ESTIMATION OF PARAMETERS OF THE
LOG-ZERO-POISSON DISTRIBUTION

by

A. Vijaya Rao and S. K. Katti

FSU Statistics Report ML65

Research Supported by

The United States Department of Agriculture
Grant 12-14-100-9183(20)

May, 1969
Department of Statistics
Florida State University
Tallahassee, Florida
SUMMARY

In this paper some estimators for the parameter of the log-zero-Poisson distribution are proposed. By a change of parameter from \((p_1, p, \lambda)\) to \((\theta, p, \lambda)\) the estimation problem is simplified. It is shown that the proportion of zeroes in the sample is the unique minimum variance unbiased estimator of the parameter \(\theta\). The efficiencies of twenty-seven estimators based on the minimum chisquare method is studied. It is shown that the estimator based on the sample frequencies \(f_0, f_1, f_2\) and the sample factorial moments \(m^{(1)}, m^{(2)}\) and \(m^{(3)}\) can be used as substitute for the maximum likelihood method in certain regions of the parameter space.
1. INTRODUCTION

Several distributions have been proposed for the statistical analysis of count data. They include the Poisson, the Poisson with zeroes, the negative binomial, the Neyman type A and the Poisson binomial distributions. Katti [1966] compared the skewness and kurtosis of the above distributions and a number of others which he himself developed in his unpublished Master's thesis. He observed that a distribution called log-zero-Poisson (l.z.P. for brevity) has a larger range for skewness and kurtosis than all the other distributions he considered. This implies that l.z.P. has more flexibility than others to describe empirical distributions and that statistical analysis of the data may be based on it.

As a check on the ability of the l.z.P. to describe empirical distributions, it was fitted to 35 empirical distributions and the fit was compared to the fits of some other theoretical distributions. It was found that the fit by the l.z.P. was comparable to the best among the fits by the other theoretical distributions in 32 of the 35 cases. Additional discussion on this may be found in Katti and Rao [1965]. Encouraged by these good empirical results, Katti and Rao [1969] proposed the use of the l.z.P. as a "one-distribution summary" of the many theoretical distributions discussed above. The principal objective of this paper is to develop and assess schemes for estimating the parameters of the l.z.P. distribution. Certain additional properties of the distribution will also be discussed.
2. EXPRESSIONS AND BOUNDS FOR THE PROBABILITIES

The probability generating function (p.g.f.) and the expressions for the probabilities of the l.z.P. distribution are given in Katti and Rao [1969]. They are given here for ready reference.

The p.g.f. of the distribution is

\[ g(z) = 1 - p_1 \ln(q-p \exp(\lambda(z-1))), \]

(2.1)

where \( 0 < p_1 < 1/\ln(q-p \exp(-\lambda)) \), \( \lambda > 0 \), \( p > 0 \) and \( q = 1 + p \).

Let \( P(i), i = 0,1,2,\ldots \) denote the probabilities. A recurrence relation for the probabilities is

\[ P(i+1) = [\lambda p_1 \pi_1(\lambda) + p \sum_{j=1}^{i} jP(j)\pi_{i-j+1}(\lambda)]/((i+1)(q-p \exp(-\lambda))) \]

where

\[ \pi_1(\lambda) = \exp(-\lambda)\lambda^i/i! \quad i = 0,1,2,\ldots \]

(2.2)

and

\[ P(0) = 1 - p_1 \ln(q-p \exp(-\lambda)) \]

(2.3)

An infinite series expression for the probabilities is:

\[ P(0) = 1 - p_1 \ln q + p_1 \sum_{j=1}^{\infty} \frac{1}{j} \frac{(p_j)^1}{q} \pi_0(j \lambda) \]

and

\[ P(i) = p_1 \sum_{j=1}^{\infty} \frac{1}{j} \frac{(p_j)^j}{q} \pi_j(j \lambda), \quad i = 1,2,\ldots \]

(2.4)
where \[ \pi_i(\lambda) = \exp(-\lambda)\lambda^i/i! \quad i = 0,1,2,\ldots. \]

Now, we will derive bounds for the probabilities with the hope that they will be useful in further theoretical developments.

**Theorem 2.1**

The probabilities \( P(i) \) in the l.z.p. distribution with parameters \( p_1, p \) and \( \lambda \) satisfy the following inequalities:

\[
\frac{p_1(\lambda)}{i! \mu} \left[ 1 - \frac{\mu}{\sqrt{2\pi(i-1)}} \exp\left(-\frac{1}{12(i-1)+1}\right) \right] < P(i) < \frac{p_1(\lambda)}{i! \mu} \left[ 1 + \mu \frac{2}{\sqrt{\pi(i-1)}} \exp\left(-\frac{1}{12(i-1)+1}\right) \right],
\]

(2.5)

where \( i > 0 \) and \( \mu = \lambda + \log(q/p) \).

**Proof:**

By (2.4), we have,

\[
P(i) = p_1 \sum_{j=1}^{\infty} \frac{1}{j! q^j} \pi_j(\lambda) = \frac{p_1(\lambda)}{i! \mu} \sum_{j=1}^{\infty} \frac{e^{-\mu j(\mu)}}{j^i}.
\]

(2.6)

The function \( \frac{e^{-\mu x}(\mu x)^i}{x} \) is an increasing function in \((0, \frac{1}{\mu})\) and decreasing in \((\frac{1}{\mu}, \infty)\). Also,

\[
\max_{x} \frac{e^{-\mu x}(\mu x)^i}{x} = \mu(1-1)^{i-1} \exp(-(1-1)).
\]

(2.7)
Let

\[ i^* = \left\lfloor \frac{i-1}{\mu} \right\rfloor, \text{ the greatest integer in } \frac{i-1}{\mu}. \]

Then, for \( j = 0, 1, 2, \ldots, i^* - 1 \),

\[
0 < \int_0^1 e^{-\mu x} \frac{(\mu x)^i}{x} \, dx < e^{-\mu} \mu^i,
\]

\[
e^{-\mu j\left(\frac{u}{\mu}\right)} j < \int_j^{j+1} e^{-\mu x} \frac{(\mu x)^i}{x} \, dx < \frac{e^{-\mu(j+1)}}{j+1},
\]

and, for \( j > i^* \),

\[
e^{-\mu(j+1)} \frac{e^{-\mu(j+1)^i}}{j+1} < \int_j^{j+1} e^{-\mu x} \frac{(\mu x)^i}{x} \, dx < \frac{e^{-\mu j(u+1)}}{j}, \tag{2.8}
\]

This leads to,

\[
\sum_{j=1}^{\infty} \frac{e^{-\mu j(u+1)} i}{j} = \sum_{j=1}^{i-1} \frac{e^{-\mu j(u+1)} i}{j} + \sum_{j=i^*}^{i+1} \frac{e^{-\mu j(u+1)} i}{j} + \sum_{j=i^*+2}^{\infty} \frac{e^{-\mu j(u+1)} i}{j} < \int_0^1 e^{-\mu x} \frac{(\mu x)^i}{x} \, dx - \int_{i^*}^{i+1} e^{-\mu x} \frac{(\mu x)^i}{x} \, dx
\]

\[
+ \sum_{j=i^*}^{i+1} \frac{e^{-\mu j(u+1)} i}{j} < \Gamma(i) + 2\mu(i-1)^{i-1} \exp(-(i-1)). \tag{2.9}
\]

We note the following Stirling's inequality:
\[ \sqrt{2\pi} \cdot n^{n+\frac{1}{2}} \cdot e^{-n} \cdot \frac{1}{12n+1} < \Gamma(n+1) < \sqrt{2\pi} \cdot n^{n+\frac{1}{2}} \cdot e^{-n} \cdot \frac{1}{12n}. \]

On combining (2.9), (2.6) and (2.10) we get the upper bound given the theorem. The lower bound follows similar lines. Hence the theorem is proved.

In equation (2.9), the sum
\[
\sum_{j=1}^{i^*} e^{-\mu_j^*(\mu_i^*)^1}
\]
should be regarded as zero whenever \( i^* \leq 1 \).
3. THE L.Z.P. DISTRIBUTION TRUNCATED AT ZERO

By (2.3),

\[ P(0) = 1 - p_1 \ln(q-p \exp(-\lambda)). \]

Therefore, the p.g.f. on the l.z.p truncated at zero is given by

\[ G(z) = \frac{g(z) - P(0)}{1-P(0)} = 1 - \frac{\ln[q-p \exp(\lambda(z-1))]}{\ln[q-p \exp(-\lambda)]}. \]

(3.1)

From here, we notice that the p.g.f. of the truncated distribution depends only on two parameters -- \( p \) and \( \lambda \). Therefore, the l.z.p. truncated at zero (l.z.p.t. for brevity) is a two parameter family. Next we note that \( G(z) \) can be obtained from \( g(z) \), the p.g.f. of l.z.p., by substituting \( 1/\ln (q-p \exp(-\lambda)) \) for \( p_1 \). This shows that the l.z.p. can be regarded as a subset of the l.z.p. family. The expressions for the probabilities and the bounds in the l.z.p.t. can be obtained from those of the l.z.p. by substituting for \( p_1 \) as above. For reference in later sections we add that an expression for the l.z.p.t. probability, obtained from (2.4) is

\[ P(i) = \sum_{j=1}^{\infty} \frac{1}{j} \frac{1}{(p/q)^j} \frac{\Gamma(j\lambda)}{\ln(q-p \exp(-\lambda))} \quad i = 1, 2, 3, \ldots . \]

(3.2)
4. PROPERTIES OF THE LIKELIHOOD FUNCTION.

Let $x_1, x_2, \ldots, x_n$, be a random sample from the $1.z.P$ distribution and let $f_0, f_1, \ldots$, be the sample frequencies of the various counts. Then the likelihood of the sample is:

$$L = \frac{n!}{f_0! f_1! \cdots} \prod_{i=0}^{\infty} \frac{f_i}{[P(i)]^i}.$$  

(4.1)

Theorem 4.1.

The likelihood $L$ factorizes as:

$$L = g_1(f_0/\theta) g_2(\Sigma i f_i | f_0, \lambda, \phi) g_3(f_1, f_2, f_n | f_0, \Sigma i f_i, \phi),$$  

(4.2)

where, $\theta = 1 - p_1 \ln(q - p \exp(-\lambda)) = P(0)$, $\phi = \frac{q}{p} \exp(\lambda)$ and $g_1, g_2, \text{ and } g_3$ are density functions of their arguments.

Proof:

Let the function $P$ denote the probability of its arguments.

The distribution of $f_0$, is clearly binomial with parameters $n$ and $\theta$.

Therefore, the probability density of $f_0$, is

$$g_1(f_0/\theta) = \binom{n}{f_0} (1-\theta)^{n-f_0} \theta^{f_0}.$$  

(4.3)

Note that (4.3) depends only on the parameter $\theta$.

Now,

$$P(f_1, f_2, \ldots, f_n | f_0) = \frac{P(f_0, f_1, \ldots, f_n, \ldots)}{P(f_0)} = \frac{(n-f_0)!}{f_0! f_1! \cdots} \prod_{i=1}^{\infty} \frac{f_i}{[\ln(q-p \exp(-\lambda))]}.$$

(4.4)
where we use the expression in (2.4) for \( P(i) \).

Clearly, the right hand side of (4.4) is the likelihood of a sample of size \((n-f_0)\) for the l.z.P.t. and does not involve \( \theta \).

Now consider the likelihood of a sample of size \((n-f_0)\) for the l.z.P.t. distribution with parameters \( p \) and \( \lambda \).

On substituting \( \phi = \frac{q}{p} \exp(\lambda) \) in the expression for the p.g.f., \( G(z) \) of l.z.P.t. (3.1) becomes

\[
G(z) = -\ln[1-(\exp(\lambda z)-1)/(\phi-1)]/\ln[(\phi-1)/(\phi-\exp(\lambda))]
\]

\[
= \sum_{j=1}^{\infty} C_j(\lambda z)^j/\ln[(\phi-1)/(\phi-\exp(\lambda))],
\]

where the \( C_j \)'s are functions of \( \phi \) only. Let \( H(z) \), denote the p.g.f. of the sum of a sample of size \((n-f_0)\) from the l.z.P.t. Then,

\[
H(z) = \{ \ln[(\phi-1)/(\phi-\exp(\lambda))] \}^{n-f_0} \{ \sum_{j=1}^{\infty} C_j(\lambda z)^j \}^{n-f_0}
\]

\[
= \{ \ln[(\phi-1)/(\phi-\exp(\lambda))] \}^{n-f_0} \{ \sum_{j=n-f_0}^{\infty} d_j(\lambda z)^j \}
\]

where, the \( d_j \)'s are functions of \( \phi \) only.

It follows that,

\[
P(f_1, f_2, \ldots | f_0, \Sigma i f_i) = \frac{(n-f_0)!}{f_1! f_2! \cdots} \prod_{i=1}^{n-f_0} (C_i)^{i} \cdot \frac{1}{d_{\Sigma i} f_i}.
\]

The right hand side depends only on \( \phi \). Therefore, we can write,

\[
P(f_1, f_2, \ldots, f_n | f_0) = g_2(\Sigma i f_i | f_0, \lambda, \phi) g_3(f_1, f_2, \ldots, f_n | \Sigma i f_i, f_0, \phi).
\]
where, the density $g_2$ of $\Sigma f_1$ given $f_0$, depends on both $\lambda$ and $\phi$, but the density $g_3$ of $f_1f_2 \ldots f_n$ given the zero frequency and the sample sum depends only on $\phi$.

On combining (4.8), (4.4) and (4.3) we get the factorization (4.2).

Hence the theorem is proved.

**Corollary 4.2.** If $\phi$ is known, then the statistic $\left( \frac{f_0}{n}, \frac{\Sigma f_1}{n} \right)$ is sufficient for the parameter $(\theta, \lambda)$.

**Proof:**

The Corollary follows easily for equation (4.8).

A consequence of the above theorem is the applicability of the notion of generalized sufficiency (Fraser [1956]) to the l.z.P distribution. This notion is defined below:

**Definition:** A statistic $t(x)$ is said to be Fraser sufficient (F-sufficient for brevity) for $\theta$ for the family $\{f(x, \theta, \eta) | (\theta, \eta) \in \Theta \times \eta \}$ if

1. the marginal distribution of $t(x)$ depends only on $\theta$, and,

2. the conditional distribution given $t(x)$ depends only on $\eta$.

Now suppose we change the parameters of the l.z.P distribution from $P_1, P_2, \lambda$ to $\theta, P, \lambda$ where $\theta = 1 - P_1 \ln (q-P \exp(-\lambda))$.

**Theorem 4.2.**

The sample zero frequency, $f_0$, is F-sufficient for $\theta$. 
Proof:

The theorem follows from equation (4.4) and the definition of F-sufficiency.

Corollary 4.3. \( T = f_0 / n \) is the unique minimum variance estimator for the parameter \( \theta \).

Proof:

Clearly, \( E(f_0 / n) = \theta \), and \( f_0 \) is F-sufficient for \( \theta \). The distribution of \( f_0 \) is binomial which is complete. Hence, by Fraser's theorem 3 (Fraser, [1956]) \( f_0 / n \) is the unique minimum variance estimator of \( \theta \).

In what follows, we will use \( \theta, p \) and \( \lambda \) as parameters of the l.z.p. distribution. Then the p.g.f. of the distribution will be

\[
g(z) = 1 - \frac{(1-\theta) \ln [q-p \exp(\lambda(z-1))] }{\ln(q-p \exp(-\lambda))} \tag{4.9}
\]

where

\[
0 < \theta < 1, \quad p > 0, \quad \lambda > 0 \quad \text{and} \quad q = 1 + p.
\]
5. LIKELIHOOD EQUATIONS

The likelihood equations for estimating the parameters $p_1$, $p$ and $\lambda$ are derived and discussed in Katti and Rao [1969]. For ready reference they are reproduced below without proof.

**Theorem 5.1.**

Let $f_0, f_1, \ldots$ be the frequencies of various counts in a sample of size $n$ from the $l.z.p$ distribution with probabilities $P(0), P(1), \ldots$ and parameters $p_1, p$ and $\lambda$ and let $\bar{x}$ be the sample mean. If $0 < f_0 < n$, then the likelihood equations are:

$$1 - p_1 \ln(q - p \exp(-\lambda)) = f_0/n,$$

(5.1)

$$\lambda p_1 p = \bar{x},$$

(5.2)

and

$$\sum_{i=0}^{\infty} (i+1)(f_i/n)(P(i+1)/P(i)) = \bar{x}. \quad (5.3)$$

If $f_0 = 0$, then the likelihood equations are:

$$\ln(q - p \exp(-\lambda)) = 1/p_1,$$

(5.4)

$$\lambda p/\ln(q - p \exp(-\lambda)) = \bar{x},$$

(5.5)

and

$$\sum_{i=1}^{\infty} \left( (i+1) \left( \frac{f_i}{n} \right) \frac{P(i+1)}{P(i)} + \lambda \frac{p \exp(-\lambda)}{(q - p \exp(-\lambda))} \right) \log(q - p \exp(-\lambda)) = \bar{x}. \quad (5.6)$$
The two cases are due to the fact that when \( f_0 = 0 \), the likelihood function becomes a monotonically increasing function of \( p_1 \) and the maximum with respect to \( p_1 \) occurs when \( p_1 \) is a maximum i.e., \( p_1 = 1/\ln(q-p \exp(-\lambda)) \).

When the likelihood with this value for \( p_1 \) is differentiated, we get (5.5) and (5.6). If \( f_0 \neq 0 \), then the maximum with respect to \( p_1 \) does not occur at this maximum attainable value of \( p_1 \) and we get a set (5.1), (5.2) and (5.3) of equations different from the set for the case \( f_0 = 0 \). In the next theorem, we will investigate the maximum likelihood scheme when the distribution is reparameterized to \((\theta, p, \lambda)\).

**Theorem 5.2.**

Let \( f_0, f_1, \ldots, f_n \ldots \) be the frequencies of various counts in a sample of size \( n \) for the \( 1.z.P \) distribution with probabilities \( P(0), P(1) \ldots \) and parameters \( \theta, p \) and \( \lambda \) and let \( \bar{x} \) be sample mean. Then, the likelihood equations are:

\[
\theta = f_0/n, \tag{5.7}
\]

\[
(1 - f_0/n)\lambda p/\ln(q-p \exp(-\lambda)) = \bar{x}, \tag{5.8}
\]

and

\[
\sum_{i=1}^{\infty} (i+1)(f_i/n)\frac{P(i+1)}{P(i)} + P(1) = \bar{x}. \tag{5.9}
\]

**Proof:**

To derive the likelihood equations we need, the expressions for the partial derivatives of the probabilities. The partial derivatives are,
\[ \frac{\partial P(0)}{\partial \theta} = 1, \]
\[ \frac{\partial P(i)}{\partial \theta} = -\frac{P(i)}{1-\theta}, \quad i = 1,2,\ldots; \]
\[ \frac{\partial P(0)}{\partial p} = 0, \]
\[ \frac{\partial P(i)}{\partial p} = (i+1)\frac{P(i+1)}{\lambda pq} - \frac{P(1)P(i)}{\lambda p(1-\theta)} \left( \exp(\lambda) - 1 \right), \quad i = 1,2,\ldots; \]
\[ \frac{\partial P(0)}{\partial \lambda} = 0, \]
and
\[ \frac{\partial P(i)}{\partial \lambda} = \left[ iP(i) - (i+1)P(i+1) \frac{P(1)P(i)}{1-\theta} \right] / \lambda, \quad i = 1,2,3,\ldots. \]

On substituting these expressions in the partial derivatives of the likelihood function we get,
\[ \frac{\partial \ln L}{\partial \theta} = \frac{f_0}{\theta} - \frac{n-f_0}{1-\theta}, \]
\[ \frac{\partial \ln L}{\partial p} = \frac{1}{\lambda pq} \sum_{i=1}^{\infty} (i+1)f \frac{P(i+1)}{P(i)} - \frac{(n-f_0)(\exp(\lambda) - 1)P(1)}{\lambda p(1-\theta)}, \]
and
\[ \frac{\partial \ln L}{\partial \lambda} = n - \frac{1}{\lambda} \sum_{i=1}^{\infty} (i+1)f \frac{P(i+1)}{P(i)} - \frac{(n-f_0)P(1)}{\lambda(1-\theta)}. \]

On equating these to zero and simplifying, we get equations (5.7), (5.8) and (5.9). Since the estimate of \( \theta \) is \( f_0/n \), whether or not \( f_0 \) is zero, these equations are the likelihood equations for both the cases \( f_0 = 0 \) and \( f_0 \neq 0 \). This is to be contrasted with the estimation of \((P_1, P, \lambda)\) where depending on \( f_0 = 0 \) or \( f_0 \neq 0 \), we had different sets of equations.

In view of this, we will use \((\theta, P, \lambda)\) as the parameters of the 1.2.3.P distribution hereafter.
Theorem 6.1.

The information matrix \( nI(\theta, p, \lambda) \) of a sample of size \( n \) from the 1. z. P. distribution with parameter \( \theta, p, \lambda \) and probabilities \( P(0), P(1), \ldots \) is

\[
nI(\theta, p, \lambda) = n \begin{pmatrix}
I_{11}(\theta, p, \lambda) & 0 & 0 \\
0 & I_{22}(\theta, p, \lambda) & I_{23}(\theta, p, \lambda) \\
0 & I_{32}(\theta, p, \lambda) & I_{33}(\theta, p, \lambda)
\end{pmatrix}
\]

where,

\[
I_{11}(\theta, p, \lambda) = \frac{1}{\theta(1-\theta)},
\]

\[
I_{22}(\theta, p, \lambda) = \frac{1}{\lambda p^2 q^2} \sum_{i=1}^{\infty} (i+1)^2 \frac{p^2(i+1)}{P(i)} + \frac{(\exp(\lambda) -1)^2}{\lambda^2 p^2 (1-\theta)},
\]

\[
I_{23}(\theta, p, \lambda) = \frac{(1-\theta)}{\ln(q-p \exp(-\lambda))} - \frac{1}{\lambda^2 pq} \sum_{i=1}^{\infty} (i+1)^2 \frac{P^2(i+1)}{P(i)}
\]

\[- \frac{(\exp(\lambda) -1)}{\lambda^2 p(1-\theta)} P^2(1),
\]

\[
I_{32}(\theta, p, \lambda) = I_{23}(\theta, p, \lambda),
\]

and

\[
I_{33}(\theta, p, \lambda) = \frac{1}{\lambda^2} \sum_{i=1}^{\infty} (i+1)^2 \frac{P^2(i+1)}{P(i)} + \frac{(1-\theta)p(1-\lambda q)}{\lambda \ln(q-p \exp(-\lambda))} \frac{P^2(1)}{\lambda^2 (1-\theta)}.
\]

Proof:

We recall the definition of information and write:
\[ nI_{11}(\theta, p, \lambda) = E\left( \frac{\partial^2 \ln L}{\partial \theta^2} \right) = E\left( \frac{\partial}{\partial \theta} \frac{f}{1-\theta} \frac{n-f}{n-1} \right) = \frac{n}{\theta(1-\theta)} . \] (6.6)

Now,
\[ \frac{\partial^2 \ln L}{\partial \theta \partial \theta} = 0, \]
and
\[ \frac{\partial^2 \ln L}{\partial p \partial \theta} = 0. \]

Therefore,
\[ I_{12}(\theta, p, \lambda) = I_{13}(\theta, p, \lambda) = 0. \] (6.7)

\[ nI_{22}(\theta, p, \lambda) = E\left( \frac{\partial^2 \ln L}{\partial p^2} \right) = E\left( \frac{\partial}{\partial p} \frac{\partial}{\partial p} \frac{\ln L}{p} \right) . \]

On substituting for \( \frac{\partial \ln L}{\partial p} \) from (5.16) and simplifying we get the expression in (6.2).

Expressions for \( I_{23} \) and \( I_{33} \) follow from similar arguments.

**Theorem 6.2.**

The information matrix \( nI(p, \lambda) \) of a sample of size \( n \) from the l.z.P.t. distribution with parameters \( \lambda \) and \( p \) and probabilities \( P^*(1), P^*(2), \ldots \), is

\[ nI(p, \lambda) = n \begin{pmatrix} I_{11}(p, \lambda) & I_{12}(p, \lambda) \\ I_{21}(p, \lambda) & I_{22}(p, \lambda) \end{pmatrix} \]

where
\[ I_{11}(p, \lambda) = \frac{1}{\lambda^2 pq} \sum_{i=1}^{\infty} (i+1)^2 \frac{P^*(i+1)}{P^*(i)} + \frac{(\exp(\lambda) - 1)^2}{\lambda^2 p^2} P^*(1), \quad (6.9) \]

\[ I_{12}(p, \lambda) = \frac{1}{\ln(q-p \exp(-\lambda))} - \frac{1}{\lambda pq} \sum_{i=1}^{\infty} (i+1)^2 \frac{P^*(i+1)}{P^*(i)} - \frac{(\exp(\lambda) - 1)^2}{\lambda^2 p} P^*(1), \quad (6.10) \]

\[ I_{21}(p, \lambda) = I_{12}(p, \lambda), \]

and

\[ I_{22}(p, \lambda) = \frac{1}{\lambda^2} \sum_{i=1}^{\infty} (i+1)^2 \frac{P^*(i+1)}{P^*(i)} + \frac{p(1-\lambda q)}{\lambda \ln(q-p \exp(-\lambda))} - \frac{P^*(1)}{\lambda^2}, \quad (6.12) \]

Proof is omitted for brevity.

**Corollary 6.3.**

Let \(|I(\theta, p, \lambda)|\) denote the information determinant of size 1 from the l.z.P. distribution with parameters \(\theta\), \(p\) and \(\lambda\) and let \(|I(p, \lambda)|\), denote the informative determinant of size 1 for the l.z.P.t. with parameters \(\lambda\) and \(p\). Then we have

\[ |I(\theta, p, \lambda)| = \frac{1-\theta}{\theta} |I(p, \lambda)|. \quad (6.13) \]

Proof:

The proof follows for the fact that if \(P(i)\) and \(P^*(i)\) are the probabilities in l.z.P and l.z.P.t., then

\[ P(i) = (i-\theta)P^*(i), \quad i = 1, 2, \ldots, \quad (6.14) \]
Theorem 6.4.

The information determinant \(|I(p, \lambda)|\) of a sample of size 1 for the 1.z.P.t. with parameter \(\lambda\) and \(p\) and probabilities \(P^*(1), P^*(2), \ldots\) has the following expression:

\[
|I(p, \lambda)| = f_1(p, \lambda) \sum_{i=1}^{\infty} (i+1)^2 \frac{P^*(i+1)}{P^*(i)} - f_2(p, \lambda),
\]

(6.15)

where,

\[
f_1(p, \lambda) = \frac{\{(1+\lambda q) \ln (q-p \exp(-\lambda)) - \lambda p\} \cdot \{\lambda^3 pq^2 [\ln(q-p \exp(-\lambda))]^2\}}{\{\lambda^3 pq^2 [\ln(q-p \exp(-\lambda))]^2\}},
\]

(6.16)

and,

\[
f_2(p, \lambda) = \frac{(\exp(\lambda) - 1)^2 P^*(1) \ln(q-p \exp(-\lambda)) + 2(\exp(\lambda) - 1) P^*(1) \lambda^2 p \ln(q-p \exp(-\lambda))}{\lambda^3 p \ln(q-p \exp(-\lambda))} - \frac{1}{[\ln(q-p \exp(-\lambda))]^2}.
\]

(6.17)

The function \(f_1(p, \lambda)\) is positive in the region \(p, \lambda > 0\).

Proof:

Derivation of equations (6.15), (6.16) and (6.17) follows by expanding \(|I(p, \lambda)|\) and simplifying.

To prove that \(f_1(p, \lambda)\) is positive it suffices to prove that the numerator of \(f_1(p, \lambda)\) is positive.

Let

\[
f_3(p, \lambda) = \lambda \ln(q-p \exp(-\lambda) - \lambda p).
\]
Now,
\[
\frac{\partial f_3(p, \lambda)}{\partial p} = \lambda \ln(q-p \exp(-\lambda)) + \frac{1-\lambda - \lambda e^{-\lambda}}{q-p e^{-\lambda}} > 0
\]

Since, for \( \lambda > 0 \), \( 1 - e^{-\lambda} - \lambda e^{-\lambda} > 0 \) and \( q - pe^{-\lambda} > 1 \).

Also,
\[
f_1(0, \lambda) = 0 \quad \text{for every} \quad \lambda > 0
\]

Therefore, \( f_1(p, \lambda) > 0 \) for \( \lambda, p > 0 \).

From the above theorems we note that the information determinant of 1.z.P. with parameter \((\theta, p, \lambda)\) is the product of a simple function of \( \theta \) and a second function involving \( p, \lambda \) but not \( \theta \). Thus the problem of tabulating a function of 3 variables is reduced to that of tabulating a function of two variables \( p \) and \( \lambda \). It should be noted that the information determinant of the 1.z.P. with parameters \( p_1, p \) and \( \lambda \) does not have this simplification.

From Corollary (6.3) we note that difficulties in the numerical evaluation of the information determinant arise in the infinite series
\[
\sum_{i=1}^{\infty} (i+1)^2 p^* (i+1)^2.
\]

In the absence of a closed expression for this series, its value may be approximated by the partial sums of the series. Since the function \( f_1(p, \lambda) \) in equation (6.14) is positive it follows that this approximation based on the partial sums will lead to a lower bound for the value of the information determinant. Since the value of the information determinant appears in the denominator of the expression for efficiency of estimates, this value of the information determinant when used will lead to an upper bound for the efficiency.
An upper bound is not sufficient to assess the merits of estimators. So we will develop lower bounds for the estimators by developing upper bounds for the information determinant.

**Theorem 6.5.**

The series

\[
\sum_{i=1}^{\infty} (i+1)^2 \frac{\mathbf{P}^*(i+1)}{\mathbf{P}^*(i)}
\]

where \(\mathbf{P}^*(i)\)'s are the probabilities in the l.z.P.t. distribution with parameters \(\lambda, p\), is convergent and the remainder after \(n\) terms of the series,

\[
R_n = \sum_{i=n+1}^{\infty} (i+1)^2 \frac{\mathbf{P}^*(i+1)}{\mathbf{P}^*(i)}
\]

satisfies the following inequality:

\[
R_n < \frac{1}{\ln(q-p \exp(-\lambda))} \left( \frac{\lambda}{\mu} \right)^{n+1} \left( 1 + \mu \sqrt{2} \frac{1}{\sqrt{n+1}} \exp(-\frac{1}{12n+13}) \right)^2
\]

\[
/((1 - \mu \exp(-\frac{1}{12n+1}))(1 - \frac{\lambda}{\mu} \frac{n+2}{n+1})).
\]

for \(n > \max(\frac{2\lambda - \mu}{\mu - \lambda}, \frac{-1}{\log(\frac{\lambda}{\mu})})\) where \(\mu = \lambda + \log \frac{q}{p}\).

**Proof:**

The bounds for \(\mathbf{P}(i)\) given in Theorem 2.1 will be used in deriving (6.17).

On substituting \(p_1 = (1-\theta)/\ln(q-p \exp(-\lambda))\) and \(\mathbf{P}(i) = (1-\theta)\mathbf{P}^*(i)\) in expression (2.5) and simplifying, we get,
\[
\frac{1}{i \ln(q-p \exp(-\lambda))} \left( \frac{\lambda}{\mu} \right) i \left[ 1 - \frac{\mu}{\sqrt{2\pi(i-1)}} \exp\left(\frac{-1}{12(i-1)+1}\right) \right] < P^*(1)
\]

\[
< \frac{1}{i \ln(q-p \exp(-\lambda))} \left( \frac{\lambda}{\mu} \right) i \left[ 1 + \frac{\sqrt{2}}{\mu \sqrt{(i-1)}} \exp\left(\frac{-1}{12(i-1)+1}\right) \right], \ i = 1, 2, \ldots
\]

On substituting these bounds in \( R_n \), we get,

\[
R_n = \sum_{i=n+1}^{\infty} \frac{(i+1)^2 P^*(i+1)}{P^*(i)} < \frac{1}{\ln(q-p \exp(-\lambda))} \sum_{i=n+1}^{\infty} a_i
\]

where

\[
a_i = \left( \frac{\lambda}{\mu} \right) i^2 \frac{[i+\mu \sqrt{\frac{2}{\pi(i-1)}} \exp\left(\frac{-1}{12(i-1)+1}\right)]^2}{[1 - \frac{\mu}{\sqrt{2\pi(i-1)}} \exp\left(\frac{-1}{12(i-1)+1}\right)]}.
\]

Now,

\[
\lim_{i \to \infty} \frac{a_{i+1}}{a_i} = \frac{\lambda}{\mu} < 1
\]

Therefore, the series

\[
\sum_{i=1}^{\infty} \frac{(i+1)^2 P^*(i+1)}{P^*(i)}
\]

is convergent.

It can be readily seen that the functions \( i \left( \frac{\lambda}{\mu} \right) i^2 \) and \( \left( \sqrt{i} \exp \left( \frac{1}{12i+1} \right) \right)^{-1} \)

are decreasing functions of \( i \) if \( i > \frac{1}{\log\left( \frac{\lambda}{\mu} \right)} \) and \( i > 1 \) respectively.
Therefore,

\[
\frac{a_{i+1}}{a_i} < \frac{\lambda}{\mu} \frac{i+1}{i} < \frac{\lambda}{\mu} \frac{n+2}{n+1} < 1 \text{ for } i > n + 1
\]

if \( n > \frac{(2\lambda - \mu)}{(\mu - \lambda)} \). It follows from here that,

\[
R_n \leq \frac{1}{\ln(q-p \exp(-\lambda))} \left( \frac{\lambda}{\mu} \right)^{n+3} (n+1) \times \left[ 1 + \mu \sqrt{\frac{2}{\pi(n+1)}} \exp\left( \frac{-1}{12n+14} \right) \right]^2 \left[ 1 - \frac{\mu}{2\ln} \exp\left( \frac{-1}{12(n+1)} \right) \right] \left[ 1 - \frac{\lambda}{\mu} \frac{n+2}{n+1} \right].
\]

(6.18)

This proves the theorem.

The right hand side of (6.18) is used in deriving an upper bound for the information determinant.

The following procedure was adopted in evaluating the information determinant of the l.z.p.t. distribution.

Let,

\[
S_n = \sum_{i=1}^{n} (i+1)^2 \frac{P^*(i+1)}{P^*(i)},
\]

\[
T_i = (i+1)^2 \frac{P^*(i+1)}{P^*(i)},
\]

\( S_n \) is taken as the approximate value of the series

\[
S = \sum_{i=1}^{\infty} (i+1)^2 \frac{P^*(i+1)}{P^*(i)} \text{ if } \frac{T_{n+1}}{S_n} < 10^{-5}
\]

and \( n \geq \max \left( \frac{(2\lambda - \mu)}{(\mu - \lambda)} \text{ and } -1/\ln(\lambda/\mu) \right) \). On substituting this value
for $S$ in (6.15) we obtain a lower bound for the information determinant. An upper bound is obtained by evaluating the bound (6.18) for $R_n$ and substituting the value $(S_n + \text{the bound for } R_n)$ in (6.15). To obtain the bounds for the 1.z.p. distribution, we need only to multiply the bounds corresponding to the 1.z.p.t. distribution by $\frac{1-\theta}{\theta}$.

For calculating the $T(i)$'s we used the recurrence relationship for the probabilities given by (2.2). It should be noted here, that the $T(i)$'s tend to zero very slowly for large $p$ and $\lambda$ and consequently the evaluation of the information determinant becomes difficult as $\lambda$ and $p$ increase.

In table 1, we give the upper and lower bounds for the information matrix for some values of $\lambda$ and $p \leq 5.0$. 
TABLE 1

LOWER AND UPPER BOUNDS FOR THE INFORMATION DETERMINANT
OF THE L.Z.P.T. DISTRIBUTION

(Against each value of \( p \), the lower bound is shown on the top line and the upper bound in the bottom line. Figures here are the values of the information determinant multiplied by \( 10^5 \).)

<table>
<thead>
<tr>
<th>( p )</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>1383</td>
<td>2103</td>
<td>2564</td>
<td>2857</td>
<td>3038</td>
<td>3133</td>
<td>3162</td>
<td>3145</td>
<td>3099</td>
<td>3032</td>
</tr>
<tr>
<td>0.5</td>
<td>1392</td>
<td>2121</td>
<td>2578</td>
<td>2878</td>
<td>3059</td>
<td>3153</td>
<td>3187</td>
<td>3176</td>
<td>3129</td>
<td>3068</td>
</tr>
<tr>
<td>1.0</td>
<td>697</td>
<td>929</td>
<td>1039</td>
<td>1091</td>
<td>1110</td>
<td>1105</td>
<td>1084</td>
<td>1052</td>
<td>1016</td>
<td>977</td>
</tr>
<tr>
<td>1.5</td>
<td>704</td>
<td>934</td>
<td>1044</td>
<td>1099</td>
<td>1119</td>
<td>1115</td>
<td>1096</td>
<td>1067</td>
<td>1033</td>
<td>996</td>
</tr>
<tr>
<td>2.0</td>
<td>422</td>
<td>514</td>
<td>548</td>
<td>557</td>
<td>554</td>
<td>541</td>
<td>522</td>
<td>501</td>
<td>479</td>
<td>456</td>
</tr>
<tr>
<td>2.5</td>
<td>425</td>
<td>518</td>
<td>552</td>
<td>563</td>
<td>560</td>
<td>548</td>
<td>532</td>
<td>511</td>
<td>490</td>
<td>469</td>
</tr>
<tr>
<td>3.0</td>
<td>280</td>
<td>322</td>
<td>332</td>
<td>330</td>
<td>323</td>
<td>311</td>
<td>298</td>
<td>284</td>
<td>269</td>
<td>254</td>
</tr>
<tr>
<td>3.5</td>
<td>283</td>
<td>325</td>
<td>335</td>
<td>335</td>
<td>328</td>
<td>318</td>
<td>306</td>
<td>292</td>
<td>279</td>
<td>266</td>
</tr>
<tr>
<td>4.0</td>
<td>199</td>
<td>218</td>
<td>219</td>
<td>215</td>
<td>208</td>
<td>199</td>
<td>188</td>
<td>178</td>
<td>168</td>
<td>158</td>
</tr>
<tr>
<td>4.5</td>
<td>201</td>
<td>220</td>
<td>222</td>
<td>219</td>
<td>212</td>
<td>204</td>
<td>195</td>
<td>186</td>
<td>177</td>
<td>168</td>
</tr>
<tr>
<td>5.0</td>
<td>147</td>
<td>156</td>
<td>153</td>
<td>149</td>
<td>142</td>
<td>135</td>
<td>128</td>
<td>120</td>
<td>112</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>149</td>
<td>158</td>
<td>157</td>
<td>152</td>
<td>147</td>
<td>141</td>
<td>134</td>
<td>127</td>
<td>121</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>113</td>
<td>116</td>
<td>113</td>
<td>108</td>
<td>103</td>
<td>97</td>
<td>91</td>
<td>85</td>
<td>79</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>114</td>
<td>118</td>
<td>115</td>
<td>112</td>
<td>107</td>
<td>102</td>
<td>97</td>
<td>92</td>
<td>87</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>89</td>
<td>89</td>
<td>35</td>
<td>81</td>
<td>77</td>
<td>72</td>
<td>67</td>
<td>62</td>
<td>57</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>91</td>
<td>88</td>
<td>85</td>
<td>81</td>
<td>77</td>
<td>73</td>
<td>69</td>
<td>66</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>71</td>
<td>70</td>
<td>66</td>
<td>63</td>
<td>59</td>
<td>55</td>
<td>51</td>
<td>47</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>73</td>
<td>72</td>
<td>69</td>
<td>66</td>
<td>63</td>
<td>60</td>
<td>57</td>
<td>54</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>56</td>
<td>53</td>
<td>49</td>
<td>46</td>
<td>43</td>
<td>39</td>
<td>36</td>
<td>33</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>58</td>
<td>56</td>
<td>53</td>
<td>50</td>
<td>48</td>
<td>45</td>
<td>43</td>
<td>41</td>
<td>39</td>
</tr>
</tbody>
</table>

Maximum likelihood estimates are obtained by solving the likelihood equations using iterative procedures. The probabilities have to be computed at every stage of the iteration. The method is very cumbersome. Therefore, we will investigate the minimum chi-square method of estimation.

Let $S = (S_1, S_2, \ldots, S_m)$, be consistent asymptotically normal estimator of $\phi = (\phi_1, \phi_2, \ldots, \phi_m)$ where $\phi$'s are functions of the unknown parameter $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$ and $m \geq k$. Let $\Sigma(\theta)$ be the covariance matrix of $S$, and $\hat{\Sigma}(\phi)$ be any consistent estimate of $\Sigma(\phi)$. Assume that the matrix $\Sigma(\phi)$ is positive definite and that the matrix $\left(\frac{\partial^2 \phi}{\partial \theta_j \partial \phi_i}\right)$ is of rank $k$.

Barankin and Gurland [1951] show that the estimator obtained by minimizing

$$\theta = (S - \phi)^{-1} \Sigma^{-1}(\phi)(S - \phi), \tag{7.1}$$

is asymptotically the best estimator in the class of estimators which are functions of $S$. The method of estimation is referred to as the minimum chi-square method based on $S$. The asymptotic covariance matrix of $\theta$ can be easily shown to be

$$V(\hat{\theta}) = \left[ \frac{\partial (\phi_1, \phi_2, \ldots, \phi_m)}{\partial (\theta_1, \ldots, \theta_k)} \right]^\prime \Sigma^{-1}(\phi) \left[ \frac{\partial (\phi_1, \phi_2, \ldots, \phi_m)}{\partial (\theta_1, \ldots, \theta_k)} \right]^{-1}, \tag{7.2}$$

where

$$\frac{\partial (\phi_1, \phi_2, \ldots, \phi_m)}{\partial (\phi_1, \phi_2, \ldots, \phi_k)} = \left( \frac{\partial \phi_i}{\partial \theta_j} \right).$$

The properties of the minimum $\chi^2$ estimators have been studied by various authors (e.g. Barankin and Gurland [1951], Ferguson [1958]). This method and its variations have been applied to the estimation of
contagious distributions by several authors including Katti and Gurland [1962], and Hinz and Gurland [1967]. The following theorem gives an interesting property of this estimator.

**Theorem 7.1** Let $S_m = (s_1, s_2, \ldots, s_m)$ and $S_{m+1} = (s_1, s_2, \ldots, s_m, s_{m+1})$. Denote by $\hat{\theta}_m$ and $\hat{\theta}_{m+1}$ the estimators of $\theta$ based on $S_m$ and $S_{m+1}$ respectively. Then

$$\text{Efficiency of } \hat{\theta}_m < \text{Efficiency of } \hat{\theta}_{m+1}$$

**Proof:**

The proof is omitted for brevity.

This theorem enables us to form a sequence of estimators $(\hat{\theta}_m)$ with increasing efficiency, by using more and more statistics. Depending on the computational difficulties, one may pick different number of statistics for different problems.

A special case of minimum chi-square method arises when the number of statistics equals number of parameters i.e. $m = k$. In this case the minimum $\chi^2$ estimators are the solutions of the equations

$$s_i = \phi_i \quad i = 1, 2, \ldots, k.$$  

The method of moments, the method of zero frequency and moments etc. which have been used in the estimation of the parameters of contagious distributions are thus the special cases of the minimum chi-square method.

**7.2 Application to the l. z. P. distribution**

Let $f_0, f_1, f_2, \ldots$ be frequencies in a sample of size $n$ for the l. z. P. distribution with parameters $\theta$, $p$ and $\lambda$. Consider the following statistics:
\[ s_1 = f_0/n \]
\[ s_i = \sum_{j=1}^{\infty} \frac{a_i(j)f_j}{n} \quad i = 2,3,\ldots,m, \quad (7.3) \]

where the \( a_i(j) \)'s are known constants. As an example if \( a_i(j) = i(i-1)\ldots(i-j+1), \) then \( s_i \) is the sample factorial moment. Now

\[ E(s_1) = \theta. \]

Let

\[ E(s_i) = \sum_{j=1}^{\infty} a_i(j)P(j) = \phi_i \quad i = 2,3,\ldots,m, \quad (7.4) \]

where \( P(i) \)'s are the probabilities of the l. z. P. distribution

It can be easily shown that

\[ \text{Cov}(s_1,s_j) = -\frac{\theta \phi_i}{n} \quad j = 2,3,\ldots,m. \quad (7.5) \]

and,

\[ \text{Cov}(s_i,s_j) = \frac{b_{ij}}{n} \phi_i \phi_j \quad i,j = 2,3,\ldots,m. \quad (7.6) \]

where

\[ b_{ij} = \sum_{k=1}^{\infty} a_i(k)a_j(k)P(k) \]

Now, by (6.14)

\[ P(i) = (1-\theta)P^*(i), \]

where \( P^*(i) \)'s are the probabilities in the l. z. P. t. distribution with parameters \( \lambda \) and \( p. \)
Let
\[ \phi_i^* = \sum_{j=1}^{\infty} a_i(j) P^*(j), \quad i = 2, 3, \ldots, n \]
and
\[ b'_{ij} = \sum_{k=1}^{\infty} a_i(k) a_j(k) P^*(k), \quad i, j = 2, 3, \ldots, m. \]

Then,
\[ \phi_i = (1-\theta) \phi_i^*, \quad i = 2, 3, \ldots, m, \quad (7.7) \]
and,
\[ b'_{ij} = (1-\theta) b'_{ij}, \quad i, j = 2, 3, \ldots, m. \quad (7.8) \]

This leads to
\[ \text{Cov}(s_i, s_j) = (1-\theta) \frac{b'_{ij} - \phi_i^* \phi_j^*}{n} + \theta (1-\theta) \frac{\phi_i^* \phi_j^*}{n} \]
\[ i, j = 2, 3, \ldots, m \quad (7.9) \]
and
\[ \text{Cov}(s_i, s_j) = -(1-\theta) \frac{\theta \phi_j^*}{n} \quad j = 2, 3, \ldots, m. \quad (7.10) \]

With the above notation, the covariance matrix of \( S = (s_1, \ldots, s_m) \) is
\[ \Sigma(\phi) = \begin{pmatrix} \frac{\theta (1-\theta)}{n} & \frac{-(1-\theta) \phi_1^*}{n} \\ \frac{-(1-\theta) \phi_1^*}{n} & \frac{\phi_1^* (1-\theta) \Sigma_1 + \theta (1-\theta) \phi_1^* \phi_1^*}{n} \end{pmatrix} \quad (7.11) \]
where \( \phi_1^* = (\phi_2^*, \phi_3^*, \ldots, \phi_m^*) \) is a \((m-1)\times1\) matrix and \( \Sigma_1 = \frac{b'_{ij} - \phi_i^* \phi_j^*}{n} \) is a \((m-1)\times(m-1)\) matrix. This matrix can be inverted to give:
$$\Sigma^{-1}(\phi) = \begin{pmatrix} \frac{n}{\theta(1-\theta)} + \frac{\phi'^1}{1-\theta} & \frac{\phi'^1}{1-\theta} \\ -1 & -1 \\ \Sigma_1 \phi_1 & \Sigma_1 \phi_1 \\ 1-\theta & 1-\theta \end{pmatrix}$$

(7.12)

**Theorem 7.2**

Let $\hat{\theta}$ be the minimum chisquare estimator of $\theta$ based on

$S = (s_1, s_2, \ldots, s_m)$. Then the generalized variance

$$V(\hat{\theta}) = n \frac{\theta}{1-\theta} \left| \frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\lambda, p)} \right|^{-1} \Sigma^{-1}(\phi) \left| \frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\lambda, p)} \right|^{-1}.$$  

(7.13)

**Proof:**

By (7.2),

$$V(\hat{\theta}) = \left( \frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\theta, \lambda, p)} \right)^{-1} \Sigma^{-1}(\phi) \left( \frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\theta, \lambda, p)} \right)^{-1}.$$  

Now,

$$\frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\theta, \lambda, p)} = \begin{pmatrix} 1 & 0 \\ -\phi'_1 & 1-\theta \end{pmatrix} \begin{pmatrix} 0 & \partial (\phi'_1, \phi'_2, \ldots, \phi'_m) \\ \partial (\lambda, p) & \partial (\lambda, p) \end{pmatrix}.$$  

Therefore,

$$V(\hat{\theta}) = \begin{pmatrix} \frac{n}{\theta(1-\theta)} \\ 0 \\ 0 \end{pmatrix} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right)^{-1} \left( \begin{pmatrix} \frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\lambda, p)} \\ \partial (\lambda, p) \end{pmatrix} \right)^{-1} \Sigma^{-1}(\phi) \left( \begin{pmatrix} \frac{\partial (\phi'_1, \phi'_2, \ldots, \phi'_m)}{\partial (\lambda, p)} \\ \partial (\lambda, p) \end{pmatrix} \right)^{-1}.$$  

(7.14)

The theorem follows from this step.
Corollary 7.3 The efficiency of \( \hat{\Theta} \) is a function of \( \lambda \) and \( p \) only.

Proof:

By definition,

\[
\text{Efficiency of } \hat{\Theta} = \frac{1}{\text{generalized variance} \times \text{information determinant}}.
\]

On substituting for the value of the information determinant from (6.13) and for the expression for generalized variance from (7.13) we get

\[
\text{Efficiency of } \hat{\Theta} = \frac{\det \left( \frac{\partial (\phi_2^*, \phi_3^*, \ldots, \phi_m^*)}{\partial (\lambda, p)} \right)}{n^2 |I(p, \lambda)|}. \quad (7.15)
\]

Since the ratio does not involve \( \Theta \), the corollary follows.

Let \( S_1 = \left( f_0, \frac{S_3}{S_2}, \frac{S_4}{S_3}, \ldots, \frac{S_m}{S_{m-1}} \right) \). Let \( \hat{\Theta}_1 \) denote the estimator of \( \Theta \) based on \( S_1 \). Therefore we have the following corollary.

Corollary 7.4

The efficiency of \( \hat{\Theta}_1 \) is a function of \( \lambda \) and \( p \) only.

Proof:

Let \( \psi_i = \frac{\phi_{i+1}}{\phi_i} = \frac{\phi_i^*}{\phi_i} \) \( i = 1, 2, \ldots, m-1 \). It can be easily seen that large sample Covariance matrix \( V(S_1) \) of \( S_1 \) is

\[
V(S_1) = \begin{pmatrix}
\frac{\Theta (1-\Theta)}{n} & 0 \\
0 & \frac{\Theta (1-\Theta)}{n} \\
\frac{\Theta (1-\Theta)}{n} & \frac{\Theta (1-\Theta)}{n}
\end{pmatrix} \quad (7.16)
\]
Therefore,

\[ V(\hat{\theta}_1) = n \frac{\theta}{1-\theta} \left| \frac{\partial (\psi_1, \psi_2, \ldots, \psi_{m-1})}{\partial (\lambda, p)} \frac{\partial (\psi_1, \ldots, \psi_{m-1})}{\partial (\phi_1, \phi_2, \ldots, \phi_m)} \right|^{-1} \]

\[ \left. \frac{1}{\partial (\psi_1, \ldots, \psi_{m-1}) \frac{\partial (\psi_1, \ldots, \psi_{m-1})}{\partial (\lambda, p)} \right|^{-1} \]

(7.17)

The determinant on the right hand side does not involve \( \theta \). Therefore, the corollary follows easily from the definition of efficiency.

Now suppose \( S_2 = (s_2^\prime, s_3^\prime, \ldots, s_m^\prime) \) have the same definition as (7.3); but the \( f \)'s are sample frequencies in a sample from the l. z. p. t. distribution with parameters \( \lambda \) and \( p \). The following can be easily verified:

\[ E(s_i^\prime) = \phi_i^\prime \]

\[ \text{Cov}(s_i^\prime, s_j^\prime) = \frac{b_{ij}^\prime - \phi_i^\prime \phi_j^\prime}{n}, \quad j, i = 2, 3, \ldots, m \]

and

\[ \text{Cov}(S_2) = \Sigma_1. \]

Let \( \hat{\theta}_2 \) denote the minimum \( \chi^2 \) estimator of \( \lambda, p \) based on \( S_2 \).

**Theorem 7.5**

Efficiency of \( \hat{\theta}_2 \) = Efficiency of \( \hat{\theta} \).

**Proof:**

It can be easily seen that
The theorem follows on comparison of (7.18) with (7.15).

This theorem implies that if the estimator for \( \theta \) is \( f_0/n \), then to study the different minimum \( \chi^2 \) estimate for \((\theta, p, \lambda)\) we need to study only the estimators of \((\lambda, p)\) based on statistics calculated from the l. z. P. t. distribution. Since \( f_0/n \) is the unique minimum variance unbiased estimator for \( \theta \), it is conjectured that the minimum \( \chi^2 \) estimator for \((\theta, p, \lambda)\) based on \((f_0/n, s_1, s_2, \ldots, s_m)\) will be more efficient than the estimator based on \((s^*, s_1, s_2, \ldots, s_m)\), where \( s^* \) is any other statistic which is not a function of \( f_0 \) alone. We have not succeeded in proving it as of the date of the writing.

It should be noted that in deriving these results the particular form of the probabilities as it occurs for the l. z. P. distribution has not been used. Hence, the results hold for more general situations.
8. EFFICIENCY OF SOME MINIMUM CHI-SQUARE ESTIMATORS

8.1 Preliminary remarks

In this section the estimator for \( \theta \) will be \( \hat{\theta} = f_0/n \). Therefore, by theorems proved in Section 7 it is sufficient to study estimators of \( \lambda \) and \( p \) based on samples for the l. z. P. t. distribution. We will use the following basic set of statistics; sample frequencies \( f_1 \) and \( f_2 \) and sample factorial moments \( m_1(1) \), \( m_2(2) \), and \( m_3(3) \). In addition to the estimator based on all the five statistics, we will also study estimators based on all combinations of two, three and four of these statistics.

The estimators will be compared on the basis of their asymptotic generalized variance. Since our objective here is to compare estimators based on the subsets of the five statistics with each other, it is sufficient to evaluate the efficiency of the subset relative to the set of all the five statistics using the formula:

\[
\frac{\text{generalized variance of the estimator based on all the five statistics}}{\text{generalized variance of the estimator based on the subset}}
\]

By the property of the minimum chi-square method discussed in theorem 7.1, this ratio will be between 0 and 1 and thus adequately reflects the loss of information due to choosing the subset. In what follows we will tabulate the asymptotic efficiency of the estimator based on the five statistics and for estimators based on subsets of these, we will tabulate relative efficiency only. If a reader is interested in the efficiency of any subset he may compute it by using the formula:
efficiency as a subset = \text{relative efficiency of the subset} \times \text{efficiency of the whole set.}

Explicit expressions for the probabilities \( P(1) \) and \( P(2) \) are given in Section 2. The expression for the first six population factorial moments for the 1. z. P. t. with parameters \( p \) and \( \lambda \) are

\[
\begin{align*}
\mu(1) &= \lambda p / \ln(q-p \exp(-\lambda)), \\
\mu(2) &= \lambda^2 p(1+p) / \ln(q-p \exp(-\lambda)), \\
\mu(3) &= \lambda^3 p(1+p)(1+2p) / \ln(q-p \exp(-\lambda)), \\
\mu(4) &= \lambda^4 p(1+p)(1+6p+6p^2) / \ln(q-p \exp(-\lambda)), \\
\mu(5) &= \lambda^5 p(1+p)(1+14p+36p^2+24p^3) / \ln(q-p \exp(-\lambda)), \\
\mu(6) &= \lambda^6 p(1+p)(1+30p+150p^2+240p^3+120p^4) / \ln(q-p \exp(-\lambda)).
\end{align*}
\]

The expressions for the covariances of the statistics considered can easily be derived.

For ready reference in later sections, we give the statistics used for each of the methods in a tabular form in table 2.

For the sake of brevity, we will use the phrase "method of \( S \)" instead of "minimum chi-square method based on the statistics \( S \)."

8.2 Efficiency of the Method of \( f_1, f_2, m(1), m(2) \) and \( m(3) \).

Lower and upper bounds for the efficiencies for values of \( \lambda, p \leq 5.0 \) are given in table 3.
<table>
<thead>
<tr>
<th>Number</th>
<th>Statistics Used</th>
<th>Number</th>
<th>Statistics Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$f_1, f_2$</td>
<td>15</td>
<td>$f_1^m(1), m(3)$</td>
</tr>
<tr>
<td>2</td>
<td>$f_1^m(1)$</td>
<td>16</td>
<td>$f_1^m(2), m(3)$</td>
</tr>
<tr>
<td>3</td>
<td>$f_1^m(2)$</td>
<td>17</td>
<td>$f_2^m(1), m(2)$</td>
</tr>
<tr>
<td>4</td>
<td>$f_1^m(3)$</td>
<td>18</td>
<td>$f_2^m(1), m(3)$</td>
</tr>
<tr>
<td>5</td>
<td>$f_2^m(1)$</td>
<td>19</td>
<td>$f_2^m(2), m(3)$</td>
</tr>
<tr>
<td>6</td>
<td>$f_2^m(2)$</td>
<td>20</td>
<td>$m(1), m(2), m(3)$</td>
</tr>
<tr>
<td>7</td>
<td>$f_2^m(3)$</td>
<td>21</td>
<td>$f_1^m(1), f_2^m(2)$</td>
</tr>
<tr>
<td>8</td>
<td>$m(1), m(2)$</td>
<td>22</td>
<td>$f_1^m(1), m(3)$</td>
</tr>
<tr>
<td>9</td>
<td>$m(1), m(3)$</td>
<td>23</td>
<td>$f_1^m(2), m(3)$</td>
</tr>
<tr>
<td>10</td>
<td>$m(2), m(3)$</td>
<td>24</td>
<td>$f_1^m(1), m(2), m(3)$</td>
</tr>
<tr>
<td>11</td>
<td>$f_1^m(1), m(1)$</td>
<td>25</td>
<td>$f_2^m(1), m(2), m(3)$</td>
</tr>
<tr>
<td>12</td>
<td>$f_1^m(2), m(1)$</td>
<td>26</td>
<td>$f_1^m(1), m(2), m(3)$</td>
</tr>
<tr>
<td>13</td>
<td>$f_1^m(2), m(3)$</td>
<td>27</td>
<td>$m(2), m(1), m(3)$</td>
</tr>
<tr>
<td>14</td>
<td>$f_1^m(1), m(2)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

35
TABLE 3

LOWER AND UPPER BOUNDS FOR THE EFFICIENCY FOR THE METHOD
$f_1, f_2, m(1), m(2)$ and $m(3)$.

(For each value of $p$, the lower bound is shown on the top line and the upper bound on the bottom line.)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$p$</th>
<th>.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>1.00*</td>
<td>.996</td>
<td>.994</td>
<td>.987</td>
<td>.974</td>
<td>.959</td>
<td>.942</td>
<td>.922</td>
<td>.901</td>
<td>.881</td>
<td>.859</td>
</tr>
<tr>
<td>1.00</td>
<td>.992</td>
<td>.981</td>
<td>.966</td>
<td>.948</td>
<td>.930</td>
<td>.910</td>
<td>.898</td>
<td>.874</td>
<td>.853</td>
<td>.833</td>
<td>.813</td>
</tr>
<tr>
<td>1.0</td>
<td>.995</td>
<td>.995</td>
<td>.997</td>
<td>.987</td>
<td>.969</td>
<td>.944</td>
<td>.912</td>
<td>.874</td>
<td>.833</td>
<td>.791</td>
<td>.748</td>
</tr>
<tr>
<td>1.00*</td>
<td>.992</td>
<td>.976</td>
<td>.951</td>
<td>.920</td>
<td>.884</td>
<td>.845</td>
<td>.804</td>
<td>.762</td>
<td>.722</td>
<td>.689</td>
<td>.652</td>
</tr>
<tr>
<td>1.5</td>
<td>.995</td>
<td>.995</td>
<td>.995</td>
<td>.997</td>
<td>.987</td>
<td>.969</td>
<td>.950</td>
<td>.913</td>
<td>.868</td>
<td>.817</td>
<td>.763</td>
</tr>
<tr>
<td>1.00*</td>
<td>.995</td>
<td>.987</td>
<td>.979</td>
<td>.969</td>
<td>.953</td>
<td>.938</td>
<td>.914</td>
<td>.863</td>
<td>.806</td>
<td>.744</td>
<td>.683</td>
</tr>
<tr>
<td>2.0</td>
<td>.996</td>
<td>.997</td>
<td>.998</td>
<td>.987</td>
<td>.982</td>
<td>.953</td>
<td>.937</td>
<td>.917</td>
<td>.865</td>
<td>.803</td>
<td>.737</td>
</tr>
<tr>
<td>1.00*</td>
<td>.998</td>
<td>.998</td>
<td>.997</td>
<td>.987</td>
<td>.968</td>
<td>.937</td>
<td>.891</td>
<td>.833</td>
<td>.768</td>
<td>.698</td>
<td>.627</td>
</tr>
<tr>
<td>2.5</td>
<td>.995</td>
<td>.995</td>
<td>1.00*</td>
<td>1.00*</td>
<td>.991</td>
<td>.958</td>
<td>.917</td>
<td>.865</td>
<td>.805</td>
<td>.737</td>
<td>.683</td>
</tr>
<tr>
<td>1.00*</td>
<td>.996</td>
<td>.986</td>
<td>.987</td>
<td>.968</td>
<td>.934</td>
<td>.887</td>
<td>.827</td>
<td>.758</td>
<td>.683</td>
<td>.608</td>
<td>.530</td>
</tr>
<tr>
<td>3.0</td>
<td>.995</td>
<td>.995</td>
<td>.986</td>
<td>.968</td>
<td>.932</td>
<td>.882</td>
<td>.820</td>
<td>.748</td>
<td>.670</td>
<td>.590</td>
<td>.512</td>
</tr>
<tr>
<td>1.00*</td>
<td>.993</td>
<td>.985</td>
<td>.987</td>
<td>.965</td>
<td>.923</td>
<td>.869</td>
<td>.805</td>
<td>.737</td>
<td>.664</td>
<td>.586</td>
<td>.508</td>
</tr>
<tr>
<td>3.5</td>
<td>.993</td>
<td>.993</td>
<td>1.00*</td>
<td>1.00*</td>
<td>.999</td>
<td>.972</td>
<td>.933</td>
<td>.877</td>
<td>.812</td>
<td>.741</td>
<td>.669</td>
</tr>
<tr>
<td>1.00*</td>
<td>.993</td>
<td>.984</td>
<td>.984</td>
<td>.963</td>
<td>.929</td>
<td>.878</td>
<td>.814</td>
<td>.738</td>
<td>.657</td>
<td>.575</td>
<td>.493</td>
</tr>
<tr>
<td>4.0</td>
<td>.992</td>
<td>.992</td>
<td>1.00*</td>
<td>1.00*</td>
<td>.991</td>
<td>.942</td>
<td>.888</td>
<td>.824</td>
<td>.752</td>
<td>.676</td>
<td>.590</td>
</tr>
<tr>
<td>1.00*</td>
<td>.992</td>
<td>.982</td>
<td>.982</td>
<td>.962</td>
<td>.926</td>
<td>.872</td>
<td>.805</td>
<td>.729</td>
<td>.645</td>
<td>.560</td>
<td>.476</td>
</tr>
<tr>
<td>4.5</td>
<td>.990</td>
<td>.990</td>
<td>.979</td>
<td>.958</td>
<td>.922</td>
<td>.866</td>
<td>.799</td>
<td>.717</td>
<td>.633</td>
<td>.546</td>
<td>.462</td>
</tr>
<tr>
<td>1.00*</td>
<td>.990</td>
<td>.979</td>
<td>.958</td>
<td>.922</td>
<td>.866</td>
<td>.799</td>
<td>.717</td>
<td>.633</td>
<td>.546</td>
<td>.462</td>
<td>.378</td>
</tr>
</tbody>
</table>

* Upper bound for efficiency exceeds 1.000. Hence the computed upper bound is not given.
The efficiency of the method exceeds .92 throughout the region \( \lambda \leq 2.5 \) and \( p \leq 5.00 \) of the tabulation. For problems where one has reason to believe that \( \lambda \) and \( p \) are in this region, this high efficiency might indeed justify using this simpler method in comparison with the likelihood estimators. Further justification for use of this method comes from the following considerations. (a) We have found from experience with empirical data that the likelihood function is very flat and the solution we get empirically may be very far from the maximum likelihood estimate. Similar difficulties arise in solving likelihood equations. (b) In actual examples, sample sizes are finite and the asymptotic efficiency given above is an approximation to the true small sample efficiency. In view of this working too hard towards additional efficiency may not be justified.

Actually, we have a feeling that inclusion of higher order moments which will increase the asymptotic efficiency might actually be causing a decrease in the small sample efficiency. A study of small sample efficiencies is currently underway at The Florida State University.

8.3 Relative Efficiency of Methods Based on Two Statistics.

Relative efficiency of using each of the ten pairs of the statistics \( f_1, f_2, m(1), m(2), m(3) \) have been tabulated. In calculating their relative efficiencies, since we do not need the value of the information determinant, the tabulation is in a larger region of the parameter space, namely \( \lambda \leq 9.5 \) and \( p \leq 9.5 \).

Table 4 gives the relative efficiency of the method of \( (f_1, m(1)) \). This method has by far the highest relative efficiency over a wide region of the parameter space.
\[
\begin{tabular}{lcccccccccc}
\hline
\multirow{2}{*}{p} & \multicolumn{10}{c}{\lambda} \\
& .5 & 1.5 & 2.5 & 3.5 & 4.5 & 5.5 & 6.5 & 7.5 & 8.5 & 9.5 \\
\hline
1.5 & .945 & .771 & .594 & .442 & .315 & .213 & .136 & .082 & .046 & .025 \\
2.5 & .960 & .803 & .631 & .482 & .385 & .257 & .175 & .113 & .068 & .038 \\
4.5 & .970 & .828 & .659 & .513 & .393 & .296 & .216 & .150 & .098 & .059 \\
5.5 & .972 & .833 & .666 & .520 & .402 & .306 & .227 & .162 & .108 & .067 \\
6.5 & .974 & .838 & .670 & .525 & .403 & .314 & .236 & .171 & .117 & .074 \\
7.5 & .975 & .841 & .674 & .529 & .413 & .320 & .243 & .179 & .124 & .080 \\
8.5 & .976 & .843 & .677 & .532 & .416 & .324 & .249 & .185 & .130 & .085 \\
9.5 & .977 & .845 & .679 & .534 & .419 & .328 & .253 & .190 & .136 & .090 \\
\hline
\end{tabular}
\]

As a summary of all the ten methods we give table 5. Here an entry against say, \( p = 1.5, \lambda = 5.5 \), 442(8) implies that the highest relative efficiency at \( p = 1.5, \lambda = 5.5 \) occurs when we use the method number 8 (cf. Table 2).

The method number 8, the method based on \( m(1) \) and \( m(2) \) has considerable advantage over the method of \( f_{(1)} \) and \( m(1) \) in the upper triangular part of the table. For large \( p \) and \( \lambda \) while the method \( (m(1), m(2)) \) is better than the method \( (f_{(1)}, m(1)) \), the efficiency of both of them is very poor. It is of interest to note that methods 1, 3, 4, 6, 7, 9 and 10 are inferior to methods 2, 5 and 8 at all but scattered points.
<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.5</th>
<th>1.5</th>
<th>2.5</th>
<th>3.5</th>
<th>4.5</th>
<th>5.5</th>
<th>6.5</th>
<th>7.5</th>
<th>8.5</th>
<th>9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.957(3)</td>
<td>0.846(8)</td>
<td>0.792(8)</td>
<td>0.753(8)</td>
<td>0.721(8)</td>
<td>0.692(8)</td>
<td>0.667(8)</td>
<td>0.644(8)</td>
<td>0.624(8)</td>
<td>0.607(8)</td>
</tr>
<tr>
<td>1.5</td>
<td>0.945(2)</td>
<td>0.771(2)</td>
<td>0.594(2)</td>
<td>0.452(8)</td>
<td>0.442(8)</td>
<td>0.445(8)</td>
<td>0.455(8)</td>
<td>0.469(8)</td>
<td>0.481(8)</td>
<td>0.490(8)</td>
</tr>
<tr>
<td>2.5</td>
<td>0.960(2)</td>
<td>0.803(2)</td>
<td>0.631(2)</td>
<td>0.482(2)</td>
<td>0.390(5)</td>
<td>0.450(5)</td>
<td>0.369(5)</td>
<td>0.390(3)</td>
<td>0.423(8)</td>
<td>0.451(8)</td>
</tr>
<tr>
<td>3.5</td>
<td>0.966(2)</td>
<td>0.819(2)</td>
<td>0.649(2)</td>
<td>0.502(2)</td>
<td>0.419(5)</td>
<td>0.451(5)</td>
<td>0.421(5)</td>
<td>0.351(5)</td>
<td>0.382(8)</td>
<td>0.423(8)</td>
</tr>
<tr>
<td>4.5</td>
<td>0.970(2)</td>
<td>0.828(2)</td>
<td>0.659(2)</td>
<td>0.513(2)</td>
<td>0.436(5)</td>
<td>0.477(5)</td>
<td>0.456(5)</td>
<td>0.393(5)</td>
<td>0.351(8)</td>
<td>0.400(8)</td>
</tr>
<tr>
<td>5.5</td>
<td>0.972(2)</td>
<td>0.833(2)</td>
<td>0.666(2)</td>
<td>0.520(2)</td>
<td>0.449(5)</td>
<td>0.495(5)</td>
<td>0.482(5)</td>
<td>0.425(5)</td>
<td>0.339(5)</td>
<td>0.381(8)</td>
</tr>
<tr>
<td>6.5</td>
<td>0.974(2)</td>
<td>0.838(2)</td>
<td>0.670(2)</td>
<td>0.525(2)</td>
<td>0.458(5)</td>
<td>0.508(5)</td>
<td>0.501(5)</td>
<td>0.450(5)</td>
<td>0.367(5)</td>
<td>0.364(8)</td>
</tr>
<tr>
<td>7.5</td>
<td>0.975(2)</td>
<td>0.841(2)</td>
<td>0.674(2)</td>
<td>0.529(2)</td>
<td>0.465(5)</td>
<td>0.519(5)</td>
<td>0.517(5)</td>
<td>0.470(5)</td>
<td>0.390(5)</td>
<td>0.350(8)</td>
</tr>
<tr>
<td>8.5</td>
<td>0.976(2)</td>
<td>0.843(2)</td>
<td>0.677(2)</td>
<td>0.532(2)</td>
<td>0.470(5)</td>
<td>0.527(5)</td>
<td>0.529(5)</td>
<td>0.487(5)</td>
<td>0.409(5)</td>
<td>0.334(8)</td>
</tr>
<tr>
<td>9.5</td>
<td>0.977(2)</td>
<td>0.845(2)</td>
<td>0.679(2)</td>
<td>0.534(2)</td>
<td>0.475(5)</td>
<td>0.534(5)</td>
<td>0.540(5)</td>
<td>0.501(5)</td>
<td>0.426(5)</td>
<td>0.329(5)</td>
</tr>
</tbody>
</table>
8.4 Relative Efficiency of the Methods Based on Three Statistics.

Relative efficiencies of all the ten methods based on these statistics have been tabulated. The method based on $f_1, f_2$ and $m_{(1)}$ has very high relative efficiency for $\lambda \leq 2.5$. The relative efficiency of this method is given in Table 6.

TABLE 6

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\lambda$</th>
<th>.5</th>
<th>1.5</th>
<th>2.5</th>
<th>3.5</th>
<th>4.5</th>
<th>5.5</th>
<th>6.5</th>
<th>7.5</th>
<th>8.5</th>
<th>9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td></td>
<td>.997</td>
<td>.949</td>
<td>.830</td>
<td>.668</td>
<td>.498</td>
<td>.346</td>
<td>.226</td>
<td>.139</td>
<td>.081</td>
<td>.045</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>1.000</td>
<td>.990</td>
<td>.947</td>
<td>.859</td>
<td>.733</td>
<td>.584</td>
<td>.433</td>
<td>.298</td>
<td>.191</td>
<td>.115</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>1.000</td>
<td>.995</td>
<td>.970</td>
<td>.912</td>
<td>.819</td>
<td>.696</td>
<td>.554</td>
<td>.410</td>
<td>.281</td>
<td>.178</td>
</tr>
<tr>
<td>3.5</td>
<td></td>
<td>1.000</td>
<td>.997</td>
<td>.980</td>
<td>.936</td>
<td>.862</td>
<td>.758</td>
<td>.630</td>
<td>.489</td>
<td>.350</td>
<td>.231</td>
</tr>
<tr>
<td>4.5</td>
<td></td>
<td>1.000</td>
<td>.998</td>
<td>.984</td>
<td>.949</td>
<td>.887</td>
<td>.797</td>
<td>.681</td>
<td>.546</td>
<td>.405</td>
<td>.275</td>
</tr>
<tr>
<td>5.5</td>
<td></td>
<td>1.000</td>
<td>.998</td>
<td>.987</td>
<td>.958</td>
<td>.904</td>
<td>.824</td>
<td>.718</td>
<td>.590</td>
<td>.450</td>
<td>.313</td>
</tr>
<tr>
<td>6.5</td>
<td></td>
<td>1.000</td>
<td>.999</td>
<td>.989</td>
<td>.963</td>
<td>.916</td>
<td>.844</td>
<td>.747</td>
<td>.624</td>
<td>.485</td>
<td>.345</td>
</tr>
<tr>
<td>7.5</td>
<td></td>
<td>1.000</td>
<td>.999</td>
<td>.990</td>
<td>.968</td>
<td>.925</td>
<td>.859</td>
<td>.769</td>
<td>.652</td>
<td>.515</td>
<td>.373</td>
</tr>
<tr>
<td>8.5</td>
<td></td>
<td>1.000</td>
<td>.999</td>
<td>.992</td>
<td>.971</td>
<td>.932</td>
<td>.871</td>
<td>.786</td>
<td>.675</td>
<td>.541</td>
<td>.398</td>
</tr>
<tr>
<td>9.5</td>
<td></td>
<td>1.000</td>
<td>.999</td>
<td>.992</td>
<td>.973</td>
<td>.937</td>
<td>.881</td>
<td>.810</td>
<td>.694</td>
<td>.563</td>
<td>.419</td>
</tr>
</tbody>
</table>

In Table 7 is a summary of all the ten methods. The entries in the table are to be interpreted as those in Table 5.

The method of $f_1, f_2$ and $m_{(1)}$ (method 11) has large relative efficiency in a large part of the parameter space tabulated. The method of $m_{(1)}, m_{(2)}, m_{(3)}$ (method 20) is far better than other methods considered here for the rectangular strip at $p = 0.5$. There is substantial region in the parameter space where this estimator is better than the rest. The method 17, based on $f_2, m_{(1)}$ and $m_{(2)}$ excels others in a small region on the parameter space of tabulation. Since there are two
The estimators are very simple to calculate, but the efficiency is very low as is seen from Table 10.

**TABLE 10**

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.5</th>
<th>1.5</th>
<th>2.5</th>
<th>3.5</th>
<th>4.5</th>
<th>5.5</th>
<th>6.5</th>
<th>7.5</th>
<th>8.5</th>
<th>9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>.515</td>
<td>.310</td>
<td>.212</td>
<td>.158</td>
<td>.127</td>
<td>.107</td>
<td>.093</td>
<td>.084</td>
<td>.077</td>
<td>.071</td>
</tr>
<tr>
<td>1.5</td>
<td>.304</td>
<td>.143</td>
<td>.085</td>
<td>.059</td>
<td>.045</td>
<td>.036</td>
<td>.031</td>
<td>.028</td>
<td>.026</td>
<td>.024</td>
</tr>
<tr>
<td>2.5</td>
<td>.198</td>
<td>.081</td>
<td>.044</td>
<td>.029</td>
<td>.022</td>
<td>.018</td>
<td>.015</td>
<td>.013</td>
<td>.013</td>
<td>.012</td>
</tr>
<tr>
<td>3.5</td>
<td>.139</td>
<td>.051</td>
<td>.027</td>
<td>.017</td>
<td>.013</td>
<td>.010</td>
<td>.009</td>
<td>.008</td>
<td>.007</td>
<td>.007</td>
</tr>
<tr>
<td>4.5</td>
<td>.101</td>
<td>.035</td>
<td>.018</td>
<td>.011</td>
<td>.008</td>
<td>.006</td>
<td>.006</td>
<td>.005</td>
<td>.005</td>
<td>.005</td>
</tr>
<tr>
<td>5.5</td>
<td>.077</td>
<td>.025</td>
<td>.013</td>
<td>.008</td>
<td>.006</td>
<td>.004</td>
<td>.004</td>
<td>.003</td>
<td>.003</td>
<td>.003</td>
</tr>
<tr>
<td>6.5</td>
<td>.060</td>
<td>.019</td>
<td>.009</td>
<td>.006</td>
<td>.004</td>
<td>.003</td>
<td>.003</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
</tr>
<tr>
<td>7.5</td>
<td>.048</td>
<td>.015</td>
<td>.007</td>
<td>.004</td>
<td>.003</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td>8.5</td>
<td>.039</td>
<td>.012</td>
<td>.006</td>
<td>.003</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td>9.5</td>
<td>.033</td>
<td>.009</td>
<td>.005</td>
<td>.003</td>
<td>.002</td>
<td>.002</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
</tbody>
</table>

Due to the low efficiency, the use of this estimator has to be restricted to preliminary experimentation where one is looking for a quick, simple method for the sole purpose of guiding one towards more careful experimentation.
9. CONCLUSIONS

The problem of estimating the parameters is simplified by changing the parameters from $p_1$, $p$ and $\lambda$ to $\theta$, $p$, and $\lambda$. The proportion of zeros in the sample, $f_0/n$ has the property of being the unique minimum variance unbiased estimator of the parameter $\theta$. The expression for the information determinant, when the parameters are $\theta$, $p$ and $\lambda$, is much simpler than when the parameters are $p_1$, $p$ and $\lambda$.

When the method of estimation is minimum chi-square and $f_0/n$ is one of the statistics used, the efficiency of the estimator does not involve $\theta$. The efficiency is equal to the efficiency of estimating $\lambda$ and $p$ using the same set of statistics from the l. z. P. t. samples, and excluding $f_0/n$.

The method of $f_1$, $f_2$, $m(1)$, $m(2)$, $m(3)$ has fairly high efficiency throughout the tabulated space. In the region $\lambda \leq 2.5$, the efficiency exceeds .92 and in the region $\lambda \leq 3.5$ the efficiency exceeds .8. Therefore, in situations where the value of $\lambda$ is known to be in the region $\lambda \leq 2.5$, this method may be profitably used in preference to the more complicated maximum likelihood method. By dropping $m(3)$ from the set of statistics, the loss in efficiency is very little when $\lambda \leq 2.5$.

The method of $f_2$, $m(1)$, $m(2)$, $m(3)$ has higher efficiency in the region in which the method of $f_1$, $f_2$, $m(1)$ and $m(2)$ has relatively lower efficiency. Among the methods based on three statistics the method of $f_1$, $f_2$ and $m(1)$ has high efficiency in a large region of the parameter space; the method of $m(1)$, $m(2)$, $m(3)$ is better for $p = .5$ and $\lambda$ and $p$ are both large. Of the methods based on two statistics, the method $f_1$, $m(1)$ is better in a large part of the parameter space, but the efficiency is small.
ACKNOWLEDGMENTS

This study was supported by The United States Department of Agriculture Grant 12-14-100-9183(20). The numerical computations were performed using the CDC 6400 at The Florida State University, supported by grant GJ 367 from the National Science Foundation.
REFERENCES


____________ [1969]. The Log-Zero-Poisson Distribution. (Accepted for publication in Biometrics).