caries are highly contagious and that the caries spread from one tooth to its nearest neighbor. The decreasing values of \( p(t) \) and \( s(t) \) indicate that the teeth become less susceptible to decay with age. The large number of caries in the mouth of an elderly person can be attributed to the facts that (i) decayed teeth do not heal and (ii) the presence of the increasing number of decayed teeth increases the probability
of a non-carious tooth decaying.

D. Effect of Plot Size and Shape

Let us first continue with the distribution of caries mentioned above and then generalize it to general contagious phenomenon.

As an extension of the model in section C for a changed plot size, we may consider the bilateral pitted surface cavitation counts. The pitted teeth of an individual may be regarded as comprised of $m$ disjoint sets of four teeth each. For example, a typical set may consist of a tooth on the left side, its counterpart on the right hand side and the two teeth opposite to them. In this case, a set can be in any one of the following seven phases: (1) no teeth carious, (2) only one tooth carious, (3) a tooth and its counterpart carious, (4) a tooth and its opposite carious, (5) a tooth and the one opposite to its counterpart carious, (6) any three teeth carious and (7) all four teeth carious. The individuals with $r$ caries are divided into subclasses, one corresponding to each of the possible vectors of the form $(r(1), r(2), \ldots, r(7))$ where $r(i)$ refers to the number of sets in the $i^{th}$ phase. It is easy to see that we can develop a model by assuming that the decaying in one set in no way affects the decaying in other sets. It should be apparent that this distribution can be very different from that of the
unilateral pitted surface cavitation count. In this perspective, it is not surprising if a family of distributions does not fit all the samples taken from a contagious phenomenon when the plot sizes for the samples differ.

In order to bring out the effect of plot size and shape on the sample data more clearly, we consider now the distribution of larvae in, say, a large rectangular corn field. Suppose that the field is divided into equal squares by drawing lines parallel to the edges of the field and that corn is planted at the vertices of these squares. Regarding the spread of the larvae, let us make the following assumptions:

(i) Egg masses are laid on the corn stalk according to a Poisson distribution with p.g.f. exp \( \lambda_1(z - 1) \).

(ii) Within a given egg mass, the distribution of larvae hatching and surviving is a Poisson distribution with p.g.f. exp \( \lambda_2(z - 1) \).

(iii) The probability that a larva stays on the plant on which it was born is \( p \) and the probability that it leaves is \( q (= 1 - p) \).

(iv) When a larva leaves a stalk, it travels with equal probability towards the eight plants which are almost symmetrically situated with respect to the plant on which the larva is presently situated.

(v) The physical strength of a larva is such that it cannot travel any further than the nearest stalk.
Consider an arbitrary cornstalk A. Assumptions (i) and (ii) imply that the distribution of larvae, hatched on A and surviving is the Neyman Type A distribution with p.g.f.

\[ \exp \left\{ \lambda_1 \left[ \exp(\lambda_0(z - 1)) - 1 \right] \right\}. \]  

Since by assumption (iii), the p.g.f. of the larva on A staying on it is the degenerate Binomial,

\[ q + pz. \]  

The p.g.f. of the number of larvae hatched on A and staying on A can be obtained by generalizing (1) with (2) as

\[ \exp \left\{ \lambda_1 \left[ \exp(\lambda_2(q + pz - 1)) - 1 \right] \right\} = \exp \left\{ \lambda_1 \left[ \exp(\lambda_2(z - 1)) - 1 \right] \right\}. \]  

Since the p.g.f. of the larvae hatched on a stalk is the one given in (1), the p.g.f. of the number of larvae hatched and surviving on the eight plants surrounding A is given by

\[ \left\{ \exp \left\{ \lambda_1 \left[ \exp(\lambda_2(z - 1)) - 1 \right] \right\} \right\}^{\frac{1}{8}} = \exp \left\{ \frac{1}{8} \lambda_1 \left[ \exp(\lambda_2(z - 1)) - 1 \right] \right\}. \]  

By assumption (iv) and (v) it is apparent that the larva hatched on any one of these plants reaching A is \( \frac{c}{8} \), the p.g.f. of the larvae, hatched on these plants and reaching A can be obtained by generalizing p.g.f. (4) with the Binomial

\[ (1 - \frac{c}{8}) + (\frac{c}{8})z \]  

as

\[ \exp \left\{ \frac{1}{8} \lambda_1 \left[ \exp(\frac{\lambda_2c}{8}(z - 1)) - 1 \right] \right\}. \]
Since the total number of larvae on A is made up of the larvae hatched and staying on A and the larvae migrating to A, the p.g.f. of the total larvae on A can be obtained by taking the product of the p.g.f.s (3) and (6) as
\[ \exp \left\{ \lambda_1 \left[ \exp \left( \frac{\lambda_2 p(z-1)}{6} \right) - 1 \right] \right\} \exp \left\{ 8 \lambda_1 \left[ \exp \left( \frac{\lambda_2 q}{6} (z-1) \right) - 1 \right] \right\} . \]

(7)

If now our plot is such that it includes \( r^2 \) stalks, which form a compact \( r \times r \) square, then the larvae found on this plot can be divided into the following three components:

1. The larvae hatched on the central \((r-\delta) \times (r-\delta)\) square which cannot leave the square due to assumption (v). The p.g.f. of their number will be
\[ \exp \left\{ \lambda_1 \left[ \exp \left( \frac{\lambda_2 (z-1)}{6} \right) - 1 \right] \right\}^{(r-\delta)^2} \]
\[ = \exp \left\{ (r-z)^2 \lambda_1 \left[ \exp \left( \frac{\lambda_2 (z-1)}{6} \right) - 1 \right] \right\} . \]

(8)

2. The larvae hatched on the bordering \(4(r-1)\) plants and not migrating outside the square. The p.g.f. of the larvae hatching is obviously
\[ \exp \left\{ 4(r-1) \lambda_1 \left[ \exp \left( \frac{\lambda_2 (z-1)}{6} \right) - 1 \right] \right\} . \]

(8a)

Assuming \( r \) to be large and neglecting the end errors, we can say that such a plant has five stalks belonging to the plot as neighbors and three stalks not belonging to the plot as neighbors. From assumptions (iv) and (v), therefore, we deduce that the probability of a larva hatched on the bordering plant
staying in the plot

= probability it will stay on the plant

+ probability it will migrate to one of the five plants belonging to the plot

= p + 5a/8.

Hence the p.g.f. of a larva staying by this process is the Binomial

\[(1 - (p + 5a/8)) + (p + 5a/8)z.\]  \hspace{1cm} (10)

The p.g.f. of the total number of larvae staying in the plot after getting hatched on the bordering plot is obtained by generalizing p.g.f. (8a) with (10) as

\[\exp\{4(r - 1)\lambda \left[\exp\left(\frac{\lambda a}{p}(p + 5a/8)(z - 1) - 1\right)\right]\}.\]  \hspace{1cm} (11)

(3) Larvae hatched on the 4(r + 1) stalks surrounding the plot and entering the plot for immigration. The p.g.f. of the larvae hatched on these plants is obviously

\[\exp\{4(r + 1)\lambda \left[\exp\left(\frac{\lambda a}{p}(z - 1) - 1\right)\right]\}.\]  \hspace{1cm} (12)

Again, by neglecting the end errors, we say that the probability of a larva on such a plant entering the plot is 3a/8 and by using arguments similar to those in section (2), we deduce the p.g.f. of the immigrants to the plot as

\[\exp\{4(r + 1)\lambda \left[\exp\left(\frac{\lambda a}{p}(z - 1)\right) - 1\right]\}.\]  \hspace{1cm} (13)
Hence the p.p.f. of the total number of larvae on
the plot is the product of the p.p.f.s (8), (11)
and (13) as
\[
\exp\left\{ (r - 2)^2 \lambda_2 \left[ \exp\left( \lambda_2 (z - 1) \right) - 1 \right] + 4(r - 1) \lambda_1 \\
\left[ \exp\left( \lambda_1 \rho + 5\alpha/2 \right) (z - 1) - 1 \right] + 4(r + 1) \lambda_1 \\
\left[ \exp\left( \lambda_2 q/8 \right) (z - 1) - 1 \right] \right\}.
\]

It should be apparent from the p.p.f.s (7) and (14)
that a change in plot size affects the distribution
of the sample considerably. It is to be noted that
with the change in shape, components (9) and (3)
mentioned above change and consequently the final
distribution. In this perspective, it may be of
interest to observe that McGuire et al. [16] also
find that the distribution which fits the corn borer
data with a plant as a plot is different from the
one which fits the data when a group of plants is a
plot.
IX. CONCLUSIONS

In Chapter IV, efficiency of certain ad hoc methods of estimation were discussed for the Neyman Type A and the Pascal distributions. When the distributions have very small mean and variance, the method of the first moment and the first frequency and that of the first moment and the ratio of the first two frequencies were found to have high efficiency and were more efficient than the method of moments. For a population with large mean and variance, the method of moments has superiority over the other two methods. But since the efficiency of this method relative to maximum likelihood is low in such cases, this relative superiority is not of any practical importance.

In view of this it was decided to combine the information on more than two statistics to obtain better methods of estimation. The method used for combining was that of minimum chi square. The efficiency of using the following three sets of statistics have been studied: (i) the first three factorial cumulants, (ii) the first two factorial cumulants and the logarithm of the first frequency and (iii) the first two factorial cumulants and the ratio of the first two frequencies. All of these methods are more efficient than the above methods which use only two statistics, throughout the parameter space. Particularly, the efficiency of the method using the set (ii) of statistics is remarkably high in a
large region in the parameter space. Since this method is much simpler than the method of maximum likelihood, it can be advantageously used in practical problems.

In Chapter V, efficiency of certain *ad hoc* methods of estimation were discussed for the three parameter families, (i) Poisson v Pascal and (ii) Pascal v Poisson. It was evident from the tables of efficiency that in a sufficiently large region near the origin, the method of the first two moments and the first frequency was very efficient and was much better than the method of moments. Since the method of maximum likelihood is highly cumbersome, the method of using the first two moments and the first frequency looks promising.

It was apparent in Chapter IV that the complexity of the method increases with the number of statistics and that when the number of statistics used is small, the methods using various sets of statistics are good in different regions. In Chapter VI, a preliminary test procedure was developed which asymptotically selected the best of the methods of estimation. It was shown that a preliminary test partition for which the mean square error of the estimate of one of the parameters is least is a second degree curve with the vector of the parameters as centre.

In Chapter VII, a measure of the reliability of a statistic was developed with the intention of selecting the statistics which rate high in reliability according to this measure.
It was found that in a number of cases, the statistics which had high reliability produced highly efficient estimates.

We hope to pursue the study of both the preliminary test and the reliability function and believe that with their help, we will be able to get simple estimates with relatively high efficiency.

In the above discussion, we dealt with the case in which there is a single sample available to the statistician. In many a contagious phenomenon, a number of samples taken at different times are available. With the help of the data on the distribution of dental caries, a method was indicated in Chapter VIII for formulating a null hypothesis using the previous samples. Since "true contagion" is visualized as something that changes the nature of the population with time, it is clear that such a method will yield more definitive information on contagion. In the particular case of the dental caries, our conclusions are that the caries are contagious and that the teeth become more and more resistant to caries with age.

It was also shown that when there is true contagion, plot size and shape can greatly influence the form of the population.

The form of the contagious distribution changes from time to time and from one system of sampling to another and is, in general, too complex to permit simple maximum likeli-
hood estimates. It is hoped that the above discussion of the efficiency of various methods of estimation, while restricted to particular families of distributions here, will be helpful at large, in formulating methods of estimation that will make the fitting of such complex families feasible.
X. LITERATURE CITED


XI. ACKNOWLEDGEMENT

The author takes great pride in acknowledging the considerable help given by Dr. John Gurland, both as the Professor in Charge of Major Work and as a benevolent friend. Dr. John Gurland's sympathetic interest in the author's work has been a continued source of inspiration to him.

The author is also grateful to Dr. Bancroft who made his stay at Ames possible, and to the staff and students for their high cooperation in all walks of life.

The financial help of the Air Force and the National Science Foundation is also greatly appreciated.