ESTIMATION OF THE PARAMETERS OF A MIXTURE OF A POISSON AND A BINOMIAL DISTRIBUTION

A Thesis

Presented To

The Faculty of Graduate Studies

University of Manitoba

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

William Allen Holtslag

April 1969

c1969

THE UNIVERSITY OF MANITOBA
ELIZABETH DAFOE LIBRARY

ABSTRACT

Title: Estimation of the Parameters of a Mixture of a Poisson and a Binomial Distribution

Author: William Allen Holtslag

After introducing the concept of mixtures of distributions, practical uses, definitions, methods of estimation in general and the method of moments in particular, the mixture of a Poisson and a Binomial distribution is studied.

Fisher-Consistent moment estimators are determined for the cases when one parameter at a time is assumed known:

α known

$$\hat{\lambda} = \frac{(n-1)F_1}{n-\alpha} + \frac{1}{n-\alpha}\left\{ \frac{(1-\alpha)n}{\alpha}\left[F_2(n-\alpha) - (n-1)F_1^2\right]\right\}^{\frac{1}{2}}$$

$$\hat{p} = \frac{F_1}{n-\alpha} - \frac{1}{n(n-\alpha)}\left\{\frac{\alpha n}{1-\alpha}\left[F_2(n-\alpha) - (n-1)F_1^2\right]\right\}^{\frac{1}{2}}$$

λ known

$$\hat{p} = \frac{(\lambda^2 - F_2) - B}{2(n-1)(\lambda - F_1)}$$

$$\hat{\alpha} = \frac{2(1-\frac{1}{n})\ F_1(\lambda - F_1) - (\lambda^2 - F_2) + B}{2(1-\frac{1}{n})\lambda\ (\lambda - F_1) - (\lambda^2 - F_2) + B}$$

p known

$$\hat{\lambda} = \frac{F_2 - b + C}{2(F_1 - a)}$$

$$\hat{\alpha} = \frac{2(F_1 - a)^2}{F_2 - b - 2a(F_1 - a) + C}$$

The variances of the moment estimators were then determined.

In order to make an empirical study of the efficiencies of the moment estimators, we found the Cramér-Rao lower bound for the variances, and determined ARE and JARE for the three cases. In general ARE $\to 0$ and JARE $\to 0$ as $n \to \infty$.

For the case of three unknown parameters, simplifying assumptions were made to solve for the moment estimators. The procedure for comparing the efficiencies was then outlined, but no study was made of empirical data.

A flow chart and program for the empirical study are presented and some tables of results are given. It is proposed that in further study, one could look for modifications of the moment estimators, which would yield unbiased estimators.

TABLE OF CONTENTS

PAGE

CHAPTER I

1.1 Introduction

In the last decade interest has been steadily growing in the area of mixtures of distributions, both in mathematical and in practical aspects. Mathematically, mixtures pose interesting problems such as in solving moment estimating equations for consistent estimators of the parameters involved in the mixture. Of particular interest would be a solution for the three parameter case of a mixture of a Poisson and a Binomial distribution. We will take a brief look at this problem in a later chapter. Also, mathematically, mixtures of normal distributions are encountered in the course of determining limit distributions of sums of interchangeable random variables. There are many practical aspects of mixtures which are receiving much attention. Mixtures occur in the theory and applications of probability and statistics. One is interested in the distribution of a random variable X, but knows only the conditional distribution of X given the values of some auxiliary random variable Y. Following from this then, the desired distribution of X is simply a mixture of the known conditional distributions. Applications for continuous or discrete distribution mixtures are very widespread including such areas of study as Biology, Chemistry, Marketing, Life and Acceptance testing, and mixtures of distributions are a fundamental aspect of the Bayesian approach to statistical inference.

In the Biological sciences there are many areas in which conditional arguments become directly involved. For example, in studying conservation, ecology, taxonomy, the investigation of natural populations is in effect a study of mixtures of distributions where often the unobservable variate is age or species. Chapman (1961) discussed thoroughly fish biology. In genetics and plant and animal breeding, inheritance and genotype studies are done through the study of mixtures. Genetic statistical analyses of populations involve both conditional distributions and the unconditional distribution of genotypes.

In other areas, Anscombe (1961) gave an application for mixtures in a marketing problem. Blischke (1962) described an ICBM weapons system and Edwards and Gurland (1961) reported an application of a mixture of bivariate Poisson distributions to an investigation of accident proneness. In Chemistry, Medgyessy (1961) used finite mixtures of normal distributions in the investigation of absorption sprectra and of electrophoretic separation of proteins. Life testing experiments were conducted by Kao (1959) on electron tubes subject to sudden and delayed failure. Acceptance testing procedures were developed by Barnard (1954) and Vagholkar (1959) for the case when the true proportion of defectives varies from batch to batch. Hald (1960) constructed sampling inspection procedures based on prior distributions. His analyses involve both the compound hypergeometric and compound binomial distributions. In the Bayesian approach to statistical inference the mixing distribution is called an a priori distribution.

The mixture we consider in this thesis, mixture of a Poisson and a Binomial, could be encountered in many of the above areas. Investigation of reaction rates in Chemistry or Life testing of machinery might involve this particular mixture.

Before going into a review of the literature, a definition of a mixture and one of identifiability will be stated. Some synonyms for a mixture are also given.

Definition 1.1

The definition of a mixture as stated by Teicher (1960) for the discrete case is: Let $P = \{ P_\lambda ; \lambda \varepsilon \Lambda \}$ be a family of discrete probability distributions, each member of P being indexed by a finite dimensional parameter $\lambda = (\lambda^{(1)}, \ldots, \lambda^{(s)})$ belonging to some measurable subset Λ of R^s. Suppose that for each $P(.;\lambda) = P_{\lambda \varepsilon} P$ the set of points to which P_λ assigns positive probability is independent of $\lambda^{(1)}, \ldots, \lambda^{(s)}$. Let $\mathcal{G} = \{ G(\lambda^{(1)}, \ldots, \lambda^{(s)}) \}$ be a class of s-dimensional distribution functions for which $\int_\Delta dG(\lambda) = 1$ for each $G \varepsilon \mathcal{G}$. Then for each $G \varepsilon \mathcal{G}$

$$P(x) = P_G(x) = \int_\Delta P(x;\lambda)\, dG(\lambda) \qquad \text{------ (1)}$$

is again a discrete probability distribution; in fact, P_G assigns positive probability to the union of points of positive probability under each λ. The distribution P_G defined in equation (1) is called a mixture, or a μ_G-mixture of distributions in P, under the condition that μ_G does not assign probability one to any single P_λ in P. Following Teicher, the family $\mathcal{H} = \mathcal{H}(P)$ of mixtures P_G resulting as G ranges over all of $\mathcal{G}$ will be called the class of mixtures of P.

Definition 1.2

The analysis of mixtures of distributions and the unique characterization of mixtures depends on the identifiability of mixtures: A mixture $P = P_G$ is said to be identifiable if the relationship

$$P = \int P(x;\lambda) dG(\lambda) = \int P(x;\lambda)\, dG^*(\lambda) \qquad \text{------ (2)}$$

implies $G^* = G$ for all G^* in $\mathcal{G} \cup \mathcal{L}$, where $\mathcal{L}$ is the family of distribution functions whose corresponding Lesbeque-Stieltjes measure assigns measure one to a single point of R^s. The class $\mathcal{H}$ is itself called identifiable if every member of it is identifiable.

Questions of estimation or hypothesis testing cannot be studied before the identifiability of the mixture being studied has been established. Suitable restrictions, on P or $\mathcal{G}$ or both can always result in identifiable families for which estimation and hypothesis testing are meaningful.

Identifiability for mixtures of Poisson distributions was shown by Feller (1943). The identifiability of mixtures of binomials follows exactly. However, if the family being mixed is $P = \{ P(x;n,p) : p \varepsilon (0,1) \}$ where n is fixed, the resulting class of mixtures is not necessarily identifiable. It is sufficient that μ_G give positive measure to not more than $r \leqslant \underline{n+1}/2$ distinct points. Since this condition is also necessary for identifiability, the estimation problem when r is unknown is quite difficult.

A listing of synonyms for mixtures of distribution would include: Compound distribution, Composite distribution, sum of distributions, Probability mixture, and superposition. Convolution is a special case of a mixture of distributions. Gurland (1957) presented a study of generalized distributions and their relational concepts.

1.2 Literature Review

The first significant work on the estimation of the parameters in a mixture of probability distributions was presented by Pearson (1894). He studied the mixtures of two normal distributions. Using the first five moments of the mixture of two normal distributions, Pearson extracted the roots of a ninth degree equation. A suitably chosen root of the nonic was used to solve the moment estimators for the five parameters: μ_1, μ_2, σ_1^2, σ_2^2, α_1

of the density

$$f(x) = (2\pi)^{-\frac{1}{2}} \sum_{i=1}^{2} \frac{\alpha_i}{\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x-u_i}{\sigma_i}\right)^2\right\}$$

where the α_i are positive numbers summing to unity. Pearson also laid the ground work in the estimation problems of mixtures of discrete distributions. For a mixture of two binomials, known n and unknown P_i, α_i, that is in the distribution

$$P(x) = \binom{n}{x} \sum_{i=1}^{2} \alpha_i\, p_i^{\,x} (1-p_i)^{n-x} \qquad x = 0,1,\ldots,n$$

Pearson (1951) constructed the moment estimators. He also studied the moment estimators for a mixture of two Poissons and generalized the moment equations for a mixture of r distributions, r known for both cases.

Schilling (1947) considered a frequency distribution represented as the sum of two Poisson distributions, the mean of neither being assumed known. A method for the dichotomy and subsequent summation is described and twenty-nine distributions to which the Poisson was supposedly a good fit are analyzed. In the following year Robbins published a paper on mixtures of distributions and the problems of estimation. Robbins (1950) gave a generalization to the method of maximum likelihood for estimating a mixing distribution. Meanwhile, Shenton (1949, 50, 62) worked on the efficiency of the method of moments, maximum likelihood and the efficiency of the method of moments and the asympototic biases of the moment estimators with an application to the negative binomial distribution. In general the properties of the moment estimators are expected to be similar to those based on the first few moments (means, covariances) when r is small and similar to the maximum likelihood estimates when r is large.

Teicher (1960, 61, 63) was the first to seriously consider the problem of identifiability of mixtures. If a mixture is not uniquely determined, one certainly cannot hope to be able to estimate parameters and thereby 'identify' the mixture nor make a choice among possible indistinguishable hypotheses. Teicher firstly concerned himself with the mixtures of distributions. He defined a weighted average of cumulative distribution functions as a mixture. That is, if $\mathcal{F} = \{F\}$ is a family of distribution functions and μ is a measure on a Borel Field of subsets of with $\mu(\mathcal{F}) = 1$, then $\int F(.)d\mu(F)$ is again a distribution function which is called a μ-mixture of $\mathcal{F}$. He then considered: convergence questions when either F_n or μ_k (or both) tend to limits where $\mathcal{F}$ is indexed by a finite number of parameters, mixtures of additively closed families which are shown to be closed under convolution, and he gave necessary and sufficient conditions for a μ-mixture of normal distributions to be normal. Generation of mixtures is discussed and the concluding remarks link the problem of mixtures of Poisson distributions to a moment problem. His subsequent papers were in connection with the problem of identifiability of mixtures (1961) and the refinement of the case of finite mixtures (1963). Theorems on identifiability of all finite mixtures of Gamma (or of normal) distributions are established. Thus the estimation of the mixing distribution on the basis of observations from the mixture is feasible in these cases. Conditions for the identifiability of binomial mixtures are also presented. Boes (1963) gave necessary and sufficient conditions for finite mixtures to be identifiable.

Rider (1961) obtained the moment estimators for the parameters of mixtures of two Poissons, two Binomials and two Weibull distributions.

Blischke (1962) extended Rider's work on moment estimators for mixtures of two binomial distributions, and then generalized the results (1964) for a mixture of r Binomials. His papers include studies of the properties of the moment estimators. A further look at Blischke's work will follow.

Katti and Gurland (1962) presented results on some methods of estimation for the Poisson binomial distribution and on the efficiency of certain methods of estimation for the Negative Binomial and Neyman type A distribution. Cohen (1963) was the next to give some estimation procedures in mixtures of discrete distributions. He considered the estimation of the parameters of a mixture of two Poissons and arrived at some simplification of Rider's work by the use of factorial moments rather than making use of the ordinary moments to form the estimating equations. He also studied estimators based on the first two sample moments and the zero sample frequency, and the mixed truncated Poisson distributions with missing zero classes.

The major part of our thesis is to consider Cohen's work on the mixture of a Poisson and a Binomial distribution along the lines of work done by Blischke on mixtures of Binomials. We will discuss some asymptotic properties and the efficiency of the estimators in different situations.

In his paper on the mixture of two Binomial distributions, Blischke (1962) discussed the method of moments for estimation of the parameters of the mixture. After deriving the estimating equations for the moment estimators, a study is made of the asymptotic behavior of these estimators. Tables are presented for the asymptotic relative efficiency (ARE) and joint asymptotic relative efficiency (JARE) of the estimators both in the general case and for the special case when the mixing proportion is known.

An extensive review of literature in the field of mixtures of distributions was done by Blischke (1963). After defining the concepts of mixtures and identifiability, he discussed various aspects of estimation and of testing of hypotheses in this connection. Perhaps, the most salient feature of the paper is the wide range of the bibliography appended to the article. Blischke (1964) generalized the moment estimation method for mixtures of r Binomial distributions. He also discussed several other methods of estimation of the parameters: Best asymptotically normal (BAN) estimators based on the moment estimators $\hat{\theta}$ as suggested by the work of Le Cam (1956), the construction of BAN estimators by Neyman's linearization technique. ARE and JARE are computed for several cases. Estimators constructed for the case when the mixing proportion is known have ARE unity for large n (and are maximum likelihood for $n = 2$).

The following chapter presents some methods of estimation for the parameters of a mixture, in particular the method of moments, which will be used for the Binomial and Poisson mixture.

CHAPTER II

2.1 Methods of Estimation

The difficulty generally in dealing with parametric estimation for finite mixtures is that the standard asymptotically efficient estimation techniques like Maximum likelihood, Minimum χ^2, yield quite intractable equations. Maximum likelihood equations often take the form of an infinite series in the unknown parameters. Katti and Gurland (1962a,b) discussed several aspects of the methods of estimation as modifications to those proposed by others such as Sprott (1958) and Gurland-Shumway (1960).

Sprott described a procedure for fitting the Poisson binomial distribution by the method of maximum likelihood. The Poisson binomial probability density function is:

$$P(x) = e^{-a} \sum_{t=o}^{\infty} \frac{a^t}{t!} \binom{nt}{x} p^x (1-p)^{nt-x} \qquad x = 0,1,2,\ldots$$

where n is known, and the parameters to be estimated are a and p. This distribution has mean apn and variance apn $[1+(n-1)p]$. The moment estimators as derived by McGuire (1957) et al. for a and p are:

$$a^* = \frac{(n-1)\bar{x}^2}{n(S^2-\bar{x})}$$

$$p^* = \frac{(S^2-\bar{x})}{(n-1)\bar{x}}$$

where $\bar{n}$ and S^2 are sample means and variance respectively. Sprott described a procedure for fitting the Poisson binomial distribution by the method of maximum likelihood and considered the efficiencies of the method of moments and the method of sample zero frequency. The maximum likelihood equations given by Sprott are:

$$\Sigma a_x \frac{\partial \log P(x)}{\partial a} = -N + \Sigma a_x \frac{S_2}{S_1} = 0$$

$$\Sigma a_x \frac{\partial \log P(x)}{\partial p} = \Sigma \frac{a_x}{S_1} \frac{\partial S_1}{\partial p} = 0$$

where a_x is the observed frequency of x and N is the total number of observations, and

$$S_1(x) = \sum_{t=0}^{\infty} \frac{a^t}{t!} \binom{nt}{x} p^x (1-p)^{nt-x}$$

$$S_2(x) = \frac{S_1}{a} = \sum_{t=0}^{\infty} \frac{ta^{t-1}}{t!} \binom{nt}{x} p^x (1-p)^{nt-x}$$

The maximum likelihood equations can be written in the form:

$$n\hat{a}\hat{p} = \bar{x}$$

$$L(\hat{p}) = \Sigma a_x f(x) - N = 0$$

where

$$F(x) = \frac{(x+1)P(x+1)}{n\hat{a}\hat{p}\, P(x)}$$

and

$$P(x+1) = \frac{nap}{(x+1)!} \sum_{t=0}^{x} \binom{x}{t} (n-1)^{[x-t]} P^{x-t} (1-P)^{n-x+t-1}\, t!\, P(t)$$

where

$$n^{[x]} = \frac{n!}{(n-x)!}$$

The method of moments may provide quite misleading estimates for the parameter p when this parameter is small, while the method of sample zero frequency remains reasonably efficient for considerably large values of p.

Likewise, Gurland and Shumway have looked at the method of moments and the method of sample frequencies with the conclusion that they give simple equations for obtaining estimates, but that the fit they provide is unsatisfactory. The parameters should therefore be estimated by an efficient method, maximum likelihood whenever possible. They then presented a simplification of the maximum likelihood estimation procedure. It involves rewriting the maximum likelihood and recurrence relations in terms of ratios of Poisson factorial moments and tabulating these ratios for values of the parameters when n = 2. The simplifying iterative procedure is to estimate a value of $\hat{p}_1$ by some relatively simple method such as sample moments or frequency. By using this value in a recurrsive formula, a value for $\hat{p}_{i+1}$ is found and substituted into the likelihood equations to obtain $L(\hat{p}_{i+1})$ and $L'(\hat{p}_{i+1})$ where

$$L(\hat{p}) = \frac{1}{n\hat{a}\hat{q}} \Sigma\, a_x\, P(x) - N = 0 \qquad \hat{a},\hat{q} \text{ MLE}$$

and differentiation yields

$$L'(\hat{p}) = \frac{1}{n^2 \hat{a}\hat{p}\hat{q}}\left[(n-1)\Sigma\, a_x P(x) - \frac{(n-1)\hat{p}+1}{\hat{q}}\, \Sigma\, a_x\, q(x)\right]$$

where

$$q(x) = p(x) \quad (P(x+1) - P(x))$$

$\hat{p}_{i+2}$ is given then by the recurrsive formula. Iteration will be discontinued when no substantial change is produced in the estimates.

Katti and Gurland (1962a) first look at Sprott's approach of using the first two moments or the first moment and the first frequency. Then they discussed a minimum χ^2 method using the first two factorial cumulants and the logarithm of the zero frequency in place of the moments and the first frequency since this leads to relatively simple equations to solve. This procedure appears to be a reasonable substitute for the highly complex asymptotically efficient methods in most practical cases.

In a further study Katti and Gurland (1962) considered the asymptotic efficiency of methods of Fisher (1941), Anscombe (1950), Evans (1953) (alternate to maximum likelihood estimation) which have been found to be very useful in the problem of fitting. The method of moments as modified by the latter three authors, namely, the first moment and the first frequency, or the first moment and the ratio of the first two frequencies, involve only two statistics and are highly efficient in very restricted regions. Katti and Gurland have introduced methods which have high efficiency over large regions in the parameter spacing using the results of Barankin and Gurland (1951) and Fergusson (1958).

For m statistics $(s_1, \ldots, s_m) = s$, Barankin and Gurland minimized the quadratic form:

$$Q = (s-\theta)\ \hat{\Omega}^{-1}\ (s-\theta)'$$

where

$$\Omega = E[(s-E\,[s])(s-E[s])']$$

is the covariance matrix of s and $\hat{\Omega}$ a nonsingular matrix is a consistent estimator of Ω. The quantity Q is referred to as a χ^2 for s and the estimates so obtained, are referred to as the minimum χ^2 estimators. The high increase in efficiency due to the inclusion of the additional statistics may well justify the additional work involved.

From Cramér (1946) minimum χ^2 estimators are constructed by grouping the data, computing (as a function of the unknown parameters) an expected frequency for each group, forming a χ^2 based on expected and observed group frequencies in the usual way, and minimizing this quantity with respect to the unknown parameters. This method generally yields asymptotically efficient estimators.

A couple of lesser known methods are those of Cassie (1954) who used a graphical method on probit paper with a 'partitioning method' which was inefficient even with respect to the method of moments, and Medgyessy (1953.54.61) who proposed a 'variance reduction method' consisting of an entirely analytic construction.

Blischke (1964) discussed estimation procedures for $n \geqslant 2r-1$ and derived efficient estimators as functions of the moment estimators by (1) expansion of χ^2 about the moment estimators yielding BAN estimators, (2) Neyman's linearization technique, (3) using Fisher's 'information' to compute a correction factor for the moment estimators.

2.2 Method of Moments

The method of moments can be used to estimate the population parameters from a large sample, Rao(1952). Consider $\{Y_i\}_1^n$ independently identically distributed random variables from a distribution with unknown parameters $\Theta_1, \ldots, \Theta_q$, let the first q raw moments of the distribution exist as explicit functions $\alpha_r(\Theta_1, \ldots, \Theta_q)$ $r = 1, \ldots, q$. If $a_r = \Sigma\, y_i^r/n$ denotes the moment functions, then the method of moments consists of equating the realized values a_{ro} in a sample and the hypothetical moments:

$$\alpha_r(\Theta_1, \ldots, \Theta_q) = a_{ro} \qquad r = 1, \ldots, q$$

and solving for $\Theta_1, \ldots, \Theta_q$.

Since a_r is the mean of n random variables and if $E[y_i^r]$, the r^{th} raw moment, exists, then by the law of large numbers: $a_r \to \alpha_r(\Theta_1, \ldots, \Theta_q)$ with probability 1, so that a_r is a consistent (besides being unbiased) estimator of α_r.

It is easy to see that if the correspondence between $\Theta_1, \ldots, \Theta_q$ and

$\alpha_1,\ldots,\alpha_q$ is one-to-one and inverse functions

$$\Theta_i = f_i(\alpha_1,\ldots,\alpha_q) \quad i = 1,\ldots,q$$

are continuous in $\alpha_1,\ldots,\alpha_q$, then

$$\hat{\Theta}_i = f_i(a_{1o},\ldots, a_{qo}) \quad i = 1,\ldots,q$$

are solution of

$$\alpha_r(\Theta_1,\ldots,\Theta_q) = a_{ro}$$

and $f_i(a_1,\ldots,a_q)$ is a consistent estimator of Θ_i, $i = 1,\ldots,q$.

The estimators obtained by this method are not generally efficient. For some numerical computations of the loss of efficiency we may refer to Fisher (1922).

Before deriving the moment estimators, one must define the factorial moments. Blischke (1962) defines sample factorial moments for the mixture of Binomials as

$$F_k = \frac{1}{m} \sum_{i=1}^{m} \frac{y_i(y_i-1)\ \ldots\ (y_i-k+1)}{n(n-1)\ \ldots\ (n-k+1)}$$

while Cohen (1963) defined the sample factorial moments for the mixture of a Poisson and a Binomial as

$$F_k = \frac{1}{m} \sum_{y=o}^{R} y_i(y_i-1)\ \ldots\ (y_i-k+1)m_y$$

where m_y is the sample cell frequency of y and

$$\sum_{y=o}^{R} m_y = m.$$

Blischke (1962) shows that the expected value of the sample factorial moment is the population factorial moment.

Moment estimators for the (2r-1) parameters of a mixture of r binomials may be constructed as:

Choose any set of (2r-1) distinct population factorial moments and consider these as equations in the parameters. Solve these (2r-1) equations for the parameters. Finally substitute the sample factorial moments for the population factorial moments in this solution to specify the moment estimators. So in general for finite mixtures

$$P(y) = \sum_{i=1}^{r} \alpha_i P(y; \lambda_i)$$

where λ_i are distinct and of s-dimension and α_i sum to unity. Here there are $N = r(s+1)-1$ parameters and the method of moments involves writing any N population moments (usually the first N are used) as equations in the N unknown parameters as functions of the moments and finally defining estimators by substituting sample moments for population moments in this solution.

For the special case of one or more parameters being known, unless the α_i are equal, simply knowing the α_i numerically is not sufficient for estimating the parameter p_i. It must be known in addition that α_i is specifically the proportion in the mixture of the population having the i^{th} smallest p. This situation prevails regardless of the estimation procedure employed. If it is unknown which α_i goes with which p_i, the estimators may not be consistent, Rider (1961). The procedure in this case for finding the moment estimators is analogous to that given for the general case above.

2.3 Asymptotic Properties of Moment Estimators

As already mentioned, moment estimators are consistent. In addition it can be shown that they are also asymptotically normal with means $\underline{\hat{\Theta}}$ and covariance matrix Σ where

$$\sigma_{ii'} = \sum_{j=1}^{2r-1} \mu_2(F_j)\delta_{ij}\,\delta_{i'j} + \sum_{\substack{j,j'=1\\ j\neq j'}}^{2r-1} \mu_{11}(F_jF_{j'})\,\delta_{ij}\,\delta_{i'j'}$$

where

$\mu_2(F_j)$ is the variance of the j^{th} factorial moment

$\mu_{11}(F_jF_{j'})$ is the covariance,

and

$$\{\delta_{ij}\} = \left\{\frac{\partial H_i}{\partial F_j}\bigg|_f\right\} \qquad \begin{matrix} i=1,\dots,r \\ j=1,\dots,r \end{matrix}$$

where the H_i are the moment estimating equations.

The variance matrix shall be derived for the special cases of one known parameter in the following chapter.

The asymptotic relative efficiency (ARE) of a consistent asymptotically normally distributed estimator $\hat{\Theta}$ of a parameter Θ relative to the maximum likelihood estimate Θ^* is computed as

$$\mathrm{ARE}(\hat{\Theta}) = \frac{\sigma_{\Theta^*}^2}{\sigma_{\hat{\Theta}}^2}$$

where $\sigma_{\Theta^*}^2$ is the Cramér-Rao lower bound and $\sigma_{\hat{\Theta}}^2/m$ is the variance in the limiting distribution of $\hat{\Theta}$. The joint asymptotic relative efficiency is given by

$$\mathrm{JARE}(\hat{\Theta}) = \frac{\det(\Sigma')}{\det(\Sigma)}$$

where we denote the Cramér-Rao lower bound for the estimator of Θ by Σ' where the inverse of the Cramér-Rao lower bound is given by

$$\Sigma\Theta^* = \left\{E\left[\frac{\partial^2 \log P_y}{\partial\Theta_i\,\partial\Theta_j}\right]\right\}$$

With these definitions of Σ and $\Sigma_\Theta*$ we can compare the performance of the moment estimators with the maximum likelihood estimators as n becomes large.

CHAPTER III

3.1 Definition and Notation

We will now consider a mixed distribution function consisting of a mixing of the Binomial and the Poisson distributions. The probability density function of the Binomial Poisson mixture is

$$P_y = P(y) = \alpha P_1(y) + (1 - \alpha)P_2(y) \quad \text{for } y = 0,1,\ldots$$

where

$$P_1(y) = \frac{e^{-\lambda}\,\lambda^y}{y!} \quad y = 0,1,\ldots$$

and

$$P_2(y) = \binom{n}{y}\, p^y\, (1-p)^{n-y} \quad y = 0,1,\ldots,n$$

The parameters of the mixture to be estimated are the mixing proportion α, the Poisson parameter λ and the Binomial parameter p, while n is assumed to be known. The mixture will be identifiable for a sample size greater than or equal to three for the case when no parameters are known. When one parameter is known, identifiability requires a sample size of two or more. To estimate the parameters, we will use the method of moments. For the three parameter case, the maximum likelihood estimators for a sample of size three can be obtained by solving the first three factorial moments for α, λ, p. The sample factorial moments shall be defined, as Cohen (1963) defined them.

The k^{th} factorial moment of a probability distribution defined in the discrete case as

$$E[y^{[k]}] = {}_{y}\; y^{[k]}\; P(y)$$

where the symbol $y^{[k]}$ denotes the factorial

$$y^{[k]} = y(y-1)\dots(y-k+1) \quad k = 1,2,\dots$$

is obtained from the factorial moment generating function by differentiating it k times with respect to t and then evaluating the result when $t = 1$.

Thus for

$$P(y) = \alpha P_1(y) + (1-\alpha)P_2(y) \quad y = 0,1,\dots$$

we have

$$E[t^y] = \sum_y t^y P(y) = \alpha e^{-\lambda(1-t)} + (1-\alpha)[1-p+pt]^n$$

Now

$$\left.\frac{\partial E[t^y]}{\partial t}\right|_{t=1} = [\alpha\lambda e^{-\lambda(1-t)} + (1-\alpha)np[1-p+pt]^{n-1}]_{t=1}$$

and therefore

$$f_1 = \alpha\lambda + (1-\alpha)np.$$

In general

$$f_k = \alpha\lambda^k + (1-\alpha)n(n-1)\dots(n-k+1)p^k$$

Also,

$$E[y^{[k]}] = \sum_y^{\infty} y^{[k]} P(y)$$

and it can be shown that

$$E[y] = f_1$$

Similarly

$$E[y^{[2]}] = f_2$$

and

$$E[y^2] = E[y^{[2]}] + E[y] = f_1 + f_2$$

It follows likewise that

$$E[y^3] = f_3 + 3f_2 + f_1$$

$$E[y^4] = f_4 + 6f_3 + 7f_2 + f_1$$

$$E[y^5] = f_5 + 10f_4 + 25f_3 + 15f_2 + f_1$$

$$E[y^6] = f_6 + 15f_5 + 65f_4 + 90f_3 + 31f_2 + f_1$$

These expected values are to be used in the next section to find the moments of the sample factorial moments.

3.2 Sample Factorial Moments

In this section we derive the moments of the sample factorial moments as these will be used later on for determining the variances of the moment estimators. The k^{th} sample factorial moment is defined by Cohen (1963) as:

$$F_k = \sum_{y=o}^{R} y(y-1)\dots(y-k+1)\,\frac{m_y}{m}$$

where m_y is the sample frequency of y and

$$\sum_{y=o}^{R} m_y = m.$$

Before we derive the first two moments of $F_k(k = 1,2,3)$, let us consider the sample cell frequency m_y and its moments.

$$E[m_y] = mP(y)$$

$$Var(m_y) = mP(y)(1-P(y))$$

and

$$Cov(m_y, m_x) = -mP(y)P(x).$$

These results are obtained directly by recalling that m_y is a multinomial random variable. Hence

$$E[F_k] = E\{\sum_{y=o}^{R} y(y-1)\ldots(y-k+1)\frac{m_y}{m}\}$$

$$= \sum_{y=o}^{R} y(y-1)\ldots(y-k+1)P(y)$$

and therefore

$$E[F_k] = f_k \qquad k = 1,2,\ldots$$

Since the variances and covariances of the F_k can not be obtained as a simple function of the population values, we will present details of the derivation for the cases $k = 1,2,3$.

$$\mathrm{Var}(F_1) = \frac{1}{m^2}\{\Sigma y^2 \mathrm{var}(m_y) + \sum_{y} \neq \sum_{x} y\, x \,\mathrm{cov}(m_y, m_x)\}$$

and from the expressions for the variance and covariance of sample cell frequencies, we have

$$\mathrm{Var}(F_1) = \frac{1}{m}\{\Sigma y^2 P_{(y)}(1 - P_{(y)}) - \sum_{y} \neq \sum_{x} y\, x\, P(y)P(x)\}$$

A simple algebraic manipulation yields the final form as

$$\mathrm{Var}(F_1) = \frac{1}{m}\{\Sigma y^2 P(y) - (\Sigma y P(y)^2\}$$

$$= \frac{1}{m}\{f_2 + f_1 - f_1^{\,2}\}$$

Similarly it follows that

$$\mathrm{Var}(F_2) = \frac{1}{m}\{2f_2 + 4f_3 + f_4 - f_2^{\,2}\}$$

$$\mathrm{Var}(F_3) = \frac{1}{m}\{f_6 + 9f_5 + 18f_4 + 6f_3 - f_3^{\,2}\}$$

Now

$$\mathrm{Cov}(F_1,F_2) = \frac{1}{m^2}\{\Sigma y^2(y-1)\,\mathrm{var}(m_y) + \sum_{y\neq x}\sum y\, x(x-1)\,\mathrm{cov}(m_y, m_x)\}$$

from which it follows

$$\mathrm{Cov}(F_1,F_2) = \frac{1}{m}\{\Sigma y^2(y-1)\, P_{(y)}(1-P_{(y)}) - \sum_{y\neq x}\sum y\, x(x-1)\, P_{(y)}\, P_{(x)}\}$$

and finally

$$\mathrm{Cov}(F_1,F_2) = \frac{1}{m}\ \{ f_3 + 2f_2 - f_1f_2\}$$

Also,

$$\mathrm{Cov}(F_1,F_3) = \frac{1}{m}\ \{ f_4 + 3f_3 - f_1f_3\}$$

$$\mathrm{Cov}(F_2,F_3 = \frac{1}{m}\ \{ f_5 + 6f_4 + 6f_3 - f_2f_3\}$$

3.3 Moment Estimators in the Two Parameter Cases

There are three cases to be considered in this section. These cases in the order that we shall discuss them are: (1) when the mixing proportion α is known, the parameters to be estimated are the Poisson parameter λ and the Binomial parameter p, (2) when λ is known, the parameters to be estimated are α and p, and (3) when p is known, the parameters α and λ will be estimated. Identifiability follows in these cases for a sample size greater than or equal to two.

3.3.1 Mixing Proportion α known: Estimators and Variances

The maximum likelihood estimators for a sample of size two will be obtained by solving the first two population factorial moments for λ, p.

The first two population factorial moments are

$$f_1 = \alpha\lambda + (1-\alpha)np \qquad \text{—— (1)}$$

$$f_2 = \alpha\lambda^2 + (1-\alpha)n(n-1)p^2 \qquad \text{—— (2)}$$

From (1) $\lambda = \dfrac{f_1 - (1-\alpha)np}{\alpha}$

From (2) $\lambda^2 = \dfrac{f_1 -(1-\alpha)n(n-1)p^2}{\alpha}$

Therefore

$$(f_1 - (1-\alpha)np)^2 = \alpha\ (f_2 - (1-\alpha)n(n-1)p^2)$$

and it follows that

$$p^2 n(n-\alpha)(1-\alpha) - 2pf_1 n(1-\alpha) + f_1^2 - \alpha f_2 = 0$$

Solving this equation for p, we get the moment estimator

$$\hat{p} = \frac{F_1}{n-\alpha} \pm \frac{1}{n-\alpha} \left\{\frac{\alpha}{n(1-\alpha)} [F_2(n-\alpha) - (n-1)\ F_1^2]\right\}^{\frac{1}{2}} \qquad \text{--- (3)}$$

Now from (1)

$$p = \frac{f_1 - \alpha\lambda}{(1-\alpha)n}$$

Substituting p into equation (2)

$$f_2 = \alpha\lambda^2 + (n-1)\ \frac{(f_1 - \alpha\lambda)^2}{(1-\alpha)n}$$

we have

$$\lambda^2 n(n-\alpha) - 2\lambda f_1 \alpha(n-1) + (n-1)f_1^2 - n(1-\alpha)\ f_2 = 0.$$

Solving this equation for λ, we get the moment estimator

$$\hat{\lambda} = \frac{F_1(n-1)}{(n-\alpha)} \pm \frac{1}{n-\alpha} \left\{\frac{n(1-\alpha)}{\alpha} [F_2(n-\alpha) - (n-1)F_1^2]\right\}^{\frac{1}{2}} \qquad \text{--- (4)}$$

Our problem now is to determine unique estimators $\hat{p}, \hat{\lambda}$ from the above equations. The criterion to be used will be that of Fisher-Consistency which is defined as follows (Rao(1952)):

Consider samples from a finite multinomial distribution with cell probabilities $\pi_1(\underline{\theta}), \ldots, \pi_k(\underline{\theta})$ depending on a vector parameter $\underline{\theta}$; n is the total sample size, n_i is the observed frequency in the i^{th} cell, $p_i = n_i/n$ is the observed proportion in the i^{th} cell and $g(\underline{\theta})$ is the parametric function to be estimated.

An estimator T is said to be Fisher-Consistent (F-C) for $g(\underline{\Theta})$ if and only if i) T is a continuous function over the set of vectors $(y_1,\ldots,y_k)$ such that $y_i > 0$ and $y_1 + \ldots + y_k = 1$ with the value of T at $y_i = p_i$ $i = 1,\ldots, k$ as the estimate of $g(\underline{\Theta})$ based on the sample, ii) the value of T at $y_i = \pi_i(\underline{\Theta})$ $i = 1,\ldots, k$ is $g(\underline{\Theta})$ for all admissible values of Θ, that is

$$T[\pi_1(\underline{\Theta}),\ldots, \pi_k(\underline{\Theta})] \equiv g(\underline{\Theta}).$$

In effect the definition demands that the estimator should be an explicit function of the observed proportions only, which may be written as $T(p_1,\ldots,p_k)$, and that it should have the true value of $g(\underline{\Theta})$ when the observed proportions happen to coincide with the true proportions.

Expanding F_1 and F_2 in terms of their population parameter values:

$$F_1 = \alpha\lambda + (1-\alpha)\, np$$

$$F_2 = \alpha\lambda^2 + (1-\alpha)\, n(n-1)p^2$$

Substituting for F_1 and F_2 in equations (3) and (4) the results are

$$\hat{p}^* = \frac{1}{n(n-\alpha)} [\{ n\alpha\lambda + n^2 (1-\alpha)p\} \pm \{ n\alpha\lambda - n(n-1)\alpha p\}] \quad \text{—— (5)}$$

and

$$\hat{\lambda}^* = \frac{1}{n-\alpha} [(n-1) \{\alpha\lambda + (1-\alpha)\, np\} \pm n(1-\alpha) \{\lambda - (n-1)p \}] \quad \text{—— (6)}$$

Considering in turn the positive and negative signs, it is easily seen that the F-C moment estimators are given by

$$\hat{p} = \frac{F_1}{n-\alpha} - \frac{1}{n(n-\alpha)} \{\frac{\alpha n}{1-\alpha} [F_2(n-\alpha)-(n-1)F_1^2]\}^{\frac{1}{2}} \quad \text{—— (7)}$$

and

$$\hat{\lambda} = \frac{(n-1)F_1}{n-\alpha} + \frac{1}{n-\alpha}\left\{\frac{(1-\alpha)}{\alpha} n\,[F_2(n-\alpha) - (n-1)F_1^2]\right\}^{\frac{1}{2}} \qquad \text{---}(8)$$

Now we shall determine expressions for the variances of the moment estimators $\hat{p}$, $\hat{\lambda}$. First set

$$\hat{\lambda} = H_1(F_1,F_2)$$

$$\hat{p} = H_2(F_1,F_2)$$

By expanding in a Taylor's series at the parametric factorial moments, ignoring terms of order higher than two, we have

$$\hat{\lambda} = H_1(F_1,F_2) = H_1(f_1,f_2) + [(F_1-f_1)\frac{\partial H_1}{\partial F_1} + (F_2 - f_2)\frac{\partial H_1}{\partial F_2}]$$

$$\hat{p} = H_2(F_1,F_2) = H_2(f_1,f_2) + [(F_1-f_1)\frac{\partial H_2}{\partial F_1} + (F_2 - f_2)\frac{\partial H_2}{\partial F_2}]$$

Now

$$H_1(F_1,F_2) - H_1(f_1,f_2) = (F_1-f_1)\,\delta_{11} + (F_2-f_2)\,\delta_{12}$$

$$H_2(F_1,F_2) - H_2(f_1,f_2) = (F_1-f_1)\,\delta_{12} + (F_2-f_2)\,\delta_{22}$$

and it readily follows that

$$Var(H_1) = Var(F_1)\,\delta_{11}^2 + Var\,(F_2)\,\delta_{12}^2 + 2\,\delta_{11}\,\delta_{12}\,Cov(F_1,F_2)$$

$$var(H_2) = Var(F_1)\,\delta_{21}^2 + Var\,(F_2)\,\delta_{22}^2 + 2\,\delta_{21}\,\delta_{22}\,Cov(F_1,F_2)$$

and

$$Cov(H_1,H_2) = Var(F_1)\,\delta_{11}\,\delta_{21} + Var(F_2)\,\delta_{12}\,\delta_{22} + Cov\,(F_1,F_2)$$
$$[\delta_{11}\,\delta_{22} + \delta_{21}\,\delta_{12}]$$

We now are required to find the matrix

$$\{\delta_{ij}\} = \{\frac{\partial H_i}{\partial F_j}\Big|_f\}$$

It follows as

$$\frac{\partial H_1}{\partial F_1}\Big|_f = \frac{\partial\hat{\lambda}}{\partial F_1}\Big|_f = \frac{\partial}{\partial F_1}\left[(\frac{n-1}{n-\alpha})\,F_1 + \frac{1}{n-\alpha}\left\{\frac{(1-\alpha)n}{\alpha}[F_2(n-\alpha)-(n-1)F_1^2]\right\}^{\frac{1}{2}}\right]_f$$

$$= \frac{n-1}{n-\alpha} + \frac{1}{n-\alpha} \{n(1-\alpha)(\lambda - (n-1)p)\}^{-1} \frac{n(1-\alpha)}{\alpha} (-(n-1)F_1)$$

$$= \frac{-(n-1)p}{\alpha(\lambda - (n-1)p)}$$

$$\left.\frac{\partial H_1}{\partial F_2}\right|_f = \frac{1}{n-\alpha} [\tfrac{1}{2} \{n^2(1-\alpha)^2(\lambda - (n-1)p)\}^{-1} \frac{n(1-\alpha)}{\alpha}(n-\alpha)]$$

$$= \frac{1}{2\,\alpha(\lambda - (n-1)p)}$$

$$\left.\frac{\partial H_2}{\partial F_1}\right|_f = \left.\frac{\partial \hat{p}}{\partial F_1}\right|_f = \frac{\partial}{\partial F_1} \left[\frac{F_1}{n-\alpha} - \frac{1}{n(n-\alpha)} \left\{ \frac{n\alpha}{1-\alpha} [F_2(n-\alpha) - (n-1)F_1^2]\right\}^{\frac{1}{2}}\right]_f$$

$$= \frac{1}{n-\alpha} \left[\frac{(1-\alpha)(\lambda - (n-1)p) + (n-1)(\alpha\lambda + (1-\alpha)np)}{n(1-\alpha)\ (\lambda - (n-1)p)}\right]$$

$$= \frac{\lambda}{n(1-\alpha)\ (\lambda - (n-1)p)}$$

$$\left.\frac{\partial H_2}{\partial F_2}\right|_f = -\tfrac{1}{2}\{n\,\alpha\,(\lambda - (n-1)p)\}^{-1} \frac{\alpha}{1-\alpha}$$

$$= -\frac{1}{2n(1-\alpha)\ (\lambda - (n-1)p)}$$

Thus

$$\{\delta_{ij}\} = \left\{ \left.\frac{\partial H_i}{\partial F_j}\right|_f \right\} = \frac{1}{\lambda-(n-1)p} \begin{bmatrix} \frac{-(n-1)p}{\alpha} & \frac{1}{2\alpha} \\ \frac{\lambda}{n(1-\alpha)} & \frac{-1}{2(1-\alpha)n} \end{bmatrix}$$

The variance matrix of H_1, H_2 can now be written as

$$\Sigma_H = \{\delta_{ij}\} \quad \{\mu(F)\} \{\delta_{ij}\}'$$

of which the general term is expressed as

$$\sigma_{ii'} = \sum_{j=1}^{2} \mu_2(F_j)\, \delta_{ij}\, \delta_{i'j} + \sum_{\substack{j,j'=1 \\ j \neq j'}}^{2} \delta_{ij}\, \delta_{i'j'} \quad \mu_{11}(F_j, F_{j'})$$

It follows then

$$\sigma_{\hat{\lambda}}^2 = \mu_2(F_1)\, \delta_{11}^2 + \mu_2(F_2)\, \delta_{12}^2 + 2\mu_{11}(F_1,F_2)\, \delta_{11}\, \delta_{12}$$

$$\sigma_{\hat{p}}^2 = \mu_2(F_1)\, \delta_{21}^2 + \mu_2(F_2)\, \delta_{22}^2 + 2\mu_{11}(F_1,F_2)\, \delta_{21}\, \delta_{22}$$

$$\sigma_{\hat{\lambda},\hat{p}} = \mu_2(F_1)\, \delta_{11}\delta_{21} + \mu_2(F_2)\, \delta_{12}\, \delta_{22} + \mu_{11}(F_1,F_2)[\delta_{11}\, \delta_{22} + \delta_{21}\, \delta_{22}]$$

where

$$\mu_2(F_i) \equiv \text{Var}(F_i)$$

$$\mu_{11}(F_j, F_{j'}) \equiv \text{Cov}(F_j, F_{j'})$$

3.3.2 Poisson Parameter λ known: Estimators and Variances

Here we are looking for moment estimators for p and α as functions of the first two factorial moments.

Now

$$f_1 = \alpha\lambda + (1-\alpha)np \qquad \text{---(1)}$$

$$f_2 = \alpha\lambda^2 + (1-\alpha)n(n-1)p^2 \qquad \text{---(2)}$$

From (1)

$$\alpha = \frac{f_1-np}{\lambda-np} \qquad \text{---(3)}$$

$$1-\alpha = \frac{\lambda-f_1}{\lambda-np}$$

Substituting for α in equation (2) we have:

$$f_2 = \{\frac{f_1-np}{\lambda-np}\}\lambda^2 + \{\frac{\lambda-f_1}{\lambda-np}\} n(n-1)p^2$$

Thus

$$f_2(\lambda - np) = \lambda^2(f_1-np) + n(n-1)(\lambda - f_1)p^2$$

and

$$p^2n(n-1)(\lambda - f_1) - np(\lambda^2 - f_2) + (\lambda^2 f_1 - f_2\lambda) = 0.$$

Solving this equation for p, we get the moment estimator

$$\hat{p} = \frac{(\lambda^2 - F_2) \pm B}{2(n-1)(\lambda - F_1)} \qquad \text{---(4)}$$

From equation (3)

$$\hat{\alpha} = \frac{F_1 - n\hat{p}}{\lambda - n\hat{p}}$$

That is

$$\hat{\alpha} = \frac{2F_1(1-\frac{1}{n})(\lambda - F_1) - (\lambda^2 - F_2) \pm B}{2(1-\frac{1}{n})\lambda(\lambda - F_1) - (\lambda^2 - F_2) \pm B} \quad \text{---(5)}$$

where $B = \{(\lambda^2 - F_2)^2 - 4(\lambda^2 F_1 - F_2\lambda)(1-\frac{1}{n})(\lambda - F_1)\}^{\frac{1}{2}}$

Our problem as before is to determine those estimators which will be Fisher-Consistent. To find the F-C moment estimators for p and α from equation (4) and (5) expand F_1 and F_2 in terms of the population parameters

$$F_1 = \alpha\lambda + (1-\alpha)np$$

$$F_2 = \alpha\lambda^2 + (1-\alpha)n(n-1)p^2$$

Substituting for F_1 and F_2 in equations (4) and (5) the results are

$$\hat{p}^* = \frac{(1-\alpha)[\lambda^2 - (1-\frac{1}{n})u^2] \pm (1-\alpha)[(1-\frac{1}{n})u^2 - 2(1-\frac{1}{n})u\lambda + \lambda^2]}{2(n-1)(1-\alpha)(\lambda - u)} \quad \text{--- (6)}$$

$$\hat{\alpha}^* = \frac{\alpha\lambda + (1-\alpha)np - n\hat{p}^*}{\lambda - n\hat{p}^*} \quad \text{--- (7)}$$

Considering in turn positive and negative signs, it is easily seen that the consistent moment estimators are given by

$$\hat{p} = \frac{(\lambda^2 - F_2) - B}{2(n-1)(\lambda - F_1)} \quad \text{---(8)}$$

$$\hat{\alpha} = \frac{2(1-\frac{1}{n})F_1(\lambda - F_1) - (\lambda^2 - F_2) + B}{2(1-\frac{1}{n})\lambda(\lambda - F_1) - (\lambda^2 - F_2) + B} \quad \text{---(9)}$$

Now we shall determine expressions for the variances of the moment estimators $\hat{p}$, $\hat{\alpha}$. First set

$$\hat{p} = H_1(F_1, F_2)$$

$$\hat{\alpha} = H_2(F_1, F_2)$$

and expanding in a Taylor's series at the parametric factorial moments as in the case of the mixing proportion being known (3.3.1) the expressions for the variances of H_1 and H_2 and the covariance are determined:

$$\mathrm{Var}(H_1) = \mathrm{Var}(F_1)\,\delta_{11}^2 + \mathrm{Var}(F_2)\,\delta_{12}^2 + 2\,\delta_{11}\,\delta_{12}\,\mathrm{Cov}(F_1,F_2)$$

$$\mathrm{Var}(H_2) = \mathrm{Var}(F_1)\,\delta_{21}^2 + \mathrm{Var}(F_2)\,\delta_{22}^2 + 2\,\delta_{21}\,\delta_{22}\,\mathrm{Cov}(F_1,F_2)$$

and

$$\mathrm{Cov}(H_1,H_2) = \mathrm{Var}(F_1)\,\delta_{11}\,\delta_{21} + \mathrm{Var}(F_2)\,\delta_{12}\,\delta_{22} + \mathrm{Cov}(F_1,F_2)\,[\delta_{11}\,\delta_{22} + \delta_{21}\,\delta_{12}]$$

We now are required to find the matrix

$$\{\delta_{ij}\} = \{ \left.\frac{\partial H_i}{\partial F_j}\right|_f \}$$

It follows as

$$\left.\frac{\partial H_i}{\partial F_i}\right|_f = \left.\frac{\partial \hat{p}}{\partial F_1}\right|_f = \frac{\partial}{\partial F_1}\left[\frac{\lambda^2 - F_2 - B}{2(n-1)(\lambda - F_1)}\right]_f$$

$$= \frac{1}{n}\left[\frac{\lambda^3 - 2\lambda^2 u + \lambda(1-\frac{1}{n})u^2 + u^3(1-\frac{1}{n}) - 2(1-\frac{1}{n})u^2\lambda + u\lambda^2}{(1-\alpha)\,(\lambda-u)(u^2(1-\frac{1}{n}) - 2(1-\frac{1}{n})u\lambda + \lambda^2)}\right]$$

where $u = np$

$$= \frac{\lambda^2 - (1-\frac{1}{n})u^2}{n(1-\alpha)(u^2(1-\frac{1}{n}) - 2(1-\frac{1}{n})u\lambda + \lambda^2)}$$

$$\left.\frac{\partial H_1}{F_2}\right|_f = \frac{-u^2 + u\lambda - (\lambda-u)\lambda}{n(1-\alpha)(u^2(1-\frac{1}{n}) - 2(1-\frac{1}{n})u\lambda + \lambda^2)(\lambda-u)}$$

$$= \frac{u - 2\lambda}{u(1-\alpha)(u^2(1-\frac{1}{n}) - 2(1-\frac{1}{n})u\lambda + \lambda^2)}$$

$$\left.\frac{\partial H_2}{\partial F_1}\right|_f = \left.\frac{\partial \hat{\alpha}}{\partial F_1}\right|_f = \frac{\partial}{\partial F_1}\left[\frac{2(1-\frac{1}{n})F_1(\lambda - F_1) - (\lambda^2 - F_2) + B}{2(1-\frac{1}{n})\lambda(\lambda - F_1) - (\lambda^2 - F_2) + B}\right]_f$$

$$= \frac{-2\,(1-\frac{1}{n})\,u}{u^2(1-\frac{1}{n}) - 2\lambda\, u(1-\frac{1}{n}) + \lambda^2}$$

$$\left.\frac{\partial H_2}{\partial F_2}\right|_f = \frac{u^2 - 2\lambda u + \lambda^2}{(\lambda - u)^2(u^2(1-\frac{1}{n}) - 2(1-\frac{1}{n})\lambda u + \lambda^2)}$$

$$= \frac{1}{u^2(1-\frac{1}{n}) - 2\lambda u(1-\frac{1}{n}) + \lambda^2}$$

Thus

$$\{\delta_{ij}\} = \{\left.\frac{\partial H_i}{\partial F_j}\right|_f\} = \frac{1}{u^2(1-\frac{1}{n}) - 2\lambda u(1-\frac{1}{n}) + \lambda^2}\begin{bmatrix} \frac{\lambda^2-(1-\frac{1}{n})u^2}{n(1-\alpha)} & \frac{u-2\lambda}{n(1-\alpha)} \\ -2(1-\frac{1}{n})u & 1 \end{bmatrix}$$

The variance matrix of H_1, H_2 can now be written as

$$\Sigma_H = \{\delta_{ij}\} \quad \{u(F)\} \quad \{\delta_{ij}\}'$$

from which it follows that

$$\sigma_{\hat{p}}^2 = u_2(F_1)\,\delta_{11}^2 + u_2(F_2)\,\delta_{12}^2 + 2u_{11}(F_1,F_2)\,\delta_{11}\,\delta_{12}$$

$$\sigma_{\hat{\alpha}}^2 = u_2(F_1)\,\delta_{21}^2 + u_2(F_2)\,\delta_{22}^2 + 2u_{11}(F_1,F_2)\,\delta_{21}\,\delta_{22}$$

$$\sigma_{\hat{p},\hat{\alpha}} = u_2(F_1)\,\delta_{11}\,\delta_{21} + u_2(F_2)\delta_{12}\delta_{22} + u_{11}(F_1,F_2)(\delta_{11}\delta_{22} + \delta_{21}\delta_{12})$$

where

$$u_2(F_i) = \mathrm{Var}(F_j)$$

$$u_{11}(F_j,F_{j'}) = \mathrm{Cov}(F_j,F_{j'})$$

3.3.3 Binomial Parameter p known: Estimators and Variances

We will solve for the moment estimators for λ and α as functions of the first two factorial moments.

Now

$$f_1 = \alpha\lambda + (1-\alpha)np \qquad \text{—(1)}$$

$$f_2 = \alpha\lambda^2 + (1-\alpha)n(n-1)p^2 \qquad \text{—(2)}$$

Setting

$$np = a$$

$$n(n-1)p^2 = b$$

and using equation (1) we have

$$\alpha = \frac{f_1 - a}{\lambda - a} \qquad \text{---(3)}$$

$$1 - \alpha = \frac{\lambda - f_1}{\lambda - a}$$

Substituting for α in equation (2) we have

$$f_2 = \{\frac{f_1 - a}{\lambda - a}\} \lambda^2 + \{\frac{\lambda - f_1}{\lambda - a}\} b$$

Now

$$f_2(\lambda - a) = \lambda^2(f_1 - a) + (\lambda - f_1)b$$

and

$$\lambda^2(f_1 - a) - \lambda(f_2 - b) + (af_2 - bf_1) = 0$$

Solving this equation for λ, we get the moment estimator

$$\hat{\lambda} = \frac{(F_2 - b) \pm C}{2(F_1 - a)} \qquad \text{---(4)}$$

From equation (3)

$$\hat{\alpha} = \frac{F_1 - a}{\lambda - a}$$

That is

$$\hat{\alpha} = \frac{2(F_1 - a)^2}{F_2 - b - 2a\ (F_1 - a) \pm C} \qquad \text{---(5)}$$

where

$$C = \{(F_2 - b)^2 - 4(F_1 - a)(aF_2 - bF_1)\}^{\frac{1}{2}}$$

Our problem once again is to determine those estimators which will be Fisher-consistent. To find the F-C moment estimators for λ and α from equations (4) and (5), expand F_1 and F_2 in terms of the population parameters

$$F_1 = \alpha\lambda + (1-\alpha)\ np$$
$$F_2 = \alpha\lambda^2 + (1-\alpha)n(n-1)p^2$$

Set

$$a = np$$
$$b = n(n-1)p^2$$

Substituting for F_1 and F_2 in equations (4) and (5) the results are

$$\hat{\lambda}* = \frac{\alpha\ (\lambda^2-b) \pm \alpha\ \{\ (\lambda^2-b)^2-4\lambda\ (\lambda\ -a)(a\lambda\ \ -b)\}}{2\alpha\ (\lambda\ -\ a)} \qquad \text{---(6)}$$

$$\hat{\alpha}* = \frac{2\ \alpha^2(\lambda\ -a)^2}{\alpha(\lambda^2\ -b)\ -\ 2a\alpha\ (\lambda\ -a) \pm \alpha(\lambda^2-\ 2a\lambda + b)} \qquad \text{---(7)}$$

Considering in turn positive and negative signs, it is easily seen that the consistent moment estimators are given by

$$\hat{\lambda} = \frac{F_2 - b + C}{2(F_1 - a)} \qquad \text{---(8)}$$

$$\hat{\alpha} = \frac{2(F_1 - a)^2}{F_2 - b-2a(F_1-a)+C} \qquad \text{---(9)}$$

We will now determine the expressions for the variances of the moment estimators $\hat{\lambda}$, $\hat{\alpha}$. First set

$$\hat{\lambda} = H_1(F_1,F_2)$$

$$\hat{\alpha} = H_2(F_1,F_2)$$

and expanding in a Taylor's series at the parametric factorial moments as in the case of the mixing proportion being known (3.3.1) the expressions for the variances of H_1 and H_2 and the covariance are determined:

$$Var(H_1) = Var(F_1)\,\delta_{11}^2 + Var(F_2)\,\delta_{12}^2 + 2\,\delta_{11}\,\delta_{12}\,Cov(F_1,F_2)$$

$$Var(H_2) = Var(F_1)\,\delta_{21}^2 + Var(F_2)\,\delta_{22}^2 + 2\,\delta_{21}\,\delta_{22}\,Cov(F_1,F_2)$$

and

$$Cov(H_1,H_2) = Var(F_1)\,\delta_{11}\,\delta_{21}\,Var(F_2)\,\delta_{12}\,\delta_{22} + Cov(F_1,F_2)$$
$$[\delta_{11}\,\delta_{22} + \delta_{21}\,\delta_{12}]$$

We are now required to find the matrix

$$\{\delta_{ij}\} = \{ \left.\frac{\partial H_i}{\partial F_j}\right|_f \}$$

It follows as

$$\left.\frac{\partial H_1}{\partial F_1}\right|_f = \left.\frac{\partial\hat{\lambda}}{\partial F_1}\right|_f = \frac{\partial}{\partial F_1}\left[\frac{F_2 - b + C}{2(F_1-a)}\right]_f$$

$$= \frac{-\lambda^3 + a\lambda^2 + b\lambda + ab}{\alpha(\lambda^2-2a\lambda+b)(\lambda-a)}$$

$$= \frac{(b-\lambda^2)}{\alpha(\lambda^2-2a\lambda+b)}$$

$$\left.\frac{\partial H_1}{\partial F_2}\right|_f = \frac{\lambda^2 - 2a\lambda + \lambda^2 + 2a^2 - 2a\lambda}{2\alpha(\lambda-a)(\lambda^2-2a\lambda+b)}$$

$$= \frac{\lambda - a}{\alpha(\lambda^2-2a\lambda+b)}$$

$$\left.\frac{\partial H_2}{\partial F_1}\right|_f = \left.\frac{\partial\hat{\alpha}}{\partial F_1}\right|_f = \frac{\partial}{\partial F_1}\left[\frac{2(F_1-a)^2}{F_2-b-2a(F_1-a)+C}\right]_f$$

$$= \frac{2(\lambda^3-2a\lambda^2+a^2\lambda)}{(\lambda-a)^2}$$

$$= 2\lambda$$

$$\left.\frac{\partial H_2}{\partial F_2}\right|_f = -\left[\frac{\alpha(\lambda^2-2a\lambda+b)+\alpha(\lambda^2+2a^2-2a\lambda-b)}{2\alpha^2(\lambda-a)^2(\lambda^2-2a\lambda+b)}\right]$$

$$= \frac{-1}{\alpha(\lambda^2-2a\lambda+b)}$$

Thus

$$\{\delta_{ij}\} = \{\frac{\partial H_i}{\partial F_j}\Big|_f\} = \frac{1}{\lambda^2-2a\lambda+b} \begin{bmatrix} \frac{b-\lambda^2}{\alpha} & \frac{\lambda-a}{\alpha} \\ 2\lambda\ (\lambda^2-2a\lambda+b) & -\frac{1}{\alpha} \end{bmatrix}$$

The variance matrix of H_1, H_2 can now be written as

$$\Sigma_H = \{\delta_{ij}\}\ \ \{u(F)\}\ \{\delta_{ij}\}'$$

from which it follows that

$$\sigma^2_{\hat{\lambda}} = u_2(F_1)\delta_{11}^2 + u_2(F_2)\ \delta_{12}^2 + 2u_{11}(F_1,F_2)\,\delta_{11}\delta_{12}$$

$$\sigma^2_{\hat{\alpha}} = u_2(F_1)\ \delta_{21}^2 + u_2(F_2)\ \delta_{22}^2 + 2u_{11}(F_1,F_2)\ \delta_{21}\ \delta_{22}$$

$$\sigma_{\hat{\lambda}\hat{\alpha}} = u_2(F_1)\ \delta_{11}\ \delta_{22} + u_2(F_2)\delta_{12}\ \delta_{22} + u_{11}(F_1,F_2)(\delta_{11}\ \delta_{22} + \delta_{21}\ \delta_{12})$$

where

$$u_2(F_i) = \mathrm{Var}(F_i)$$

$$u_{11}(F_j,F_{j'}) = \mathrm{Cov}(F_j,F_{j'})$$

3.4 Asymptotic Relative Efficiency

In order to compare the efficiencies of estimators, we must first derive the elements of the Cramér-Rao lower bound matrix. We will denote the inverse of the Cramér-Rao lower bound matrix by

$$\Sigma_\Theta' = \{\sigma'_{ij}\}$$

where

$$\Theta' = (\lambda',\ p',\ \alpha')$$

In view of symmetry, only the elements of the upper right triangle of the matrix need be determined. They are

σ_{11}', σ_{12}', σ_{22}', σ_{13}', σ_{23}', σ_{33}'. Now

$$\sigma_{11}' = E\left[\frac{\partial \log P(y)}{\partial\lambda}\right]^2$$

That is

$$\sigma_{11}' = \sum_{y=o}^{\infty} [\frac{\partial P(y)}{\partial \lambda} \frac{1}{P_y}]^2 P(y)$$

and it follows that

$$\sigma_{11}' = \alpha^2 \sum_{y=o}^{\infty} [\frac{C^{-\lambda} \lambda^{y-1}}{y!} (y-\lambda)]^2 \frac{1}{P(y)}$$

Simplifying, we get

$$\sigma_{11}' = \frac{\alpha^2}{\lambda^2} \sum_{y=o}^{\infty} \frac{[P_1(y)(y-\lambda)]^2}{P(y)}$$

Similarly, we have

$$\sigma_{12}' = \frac{\alpha(1-\alpha)}{P(1-P)} \sum_{y=o}^{n} \frac{[P_1(y)(y-\lambda)P_2(y)(y-np)]}{P(y)}$$

$$\sigma_{22}' = \frac{(1-\alpha)^2}{P^2(1-P)^2} \sum_{y=o}^{n} \frac{[P_2(y)(y-np)]^2}{P(y)}$$

$$\sigma_{13}' = \frac{\alpha}{\lambda} \sum_{y=o}^{\infty} \frac{[P_1(y)(y-\lambda)(P_1(y) - P_2(y))]}{P(y)}$$

$$\sigma_{23}' = \frac{1-\alpha}{P(1-P)} \sum_{y=o}^{n} \frac{[P_2(y)(y-np)(P_1(y)- P_2(y))]}{P(y)}$$

$$\sigma_{33}' = \sum_{y=o}^{\infty} \frac{[P_1(y) - P_2(y)]^2}{P(y)}$$

It is to be noted that σ_{11}', σ_{13}', σ_{33}' are summed to infinity, while σ_{12}', σ_{22}', σ_{23}' are summed to n, since the contribution by the Binomial becomes zero for y greater than n.

Letting $\Sigma\Theta^*$ denote $\Sigma\Theta'^{-1}$, the Cramér-Rao lower bound matrix, the elements of the upper right triangle will be σ_{11}^*, σ_{12}^*, σ_{13}^*, σ_{22}^*, σ_{23}^*, σ_{33}^*.

3.4.1. Asymptotic Relative Efficiency for α known

$$ARE(\hat{\lambda}) = \frac{\sigma_{11}^{*}}{\sigma_{\hat{\lambda}}^{2}}$$

$$ARE(\hat{p}) = \frac{\sigma_{22}^{*}}{\sigma_{\hat{p}}^{2}}$$

$$JARE(\hat{\lambda},\hat{p}) = \frac{\det|\Sigma\Theta^{*}|}{\det|\Sigma\hat{\Theta}|}$$

where

$$\Sigma\Theta^{*} = \begin{bmatrix} \sigma_{11}^{*} & \sigma_{12}^{*} \\ \cdots & \sigma_{22}^{*} \end{bmatrix}$$

and

$$\Sigma\hat{\Theta} = \begin{bmatrix} \sigma_{\hat{\lambda}}^{2} & \sigma_{\hat{\lambda} p} \\ \cdots & \sigma_{\hat{p}}^{2} \end{bmatrix}$$

3.4.2 Asymptotic Relative Efficiency for λ known

$$ARE(\hat{p}) = \frac{\sigma_{22}^{*}}{\sigma_{p}^{2}}$$

$$ARE(\hat{\alpha}) = \frac{\sigma_{33}^{*}}{\sigma_{\hat{\alpha}}^{2}}$$

$$JARE(\hat{p},\hat{\alpha}) = \frac{\det|\Sigma\Theta^{*}|}{\det|\Sigma\hat{\Theta}|}$$

where

$$\Sigma\Theta^{*} = \begin{bmatrix} \sigma_{22}^{*} & \sigma_{23}^{*} \\ \cdots & \sigma_{33}^{*} \end{bmatrix}$$

and

$$\Sigma\hat{\Theta} = \begin{bmatrix} \sigma_{\hat{p}}^{2} & \sigma_{\hat{p}\hat{\alpha}} \\ \cdots & \sigma_{\hat{\alpha}}^{2} \end{bmatrix}$$

3.4.3 Asymptotic Relative Efficiency for p known

$$ARE(\hat{\lambda}) = \frac{\sigma_{11}^{*}}{\sigma_{\hat{\lambda}}^{2}}$$

$$ARE(\hat{\alpha}) = \frac{\sigma_{33}^{*}}{\sigma_{\hat{\alpha}}^{2}}$$

$$\text{JARE}(\hat{\lambda}, \hat{\alpha}) = \frac{\det|\Sigma\theta^*|}{\det|\Sigma\hat{\theta}|}$$

where

$$\Sigma\theta^* = \begin{bmatrix} \sigma_{11}^* & \sigma_{13}^* \\ \cdots & \sigma_{33}^* \end{bmatrix}$$

and

$$\Sigma\hat{\theta} = \begin{bmatrix} \sigma_{\hat{\lambda}}^2 & \sigma_{\hat{\lambda}\hat{\alpha}} \\ \cdots & \sigma_{\hat{\alpha}}^2 \end{bmatrix}$$

3.5 Three Parameter Case

The maximum likelihood estimates for a sample of size three can be obtained by solving the first three factorial moments for α, λ, and p.

Now

$$f_1 = \alpha\lambda + (1-\alpha)np \qquad \text{---(1)}$$

$$f_2 = \alpha\lambda^2 + (1-\alpha)n(n-1)p^2 \qquad \text{---(2)}$$

$$f_3 = \alpha\lambda^3 + (1-\alpha)n(n-1)(n-2)p^3 \qquad \text{---(3)}$$

If we make the following simplifying assumptions for large n:

$$np = u$$

$$n(n-1)p^2 \approx u^2$$

$$n(n-1)(n-2)p^3 \approx u^3$$

and divide equation (2) by equation (1) we get

$$\frac{f_2 - u^2}{f_1 - u} = \frac{\lambda^2 - u^2}{\lambda - u} = \lambda + u \qquad \text{---(4)}$$

Dividing equation (3) by equation (1) we get

$$\frac{f_3 - u^3}{f_1 - u} = \frac{\lambda^3 - u^3}{\lambda - u} = \lambda^2 + \lambda u + u^2 \qquad \text{---(5)}$$

Now, from equation (4)

$$f_2 = \lambda f_1 - \lambda u + uf_1$$

and

$$\lambda = \frac{f_2 - uf_1}{f_1 - u} \qquad \text{---(6)}$$

From equation (5)

$$f_3 = \lambda^3(f_1 - u) + \lambda u\ (f_1 - u) + u^2 f_1$$

and

$$\frac{f_3 - u^2 f_1}{f_1 - u} = \lambda^2 + \lambda u \qquad \text{---(7)}$$

Substituting for λ we obtain

$$f_3 - u^2 f_1 = \frac{(f_2 - uf_1)^2}{f_1 - u} + (f_2 - uf_1)u$$

and

$$0 = u^2(f_1^{\ 2} - f_2) + u\ (f_3 - f_1 f_2) + f_2^{\ 2} - f_1 f_3$$

Solving for u we obtain

$$u = \frac{(f_1 f_2 - f_3) \pm R}{2(f_1^{\ 2} - f_2)} \qquad \text{---(8)}$$

where

$$R = \{ (f_1 f_2 - f_3)^2 - 4(f_1^{\ 2} - f_2)(f_2^{\ 2} - f_1 f_3) \}^{\frac{1}{2}}$$

Resolving equation (4) we get

$$u = \frac{f_2 - \lambda f_1}{f_1 - \lambda} \qquad \text{---(9)}$$

and rewriting equation (5) we have

$$\frac{f_3 - \lambda^2 f_1}{f_1 - \lambda} = u^2 + u\lambda \qquad \text{---(10)}$$

Substituting for u we obtain

$$f_3 - \lambda^2 f_1 = \frac{(f_2 - \lambda f_1)^2}{f_1 - \lambda} + (f_2 - \lambda f_1)\lambda$$

and

$$0 = \lambda^2 (f_1^2 - f_2) + \lambda(f_3 - f_2 f_1) + f_2^2 - f_1 f_3$$

Solving for λ we get

$$\lambda = \frac{(f_1 f_2 - f_3) \pm R}{2(f_1^2 - f_2)} \qquad \text{---(11)}$$

Now with $\mu = np$ and by assuming $n(n-1)p^2 \approx u^2$ (for large n) we have $\hat{u}$ and $\hat{\lambda}$ as the roots of the equation

$$\frac{F_1 F_2 - F_3 \pm R}{2(F_1^2 - F_2)} \qquad \text{---(12)}$$

and

$$\hat{\alpha} = \frac{F_1 - \hat{u}}{\hat{\lambda} - \hat{u}}.$$

We are now faced with the problem of a proper choice of sign of the radical to obtain the estimators. The simplest approach is to assume $\hat{u} > \hat{\lambda}$ (or vice versa) and take the appropriate sign. However, there is no concrete reasoning behind this, and since the equations were derived under simplifying assumptions, these estimators need not necessarily be Fisher-Consistent.

If we were to continue the study of the three parameter case from this point, we would study the asymptotic properties of the moment estimators. The matrix

$$\{\delta_{ij}\} = \{ \left. \frac{\partial H_i}{\partial F_j} \right|_{f} \}$$

would be determined firstly, as in the previous cases where one parameter is known. For the three parameter case we have a 3x3 matrix of differentials.

The covariance matrix of the moment estimators can be expressed as

$$\Sigma_H = \{\delta_{ij}\} \ \{u(F)\} \ \{\delta_{ij}\}'$$

where u(F) is the 3x3 covariance matrix of the sample factorial moments. The asymptotic efficiencies can now be computed following the method of the previous cases.

CHAPTER IV

4.1 Flow Chart and Program

Flow chart of program for an empirical study of the three cases when one parameter is known follows a listing of symbols and their code.

4.1.1 Listing of symbols and coding used:

Symbol	Code	Definition
n	B	Binomial n
α	A	Mixing Proportion
λ	PM	Poisson Parameter
P	P	Binomial Parameter
F_1	F1	1st Population factorial moment
F_2	F2	2nd "
F_3	F3	3rd "
F_4	F4	4th "
$\mu_2(F_1)$	VARF1	Variance of 1st pop^n fact'l moment
$\mu_2(F_2)$	VARF2	" 2nd "
$\mu_{11}(F_1,F_2)$	COV12	Covariance
$\partial H_1/\partial F_1\|f$	D_{11},C_{11},Q_{11}	2X2 Matrices of differentials
$\partial H_1/\partial F_2\|f$	D_{12},C_{12},Q_{12}	resp'ly for α known
$\partial H_2/\partial F_1\|f$	D_{21},C_{21},Q_{21}	λ known
$\partial H_2/\partial F_2\|f$	D_{22},C_{22},Q_{22}	np known
$Var(\hat{\lambda})$	VARPM	Variance of moment estimators α known
$Var(\hat{p})$	VARNP	"

Symbol	Code	Definition
$Cov(\hat{\lambda},\hat{p})$	VARPMP	Variance of moment estimators α known
$p(\hat{\lambda},\hat{p})$	CORA	Correlation
$Var(\hat{p})$	VNP	Variance of moment estimators λ known
$Var(\hat{\alpha})$	VA	"
$Cov(\hat{p},\hat{\alpha})$	VANP	"
$p(\hat{p},\hat{\alpha})$	CORNP	Correlation
$Var(\hat{\lambda})$	VAPM	Variance of moment estimators np known
$Var(\hat{\alpha})$	VAA	"
$Cov(\hat{\lambda},\hat{\alpha})$	VAAPM	"
$p(\hat{\lambda},\hat{\alpha})$	CORNP	Correlation

Symbol	Code	Symbol	Code	Symbol	Code
α known		λ known		np known	
$\det(\Sigma\hat{\lambda}\hat{p})$	VETA	$\det(\Sigma\hat{p}\hat{\alpha})$	VETPM	$\det(\Sigma\hat{\lambda}\hat{\alpha})$	VETNP
$\det(\Sigma^{-1}\lambda*p*)$	DETA	$\det(\Sigma^{-1}p*\alpha*)$	DETPM	$\det(\Sigma^{-1}\lambda*\alpha*)$	DETNP
$CRLB(\hat{\lambda})$	CRL11	$CRLB(\hat{p})$	CRLPM1	$CRLB(\hat{\lambda})$	CRLNP1
$CRLB(\hat{p})$	CRL22	$CRLB(\hat{\alpha})$	CRLPM2	$CRLB(\hat{\alpha})$	CRLNP2
$ARE(\hat{\lambda})$	AREA11	$ARE(\hat{p})$	AREM11	$ARE(\hat{\lambda})$	AREP11
$ARE(\hat{p})$	AREA22	$ARE(\hat{\alpha})$	AREM22	$ARE(\hat{\alpha})$	AREP22
$JARE(\hat{\lambda}\hat{p})$	CAREA	$JARE(\hat{p}\hat{\alpha})$	CAREM	$JARE(\hat{\lambda}\hat{\alpha})$	CAREP

det $\equiv$ determinant

CRLB $\equiv$ Cramér-Rao Lower Bound

ARE $\equiv$ Asymptotic relative efficiency

JARE $\equiv$ Joint asymptotic relative efficiency

4.1.2 Flow Chart

```
(Start)
   |
[Integer B]
   |
1  [Read A, PM, P, X]
   |
<DO  10  I = 2,10>
   |
B = I
F1 =
F2 =
F3 =
F4 =
   |
VARF1 =
VARF2 =
COV12 =
   |
C(I) = 1./(PM-(B-1)*P)
D11 =
D12 =
D21 =
D22 =
   |
VARPM  = Var(λ̂)
VARNP  = Var(np̂)
VARNP  = Var(p̂)
VARPMP = Cov(λ̂,np̂)
VARPMP = Cov(λ̂,p̂)
CORA   = Corr(λ̂,p̂)
   |
F    =
D(I) = 10/(P*B*P*B*F - 2. *F*PM*B*P+PM*PM)
C11  =
C12  =
C21  =
C22  =
   |
VNP   = Var(np̂)
VNP   = Var(p̂)
VA    = Var(α̂)
VANP  = Cov(np̂,α̂)
VANP  = Cov(p̂,α̂)
CORNP = Corr(p̂,α̂)
   |
   v
```

```
E      =
G      =
H(I)   = 1./(PM*PM-Z.*E*PM+G)
Q11    =
Q12    =
Q21    =
Q22    =
```

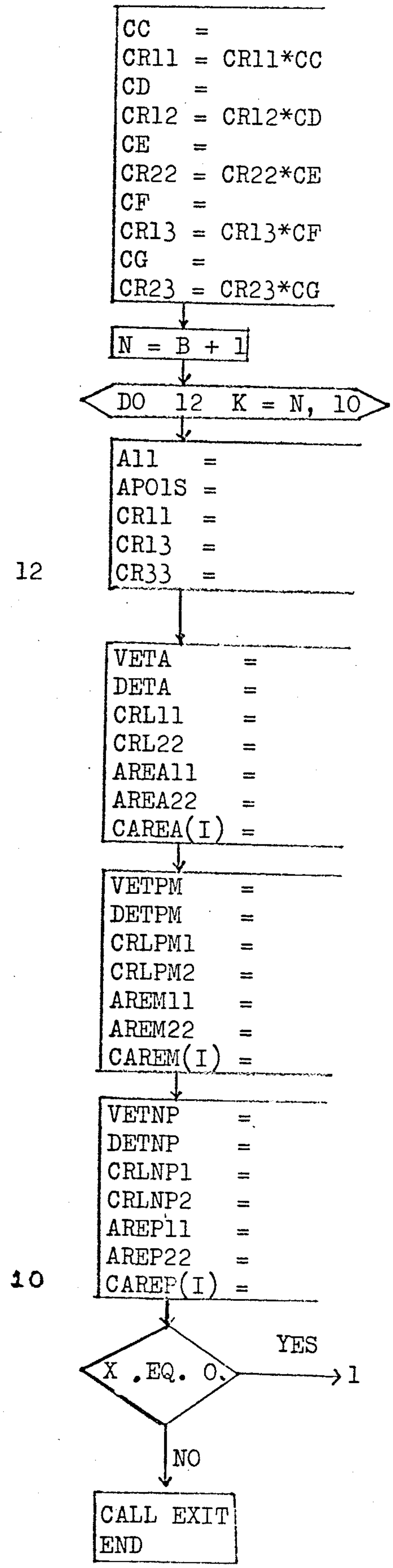
```
VAPM   = Var(λ̂)
VAA    = Var(α̂)
VAAPM  = Cov(λ̂,α̂)
CORNP  = Corr(λ̂,α̂)
```

```
CR11   =
CR12   =
CR22   =
CR13   =
CR23   =
CR33   =
PY     =
```

```
< DO  11  J = 1, B >
```

```
A1       =
A2       =
POIS     =
BIN      =
TPY(J)   =
```

```
   CR1(J) =
   CR11 = CR11 + CR1(J)
   CR2(J) =
   CR12 = CR12 + CR2(J)
   CR3(J) =
   CR13 = CR13 + CR3(J)
   CR4(J) =
   CR22 = CR22 + CR4(J)
   CR5(J) =
   CR23 = CR23 + CRS(J)
   CR6(J) =
11 CR33 = CR33 + CR6(J)
```



CC =
CR11 = CR11*CC
CD =
CR12 = CR12*CD
CE =
CR22 = CR22*CE
CF =
CR13 = CR13*CF
CG =
CR23 = CR23*CG
N = B + 1
DO 12 K = N, 10
A11 =
AP01S =
CR11 =
CR13 =
12
CR33 =
VETA =
DETA =
CRL11 =
CRL22 =
AREA11 =
AREA22 =
CAREA(I) =
VETPM =
DETPM =
CRLPM1 =
CRLPM2 =
AREM11 =
AREM22 =
CAREM(I) =
VETNP =
DETNP =
CRLNP1 =
CRLNP2 =
AREP11 =
AREP22 =
10
CAREP(I) =
X .EQ. 0.
YES
1
NO
CALL EXIT
END

4.1.3 Program


```
      DIMENSION CR1(100),CR2(100),CR3(100),CR4(100),CR5(100),CR6(100),
     1TPY(100),C(10),D(10),H(10), CAREA(10),CAREM(10), CAREP(10),
     2AREA11(10),AREA22(10),AREM11(10),AREM22(10),AREP11(10),AREP22(10)
      INTEGER B
C  READ IN PARAMETERS  A,PM,P
      L = 0
    1 READ(5,100) A,PM,P,X
  100 FORMAT (4F5.1)
      L = L + 1
      WRITE(6,102) L,A,PM,P
  102 FORMAT('1'10(/),15X,' 4.2.'I1,3X,'TABLE OF ASYMPTOTIC RELATIVE EFF
     1ICIENCY FOR A = 'F3.1,' PM = 'F3.1,' P = 'F3.1//34X,'    A  KNOWN
     2            PM  KNOWN                 P  KNOWN'/'  N      C        D
     3     H'//)
      DO 10 I = 2,10
      B = I
C  FACTORIAL MOMENTS AND THEIR VARIANCE
      F1 = A*PM + (1.-A)*B*P
      F2 = A*PM*PM + (1.-A)*B*(B-1.)*P*P
      F3 = A*PM*PM*PM + (1.-A)*B*(B-1.)*(B-2.)*P*P*P
      F4 = A*PM*PM*PM*PM + (1.-A)*B*(B-1.)*(B-2.)*(B-3.)*P*P*P*P
      VARF1 = F1 + F2 - F1*F1
      VARF2 = 2.*F2 + 4.*F3 + F4 -F2*F2
      COV12 = F3 + 2.*F2 -F1*F2
C  MIXING PROPORTION KNOWN
C  MATRIX OF DIFFERENTIALS   MIXING PROPORTION KNOWN
      C(I) = 1./(PM-(B-1.)*P)
      D11 = -C(I)*(B-1.)*P/A
      D12 = C(I) /(2.*A)
      D21 = C(I)*PM/(1.-A)
      D22 = -C(I)/(2.*(1.-A))
C  VARIANCE OF MOMENT ESTIMATORS,   MIXING PROPORTION KNOWN
      VARPM = VARF1*D11*D11 + VARF2 *D12*D12 + 2.*COV12*D11*D12
      VARNP = VARF1*D21*D21 + VARF2*D22*D22 + 2.*COV12*D21*D22
      VARNP = VARNP/(B*B)
      VARPMP = VARF1*D11*D21 + VARF2*D12*D22 + COV12*(D11*D22 +D21*D12)
      VARPMP = VARPMP/B
      CORA = VARPMP/SQRT(VARPM*VARNP)
C  POISSON MEAN KNOWN
C  MATRIX OF DIFFERENTIALS   POISSON MEAN KNOWN
      F = (1.-1./B)
      D(I) =    1./(P*B*P*B*F-2.*F*PM*B*P + PM*PM)
      C11 = D(I)*(PM*PM-F*B*B*P*P)/(1.-A)
      C12 = D(I)*(B*P-2.*PM)/(1.-A)
      C21 = -D(I)*2.*F*P*B
      C22 = D(I)
C  VARIANCE OF MOMENT ESTIMATORS,  POISSON MEAN KNOWN
      VNP = VARF1*C11*C11 + VARF2*C12*C12 + 2.*COV12*C11*C12
      VNP = VNP/(B*B)
      VA  = VARF1 *C21*C21 + VARF2 *C22*C22 + 2.*COV12*C21*C22
      VANP = VARF1*C11*C21 + VARF2*C12*C22 + COV12*(C11*C22 +C21*C12)
      VANP = VANP/B
      CORPM = VANP/SQRT(VNP*VA)
C  BINOMIAL MEAN KNOWN
C  MATRIX OF DIFFERENTIALS   BINOMIAL MEAN KNOWN
      E = B*P
      G = B*(B-1.)*P*P
      H(I) =    1./(PM*PM-2.*E*PM + G)
```

```
      Q11 = H(I)*(G-PM*PM)/A
      Q12 = H(I)*(PM-E)/A
      Q21 = 2.*PM
      Q22 = -H(I)/A
C   VARIANCE OF MOMENT ESTIMATORS,  BINOMIAL MEAN KNOWN
      VAPM = VARF1*Q11*Q11 + VARF2*Q12*Q12 + 2.*COV12*Q11*Q12
      VAA  = VARF1*Q21*Q21 + VARF2*Q22*Q22 + 2.*COV12*Q21*Q22
      VAAPM = VARF1*Q11*Q21 + VARF2*Q12*Q22 + COV12*(Q11*Q22+Q21*Q12)
      CORNP = VAAPM/SQRT(VAPM*VAA)
C   ELEMENTS OF INVERSE OF CRLB MATRIX
      CR11 = (-PM*EXP(-PM))**2/(A*EXP(-PM) + (1.-A)*(1.-P)**B)
      CR12 = PM*B*P*(1.-P)**B*EXP(-PM)/(A*EXP(-PM)+(1.-A)*(1.-P)**B)
      CR22 = ((1.-P)**B*(-B*P))**2/(A*EXP(-PM)+(1.-A)*(1.-P)**B)
      CR13 = (-PM)*EXP(-PM)*(EXP(-PM)-(1.-P)**B)/(A*EXP(-PM)+(1.-A)*
     1(1.-P)**B)
      CR23 = (1.-P)**B*(-B*P)*(EXP(-PM)-(1.-P)**B)/(A*EXP(-PM)+(1.-A)*
     1(1.-P)**B)
      CR33 = (EXP(-PM)-(1.-P)**B)**2/(A*EXP(-PM)+(1.-A)*(1.-P)**B)
      PY = A*EXP(-PM) + (1.-A)*(1.-P)**B
      DO 11 J = 1,B
      A1 = (J-PM)
      A2 = (J-B*P)
      POIS = EXP(-PM)*PM**J/FACT(J)
      BIN = COMB(B,J)*P**J*(1.-P)**(B-J)
      TPY(J) = A*POIS + (1.-A)*BIN
      PY = PY + TPY(J)
      CR1(J) = (POIS*A1)**2/TPY(J)
      CR11 = CR11 + CR1(J)
      CR2(J) = (POIS*BIN*A1*A2)/TPY(J)
      CR12 = CR12 + CR2(J)
      CR3(J) = POIS*A1*(POIS-BIN)/TPY(J)
      CR13 = CR13 + CR3(J)
      CR4(J) = (BIN*A2)**2/TPY(J)
      CR22 = CR22 + CR4(J)
      CR5(J) = BIN*A2*(POIS-BIN)/TPY(J)
      CR23 = CR23 + CR5(J)
      CR6(J) = (POIS-BIN)**2/TPY(J)
   11 CR33 = CR33 + CR6(J)
      CC= A*A/(PM*PM)
      CR11 = CR11*CC
      CD= A*(1.-A)/(PM*P*(1.-P))
      CR12 = CR12 *CD
      CE= (1.-A)*(1.-A)/(P*P*(1.-P)*(1.-P))
      CR22 = CR22*CE
      CF= A/PM
      CR13 = CR13*CF
      CG= (1.-A)/(P*(1.-P))
      CR23 = CR23*CG
      PYT = PY
      N = B + 1
      DO 12 K = N,10
      A11 = K-PM
      APOIS = EXP(-PM)*PM**K/FACT(K)
      PYT = PYT + A*APOIS
      CR11 = CR11 + CC*APOIS*A11*A11/A
      CR13 = CR13 + APOIS*A11*CF/A
   12 CR33 = CR33 + APOIS/A
C   WE NOW HAVE THE INVERSE OF CRLB MATRIX
```

```
C   FIND THE INVERSES OF THE INVERSE OF 2X2 CLLB MATRICES AND COMPARE
C   WITH APPROPRIATE VARIANCE OF THE ESTIMATORS  FOR THE ARE.
C   FIND DETERMINANTS OF THE COVARIANCE MATRIX OF THE ESTIMATORS AND
C   COMPARE WITH DET. OF THE INVERSE OF THE CRLB MATRIX FOR JARE.
      VETA = VARPM*VARNP-VARPMP*VARPMP
      DETA = CR11*CR22-CR12*CR12
      CRL11 = CR22/DETA
      CRL22 = CR11/DETA
      AREA11(I) = CRL11/VARPM
      AREA22(I) = CRL22/VARNP
      CAREA(I) = 1./(VETA*DETA)
      VETPM = VNP*VA-VANP*VANP
      DETPM = CR22*CR33-CR23*CR23
      CRLPM1 = CR33/DETPM
      CRLPM2 = CR22/DETPM
      AREM11(I) = CRLPM1/VNP
      AREM22(I) = CRLPM2/VA
      CAREM(I) = 1./(VETPM*DETPM)
      VETNP = VAPM*VAA-VAAPM*VAAPM
      DETNP = CR11*CR33-CR13*CR13
      CRLNP1 = CR33/DETNP
      CRLNP2 = CR11/DETNP
      AREP11(I) = CRLNP1/VAPM
      AREP22(I) = CRLNP2/VAA
      CAREP(I) = 1./(VETNP*DETNP)
   10 WRITE(6,103) I,C(I),D(I),H(I),AREA11(I),AREA22(I),CAREA(I),
     1AREM11(I),AREM22(I),CAREM(I),AREP11(I),AREP22(I),CAREP(I)
  103 FORMAT(I3,3(1X,F8.4),9(2X,F5.4)/)
      IF (X.EQ.0.) GO TO 1
      CALL EXIT
      END


      FUNCTION FACT(K)
      FACT = 1.
      DO 1 I = 1,K
    1 FACT = FACT*I
      RETURN
      END


      FUNCTION COMB(J,K)
      I = J-K
      COMB = FACT(J)/(FACT(K)*FACT(I))
      RETURN
      END
```

4.2.1 TABLE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR A = 0.4 PM = 1.5 P = 0.3

N	C	D	H	A KNOWN $(\hat{\lambda})$	A KNOWN $(\hat{\rho})$	A KNOWN $(\hat{\lambda},\hat{\rho})$	PM KNOWN $(\hat{\rho})$	PM KNOWN $(\hat{\alpha})$	PM KNOWN $(\hat{\rho},\hat{\alpha})$	P KNOWN $(\hat{\lambda})$	P KNOWN $(\hat{\alpha})$	P KNOWN $(\hat{\lambda},\hat{\alpha})$
2	0.8333	0.6536	1.5873	.9149	.7956	.7771	.0477	.6294	*****	.7890	.0753	.0236
3	1.1111	1.0101	11.1110	.9285	.9023	.8891	.0395	.7861	*****	.0926	.0056	.0000
4	1.6667	1.5873	-3.7037	.8558	.8491	.8316	.0202	.8980	.0808	.7049	.0290	.0048
5	3.3333	2.2222	-2.2222	.4764	.4782	.4581	.0106	.9634	.0189	.9623	.0304	.1477
6	********	2.2222	-2.2222	.0000	.0000	.0000	.0205	.9885	.0118	.8269	.0233	.0707
7	-3.3333	1.5873	-3.7037	.4425	.4575	.4308	.0496	.9748	.0166	.4117	.0122	.0016
8	-1.6667	1.0101	11.1116	.7700	.7962	.7554	.0986	.9723	.0332	.0388	.0014	.0000
9	-1.1111	0.6536	1.5873	.8605	.8944	.8460	.1688	.9744	.0698	.4482	.0238	.0026
10	-0.8333	0.4444	0.7407	.8758	.9246	.8596	.2600	.9725	.1412	.5641	.0483	.0140

**** For the choice of parametric region, results unexpectedly large.

4.2.2 TABLE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR A = 0.6 PM = 1.5 P = 0.3

N	C	D	H	A KNOWN			PM KNOWN			P KNOWN		
				$(\hat{\lambda})$	$(\hat{p})$	$(\hat{\lambda},\hat{p})$	$(\hat{p})$	$(\hat{\alpha})$	$(\hat{p},\hat{\alpha})$	$(\hat{\lambda})$	$(\hat{\alpha})$	$(\hat{\lambda},\hat{\alpha})$
2	0.8333	0.6536	1.5873	.8968	.7076	.6992	.0342	.5853	*****	.8703	.1857	.1073
3	1.1111	1.0101	11.1110	.9249	.8689	.8641	.0292	.7392	*****	.1078	.0121	.0001
4	1.6667	1.5873	-3.7037	.8952	.8712	.8664	.0157	.8643	.0775	.6849	.0498	.0093
5	3.3333	2.2222	-2.2222	.5306	.5081	.5014	.0094	.9401	.0181	.9862	.0588	.3357
6	********	2.2222	-2.2222	.0000	.0000	.0000	.0200	.9649	.0111	.8964	.0510	.1338
7	-3.3333	1.5873	-3.7037	.4661	.4544	.4423	.0494	.9721	.0161	.4637	.0290	.0036
8	-1.6667	1.0101	11.1116	.8234	.8185	.7994	.0979	.9799	.0331	.0451	.0038	.0000
9	-1.1111	0.6536	1.5873	.9281	.9318	.9126	.1617	.9751	.0698	.5452	.0750	.0093
10	-0.8333	0.4444	0.7407	.9510	.9613	.9381	.2364	.9601	.1401	.7047	.1930	.0933

**** For the choice of parametric region, results unexpectedly large.

4.2.3 TABLE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR A = 0.4 PM = .5 P = 0.1

N	C	D	H	A KNOWN $(\hat{\lambda})$	A KNOWN $(\hat{p})$	A KNOWN $(\hat{\lambda},\hat{p})$	PM KNOWN $(\hat{p})$	PM KNOWN $(\hat{\alpha})$	PM KNOWN $(\hat{p},\hat{\alpha})$	P KNOWN $(\hat{\lambda})$	P KNOWN $(\hat{\alpha})$	P KNOWN $(\hat{\lambda},\hat{\alpha})$
2	2.5000	5.8823	14.2857	.9345	.9164	.9097	.0918	.8137	*****	.9560	.1037	.0150
3	3.3333	9.0909	99.9994	.9832	.9822	.9807	.0645	.9446	*****	.2398	.0245	.0001
4	5.0000	14.2857	-33.3333	.9376	.9381	.9347	.0255	.9851	.0910	.9113	.0844	.0076
5	10.0000	19.9999	-20.0000	.6907	.6928	.6880	.0068	.9910	.0201	.9947	.0821	.1576
6	********	19.9999	-20.0001	.0000	.0000	.0000	.0336	.9994	.0123	.9361	.0715	.1383
7	-10.0000	14.2857	-33.3336	.7144	.7178	.7131	.1256	.9873	.0172	.7345	.0536	.0036
8	-5.0000	9.0909	100.0019	.8947	.8996	.8932	.2967	.9789	.0340	.1334	.0097	.0000
9	-3.3333	5.8824	14.2858	.9201	.9280	.9182	.5523	.9749	.0708	.7137	.0531	.0027
10	-2.5000	4.0000	6.6667	.9148	.9285	.9120	.8794	.9710	.1432	.7736	.0603	.0072

**** For the choice of parametric region, results unexpectedly large.

4.2.4 TABLE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR A = 0.4 PM = 1.0 P = 0.1

N	C	D	H	A KNOWN $(\hat{\lambda})$	A KNOWN $(\hat{p})$	A KNOWN $(\hat{\lambda},\hat{p})$	PM KNOWN $(\hat{p})$	PM KNOWN $(\hat{\alpha})$	PM KNOWN $(\hat{p},\hat{\alpha})$	P KNOWN $(\hat{\lambda})$	P KNOWN $(\hat{\alpha})$	P KNOWN $(\hat{\lambda},\hat{\alpha})$
2	1.1111	1.2195	1.6129	.8817	.7153	.6965	.0622	.5895	.8147	.9450	.1070	.1695
3	1.2500	1.5152	2.1739	.9069	.8323	.8173	.0706	.6873	*****	.9731	.1016	.0697
4	1.4286	1.9231	3.1250	.9385	.9109	.9016	.0703	.7864	*****	.9737	.0925	.0297
5	1.6667	2.5000	5.0000	.9673	.9596	.9553	.0616	.8735	.8627	.9384	.0802	.0114
6	2.0000	3.3333	9.9999	.9843	.9829	.9811	.0470	.9389	.1719	.8112	.0619	.0029
7	2.5000	4.5454	49.9977	.9802	.9800	.9782	.0303	.9763	.0491	.2120	.0144	.0000
8	3.3333	6.2500	-25.0001	.9424	.9429	.9396	.0155	.9881	.0164	.7252	.0443	.0009
9	5.0000	8.3333	-12.5000	.8387	.8401	.8357	.0055	.9891	.0065	.9821	.0545	.0122
10	10.0000	9.9998	-10.0000	.5386	.5401	.5367	.0020	.9954	.0034	.9946	.0509	.1583

**** For the choice of parametric region, results unexpectedly large.

4.2.5 TABLE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR A = 0.6 PM = 0.5 P = 0.1

N	C	D	H	A KNOWN $(\hat{\lambda})$	A KNOWN $(\hat{p})$	A KNOWN $(\hat{\lambda},\hat{p})$	PM KNOWN $(\hat{p})$	PM KNOWN $(\hat{\alpha})$	PM KNOWN $(\hat{p},\hat{\alpha})$	P KNOWN $(\hat{\lambda})$	P KNOWN $(\hat{\alpha})$	P KNOWN $(\hat{\lambda},\hat{\alpha})$
2	2.5000	5.8823	14.2857	.9117	.8711	.8681	.0786	.7788	*****	.9837	.2115	.0365
3	3.3333	9.0909	99.9994	.9740	.9687	.9680	.0550	.9079	*****	.2829	.0576	.0003
4	5.0000	14.2857	−33.3333	.9749	.9733	.9728	.0222	.9762	.0902	.8740	.1670	.0161
5	10.0000	19.9999	−20.0000	.7638	.7613	.7598	.0063	.9945	.0202	.9980	.1795	.3574
6	********	19.9999	−20.0001	.0000	.0000	*****	.0336	.9945	.0122	.9634	.1649	.3082
7	−10.0000	14.2857	−33.3336	.7211	.7203	.7180	.1262	.9883	.0172	.7510	.1249	.0081
8	−5.0000	9.0909	100.0019	.9182	.9182	.9153	.2962	.9899	.0343	.1387	.0230	.0001
9	−3.3333	5.8824	14.2858	.9546	.9554	.9517	.5352	.9927	.0721	.7632	.1298	.0066
10	−2.5000	4.0000	6.6667	.9592	.9614	.9556	.8144	.9935	.1464	.8373	.1492	.0185

**** For the choice of parametric region, results unexpectedly large.

4.2.6 TABLE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR A = 0.6 PM = 1.0 P = 0.1

N	C	D	H	A KNOWN $(\hat{\lambda})$	A KNOWN $(\hat{p})$	A KNOWN $(\hat{\lambda},\hat{p})$	PM KNOWN $(\hat{p})$	PM KNOWN $(\hat{\alpha})$	PM KNOWN $(\hat{p},\hat{\alpha})$	P KNOWN $(\hat{\lambda})$	P KNOWN $(\hat{\alpha})$	P KNOWN $(\hat{\lambda},\hat{\alpha})$
2	1.1111	1.2195	1.6129	.8609	.6376	.6286	.0525	.5635	.7196	.9124	.2301	.8484
3	1.2500	1.5152	2.1739	.8793	.7505	.7428	.0574	.6446	*****	.9551	.2109	.2368
4	1.4286	1.9231	3.1250	.9078	.8425	.8371	.0568	.7321	*****	.9814	.1900	.0854
5	1.6667	2.5000	5.0000	.9405	.9132	.9102	.0505	.8177	.7914	.9808	.1676	.0303
6	2.0000	3.3333	9.9999	.9705	.9622	.9610	.0395	.8927	.1624	.8969	.1367	.0075
7	2.5000	4.5454	49.9977	.9882	.9864	.9860	.0263	.9487	.0477	.2521	.0347	.0001
8	3.3333	6.2500	-25.0001	.9769	.9753	.9748	.0138	.9806	.0163	.6842	.0862	.0019
9	5.0000	8.3333	-12.5000	.8952	.8926	.8914	.0050	.9929	.0065	.9661	.1134	.0264
10	10.0000	9.9998	-10.0000	.5882	.5860	.5850	.0019	.9974	.0034	.9979	.1113	.3583

**** For the choice of parametric region, results unexpectedly large.

4.3 Discussion of Results

A computer oriented empirical study was completed for a chosen set of combinations of the parameters α, λ and p. In particular, we looked at the following twenty-four cases

$$\begin{Bmatrix} \alpha = 0.4 \\ \alpha = 0.6 \end{Bmatrix} \qquad \begin{Bmatrix} \lambda = 0.5 \\ \lambda = 1.0 \\ \lambda = 1.5 \end{Bmatrix} \qquad \begin{Bmatrix} p = 0.3 \\ p = 0.6 \\ p = 0.1 \\ p = 0.8 \end{Bmatrix}$$

for binomial parameter n equal to two to ten. We shall discuss some of the results and suggest areas requiring further study.

From the results it was seen that in general there was a pattern in the comparison of the ARE. Generally ARE and JARE approach zero for large values of n. When α is assumed known, it was seen that the ARE($\hat{\lambda}$) and ARE($\hat{p}$) are almost equal. When λ is assumed known, ARE($\hat{p}$) is much less than ARE($\hat{\alpha}$), and for the case p known, ARE($\hat{\alpha}$) is much less than ARE($\hat{\lambda}$). From the tables given, the JARE for the three cases follow the pattern:

$$\text{JARE}(\hat{\lambda},\hat{p}) > \text{JARE}(\hat{p},\hat{\alpha}) > \text{JARE}(\hat{\lambda},\hat{\alpha}).$$

For our particular choices of parametric values, the JARE($\hat{p},\hat{\alpha}$) for certain cases of n = 2,3, or 4 was unexpectedly large. This could possibly be due to the bias in the estimators.

Since the methods of moments is at best a crude, but widely used, method of estimation, the only condition that we can hope to have satisfied is that of consistency. In the method studied here the estimators were chosen to be Fisher-Consistent. Also with the method of moments, it is not possible to know whether the estimators are unbiased. We therefore propose that in pursuing the problem further, one may look for modifications of the estimators which would yield unbiased estimators for the parameters. Several suggestions for such study are: an examination of the radical in the estimating equation in order that an approximation might be found so that the estimator is unbiased, a study of the parametric region and the cause and effect of the terms

$$C = 1/(\lambda - (n-1)p)$$

$$D = 1/(n^2p^2(1-1/n) - 2(1-1/n)np\lambda + \lambda^2)$$

$$H = 1/(\lambda^2 - 2np\lambda + n(n-1)p^2)$$

on the variance of the estimators. It was shown in the empirical study that, for most combinations of λ and p, C,D and H tended to stabilize toward zero for values of n greater than or equal to five. Comparisons of the variances of the moment estimators with the Cramér-Rao lower bound would be more meaningful if the moment estimators (or their modifications) are unbiased. In the presence of bias mean squared error might be a more suitable criterion for comparison.

BIBLIOGRAPHY

Anscombe, F.J.(1950): "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions." Biometrika, 37, 358-362.

———, (1961): "Estimating a Mixed-Exponential Response Law." J. Amer. Stat. Assoc., 56, 493-502.

Barankin, E.W. and Gurland, J. (1951):

Barnard, G.A.(1954): "Sampling Inspection and Statistical decisions." J. Roy. Stat. Soc., 16B, 151-174.

Blischke, W.R. (1962): "Moment Estimators for the Parameters of a Mixture of Two Binomial Distributions." Ann. Math. Stat., 33, 444-454.

———, (1963): "Mixtures of Discrete Distributions." Proceedings of the International Symposium held at Montreal. Statistical Publishing Society. Calcutta and Pergamon Press Oxford. 351-372.

———, (1964): "Estimating the Parameters of Mixtures of Binomial Distributions. J. Amer. Stat. Assoc., 59, 510-528.

Boes, D.C. (1963): "On the Estimation of Mixing Distributions." (Abstract), Ann. Math. Stat., 34, 679.

Cassie, R.M. (1954): "Some Uses of Probability Paper in the Analysis of Size Frequency Distributions." Austral. Jour. Marine and Freshwater Research, 5, 513-523.

Chapman, D.G.(1961): "Statistical Problems in Dynamics of Exploited Fisheries Populations." Proc. Fourth Berkeley Symp. Math. Statist. and Prob., 4, 153-168.

Cohen, A.C. (1963): "Estimation in Mixtures of Discrete Distributions." Proceedings of the International Symposium on Classical and Contagious Discrete Distributions, Statistical Publishing Society, Calcutta, 373-378.

Cramer, Harald (1946): Mathematical Methods of Statistics, Princeton Univ. Press, Princeton.

Edwards, Carol B., and Gurland, John (1961): "A Class of Distributions Applicable to Accidents." J. Amer. Stat. Assoc., 56, 503-517.

Evans, D.A. (1953): "Experimental Evidence Concerning Contagious Distributions in Ecology. Biometrika, 40, 186-200.

Feller, W. (1943): "On a General Class of 'Contagious' Distributions." Ann. Math. Stat., 14, 389-400.

Fergusson, T. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities." Ann. Math. Stat., 29, 1046-62.

Fisher, R.A. (1922):

________, (1941): "The Negative Binomial Distribution." Ann. Eugenics, 11, 182-187.

Gurland, J. (1957): "Some Interrelations Among Compound and Generalized Distributions." Biometrika, 44, 265-268.

Gurland, J. and Shumway, R. (1960): "Fitting the Poisson Binomial Distribution." Biometrics, 16, 522-533.

Hald, A. (1960): "The Compound Hypergeometric Distribution and a System of Single Sampling Inspection Plans Based on Prior Distributions and Costs." Technometrics, 2, 275-340.

Kao, J.H.K. (1959): "A Graphical Estimation of Mixed Weibull Parameters in Life-Testing of Electron Tubes. Technometrics, 1, 389-407.

Katti, S.K. and Gurland, J. (1962a): "Some Methods of Estimation for the Poisson Binomial Distribution." Biometrics, 18, 42-51.

________, (1962b): "Efficiency of Certain Methods of Estimation for the Negative Binomial and the Neyman Type A Distributions." Biometrika, 49, 215-226.

Le Cam, L. (1956): "On the Asymptotic Theory of Estimation and Testing Hypotheses." Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1. Berkeley, California: U. of California Press, 129-156.

McGuire, J.U., Brindley, T.A. and Bancroft, T.A. (1957): "The Distribution of European Corn Borer Larvae Ryrausta Nubilalis (HBN), in Field Corn." Biometrics, 13, 65-78.

Medgyessy, P. (1961):* Decomposition of Superpositions of Distribution Functions, Publishing House of the Hungarian Academy of Sciences, Budapest.

Pearson, K. (1894): "Contributions to the Mathematical Theory of Evolution." Bhil. Trans. Roy. Soc. 185A, 71-110.

________, (1915): "On Certain Types of Compound Frequency Distributions in which the Components Can be Individually Described by Binomial Series." Biometrika, 11, 139-144.

Rao, C.R. (1952): Advanced Statistical Methods in Biometric Research John Wiley and Soon, Inc., New York.

_______ ,(1965): Linear Statistical Inference and Its Applications John Wiley and Sons, Inc.. New York.

Rider, P.R. (1961): "Estimating the Parameters of Mixed Poisson, Binomial, and Weibull Distributions by the Method of Moments." Bull, Inst. Inter. Stat., 39(Part 2), 143-147.

Robbins, H. (1948): "Mixture of Distributions." Ann. Math. Stat., 19, 360-369.

_______ ,(1950): "A Generalization of the Method of Maximum Likelihood: Estimating a Mixing Distribution." (Preliminary Report). (Abstract). Biometrics, 6, 194.

Schilling, W. (1947): "A Frequency Distribution Represented as the Sum of Two Poisson Distributions." J. Amer. Stat. Assoc., 42,407-424.

Shenton, L.R. (1949): "On the Efficiency of the Method of Moments and Neyman's Type A Distribution." Biometrika, 36, 450-54e.

_______ ,(1950): Maximum likelihood and the Efficiency of the Method of Moments." Biometrika, 37, 111-116.

_______ ,and Wallingthon, P.A. (1962): "The Bias of Moment Estimators with an Application to the Negative Binomial Distribution." Biometrika, 49, 193-204.

Sprott, D.A. (1958): "The Method of Maximum Likelihood Applied to the Poisson Binomial Distribution." Biometrics, 14, 97-106.

Teicher, H. (1960): "On the Mixture of Distributions." Ann. Math. Stat., 31, 55-73.

_______ , (1961): "Identifiability of Mixtures." Ann. Math. Stat., 32, 244-248.

_______ , (1963): "Identifiability of Finite Mixtures." Ann. Math. Stat., 34, 1265-1269.

Vagholkar, M.K. (1959): "The Process Curve and the Equivalent Mixed Binomial with Two Components." Jour. Roy. Stat. Soc. Notices, 8, 569.