The Poisson Distribution

by

Raymond Sproule

A Thesis submitted to the faculty of Graduate Studies
and Research in partial fulfillment of the
requirements for the Degree of Master
of Science

Department of Mathematics,
McGill University,
Montreal.

April 1963

CONTENTS

Chapter One

Introduction

In 1837 a paper by Poisson [61] was published in France, establishing the Poisson distribution. Since then a great deal of research has been done on this distribution and the distributions stemming from it. It enjoys remarkable application in medecine, psychology, genetics, biology, physics, economics and other areas. This work is an attempt to present, in an organized manner, the existing knowledge of the Poisson distribution, and so, is basicly expository in nature, and definitely not original. Complete references are given throughout.

The Poisson distribution and four generalized Poisson distributions are first introduced and various properties are developed, properties which are used in the later chapters on estimation and hypothesis testing.

A detailed account of point estimates of the parameter λ of the complete Poisson distribution occupies chapter 4. The theory has been developed in this area to such an extent that the uniformly minimum variance unbiased (U. M. V. U.) estimate of a wide class of functions of λ, $g(\lambda)$, has been obtained. The chapter closes with a brief treatment of the Poisson process.

Truncated and censored Poisson distributions are examined with respect to point estimation of parameters. A great deal of papers have been written about the truncated Poisson distribution and only a few about the censored Poisson distribution. A special emphasis lies on the case of truncation away from the zero point.

The generalized Poisson distributions, the Pascal, the Neyman Type A, the Poisson v Binomial and the Poisson v Pascal, were developed to provide better fits to data where simpler distributions, such as the Poisson proved unsatisfactory. Chapters 7, 8, 9 and 10 deal with estimation of the parameters of these distributions, with major emphasis on maximum likelihood estimation. The resulting estimation procedures are generally long and tedious, however, tables have been developed to save time and labour.

In chapter 11, two basic types of confidence intervals for the parameter λ of a complete Poisson distribution, randomized and non-randomized, are discussed. The subject has been dealt with in detail and references are given to the excellent tables available. A short treatment of the Poisson process due to Birnbaum [4], [5] concludes the chapter.

In chapter 12, goodness of fit and homogeneity tests involving the Poisson distribution, both complete and truncated away from zero are discussed and compared. A short treatment of the Poisson process is also given. The problem of testing hypotheses and combining tests for discrete distributions in general is examined in sections 12.6 and 12.7.

The final chapter presents a few ideas for research topics with the major emphasis on the truncated Poisson distribution.

Chapter Two

The Poisson Distribution

2.1 Introduction

In this chapter a formal definition of the Poisson distribution is given and various properties are developed. Most of these properties are used in later chapters but not all. The simple structure of the Poisson distribution is evident and is one of the reasons why it has widespread application. The distribution appears in a great variety of situations, often as a limiting approximation to a much more complicated distribution. Examples of this are given in this chapter, and even though some are not referred to in later chapters, they are sufficiently intriguing so as to earn a place here.

2.2 The Poisson Distribution

Let X denote a discrete random variable taking on values x belonging to the infinite sample space consisting of the non-negative integers, and having probability density function (p. d. f.)

$$p(x;\lambda) = \begin{cases} e^{-\lambda}\dfrac{\lambda^{x}}{x!} & x = 0,1,2,\cdots \\ 0 & \text{elsewhere} \end{cases}$$

where $\lambda > 0$ is a real number. Then X is said to have a "Poisson distribution" with parameter λ. We can easily see that the two requirements for a function of a random variable to be a p. d. f. are satisfied, by noting that

$$1)\quad e^{-\lambda}\frac{\lambda^x}{x!} > 0 \qquad\qquad x = 0,1,2,\cdots$$

$$2)\quad \sum_{x=0}^{\infty} p(x;\lambda) = 1$$

The cumulative distribution function (c. d. f.) of X is

$$F(x;\lambda) = \sum_{r=0}^{x} p(r;\lambda) \qquad\qquad x = 0,1,2,\cdots$$

Both $p(x;\lambda)$ and $F(x;\lambda)$ have been tabulated by Molina [51] for λ ranging from 0.001 to 100. The Poisson distribution has also been tabulated by Soper [72] , Whitaker [84] and Kitagawa [45] .

2.3 Structural Properties

The p. d. f. $p(x;\lambda)$ obeys the simple recurrence relation

$$p(x+1;\lambda) = \frac{\lambda}{(x+1)}\, p(x;\lambda) \qquad (1)$$

From (1) we obtain the mode x_0 of the Poisson distribution. For, note that

$$p(x_0+1;\lambda) \leq p(x_0;\lambda)$$

and $\quad p(x_0 - 1; \lambda) \leq p(x_0; \lambda)$

lead to the inequalities, $\lambda - 1 \leq x_0$ and $x_0 \leq \lambda$, respectively. Since x_0 must be an integer the inequalities combine to give $x_0 = [\lambda]$ where $[\lambda]$ denotes the greatest integer less than or equal to λ. The mode x_0 is unique except for the case where λ is an integer. In that case x_0 takes on the two consecutive values $\lambda - 1$ and λ.

Let X be a Poisson random variable with p. d. f., $p(x;\lambda)$, then for a positive integer n

$$\sum_{x=0}^{n-1} p(x;\lambda) = \frac{1}{(n-1)!}\int_{\lambda}^{\infty} e^{-t} t^{n-1}\,dt \qquad (2)$$

$$= 1 - \frac{1}{(n-1)!}\int_{0}^{\lambda} e^{-t} t^{n-1}\,dt$$

This can be easily verified by integrating the right side of (2) by parts and then summing. If T has a gamma distribution with parameter n, and χ^2_{2n} has a chi-square distribution with 2n degrees of freedom, then either side of (2) equals $P(T \geq \lambda)$ and also, $P(\chi^2_{2n} \geq 2\lambda)$. Since there are extensive gamma and chi-square tables available, (2) permits one to obtain Poisson sums quickly. This property is particularly useful when investigating confidence intervals and testing hypotheses.

Let X and $Y-X$ be two Poisson random variables having p. d. f.'s $p(x;\lambda_1)$ and $p(y-x;\lambda_2)$ respectively. Then we have

$$\sum_{x=0}^{y} p(x;\lambda_1)\,p(y-x;\lambda_2) = \sum_{x=0}^{y} e^{-(\lambda_1+\lambda_2)} \frac{\lambda_2^{y}}{y!}\binom{y}{x}\left(\frac{\lambda_1}{\lambda_2}\right)^{x}$$

$$= e^{-(\lambda_1+\lambda_2)} \frac{\lambda_2^{y}}{y!} \sum_{x=0}^{y}\binom{y}{x}\left(\frac{\lambda_1}{\lambda_2}\right)^{x}$$

$$= e^{-(\lambda_1+\lambda_2)} \frac{\lambda_2^{y}}{y!}\left(1+\frac{\lambda_1}{\lambda_2}\right)^{y}$$

$$= p(y;\lambda_1+\lambda_2)$$

We conclude this section of the chapter by stating an inequality and an equality involving the Poisson distribution. The proofs are long, yet not difficult, and are omitted. Teicher [18] proves the interesting inequality

$$\sum_{r=0}^{[\lambda]} \frac{\lambda^{r}}{r!} > \begin{cases} e^{\lambda-1} & \text{for all } \lambda>0 \\ \dfrac{e^{\lambda}}{2} & \text{for all integral } \lambda>0 \end{cases}$$

Crow and Gardner [19], in the course of developing confidence intervals for the parameter λ of a Poisson distribution, use the following equality. Let λ_{r_1, r_2}, where $r_1 \leq r_2$ and both are fixed, be the value of λ that maximizes

$$\sum_{x=r_1}^{r_2} p(x;\lambda)$$

treated as a function of λ. It is easy to show that

$$\lambda_{r_1,r_2} = \left[r_1(r_1+1) \cdots r_2 \right]^{\frac{1}{r_2-r_1+1}}$$

Crow and Gardner [19] prove that for fixed $k=0,1,2,\ldots$

$$\operatorname*{Max}_{r=0,1,2,\cdots} \sum_{x=r}^{r+k} p(x;\lambda) = \sum_{x=r_1}^{r_2} p(x;\lambda) \tag{3}$$

when $\lambda_{r_1,r_1+k} \leq \lambda \leq \lambda_{r_1+1,r_1+k+1}$. This is true for $r_1 = 0,1,2,\ldots$.

2.4 Generating Functions

If X is a Poisson random variable with parameter λ, then X has

1) characteristic function

$$\phi(t) = E\left\{e^{itX}\right\} = \sum_{x=0}^{\infty} e^{itx} p(x;\lambda) = e^{\lambda(e^{it}-1)} \tag{4}$$

where $i = \sqrt{-1}$.

2) moment generating function, (m. g. f.),

$$\psi(t) = E\left\{e^{tX}\right\} = \phi\left(\frac{t}{i}\right) = e^{\lambda(e^{t}-1)}$$

3) probability generating function, (p. g. f.),

$$\theta(t) = \psi(\ln t) = E\left\{t^X\right\} = e^{\lambda(t-1)} \tag{5}$$

4) factorial moment generating function

$$u(t) = E\left\{(1+t)^X\right\} = e^{\lambda t} \tag{6}$$

5) cumulant generating function

$$v(t) = \ln \phi(t) = \lambda\left(e^{it} - 1\right) \tag{7}$$

The r th central moment of X or the moment about the origin of X denoted by μ'_r, is the coefficient of $(it)^r/r!$ in the expansion of (4) and may also be written as

$$\mu'_r = E\left\{X^r\right\} = \sum_{x=0}^{\infty} x^r\, p(x;\lambda) \tag{8}$$

The r th moment about the mean of X, denoted by μ_r, may be written as

$$\mu_r = E\left\{(X-\lambda)^r\right\} = \sum_{x=0}^{\infty} (x-\lambda)^r\, p(x;\lambda) \tag{9}$$

The r th factorial moment of X, denoted by $\mu_{[r]}$, is the coefficient of $t^r/r!$ in the expansion of (6), and may also be written as

$$\mu_{[r]} = E\left\{X^{[r]}\right\} = \sum_{x=0}^{\infty} x^{[r]}\, p(x;\lambda) = \lambda^r$$

where $X^{[r]} = X(X-1)\ldots(X-r+1)$. The r th cumulant (or semi - invariant) of X, denoted by k_r, is the coefficient of $(it)^r/r!$ in the expansion of (7), and so for all r is λ.

More specifically, we may use (8) and (9) to obtain the following

$$\mu_1' = \kappa_1 = \lambda$$

$$\mu_2' = \kappa_2 + \kappa_1 = \lambda + \lambda^2$$

$$\mu_3' = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = \lambda + 3\lambda^2 + \lambda^3$$

$$\vdots$$

and

$$\mu_1 = \lambda$$

$$\mu_2 = \lambda$$

$$\mu_3 = \lambda$$

$$\mu_4 = \lambda\,(1+3\lambda)$$

$$\mu_5 = \lambda\,(1+10\lambda)$$

$$\mu_6 = \lambda\,(1+25\lambda+15\lambda^2)$$

$$\vdots$$

It should be noted that the mean, defined by $E(X)$, is λ, and that the variance, defined by $\sigma^2 = E\{(X-\lambda)^2\}$, is also λ.

The following relations involving the moments about the mean, μ_r, and the central moments, μ_r', are given. Their proofs are straightforward and are omitted to save space.

$$\mu_r' = \lambda \sum_{j=0}^{r-1} \binom{r-1}{j} \mu_j'$$

$$\mu_r = \lambda \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j$$

$$\mu_{r+1}' = \lambda\,\mu_r' + \lambda\,\frac{d\mu_r'}{d\lambda}$$

$$\mu_{r+1} = r\lambda\,\mu_{r-1} + \lambda\,\frac{d\mu_r}{d\lambda}$$

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} (-\lambda)^{r-j} \mu'_j$$

A very useful property of the Poisson distribution is the additivity property. Let $X_1, \ldots, X_n$ be n independent random variables having parameters $\lambda_1, \ldots, \lambda_n$, respectively. Then from (4) we have that the characteristic function of the statistic $\sum_{j=1}^{n} X_j$ is

$$E\left\{e^{it\sum_{j=1}^{n} X_j}\right\} = \prod_{j=1}^{n} e^{\lambda_j(e^{it}-1)} = e^{\sum_{j=1}^{n}\lambda_j(e^{it}-1)}$$

This is the characteristic function of a Poisson random variable with parameter $\sum_{j=1}^{n} \lambda_j$. Thus the sum of n independent Poisson random variables is again a Poisson random variable.

2.5 The Limit of the Binomial Distribution

The Poisson distribution is usually introduced in textbooks as a limit of the binomial distribution. Consider the binomial distribution having parameters n and p and probability generating function (p. g. f.), $(q + pz)^n$ where $p > 0$, $q > 0$ and $p + q = 1$. Then, in the limit as $p \to 0$ and $n \to \infty$ such that $np = \lambda$

where $\lambda > 0$ is a real number, we have

$$\begin{aligned}\lim (q+pz)^n &= \lim (1-p)^n \left(1 + \frac{pz}{1-p}\right)^n \\ &= \lim \left(1-\frac{\lambda}{n}\right)^n \left(1 + \frac{\lambda z}{n(1-p)}\right)^n \\ &= e^{\lambda(z-1)}\end{aligned}$$

From (5) we see that this is the p. g. f. of a Poisson random variable having parameter λ. Mainly due to this approximation, used when p is very small, the Poisson law is known to many statisticians as the law of small numbers or rare events. Raff [63] has made a study of the six best approximations to the cumulative binomial probability and classifies these into two groups. The simple, less accurate approximations consist of the normal, the arcsine and the Poisson, while the "advanced" or more accurate approximations are the normal Gram - Charlier, the Poisson Gram - Charlier and the Camp - Paulson. The Poisson Gram - Charlier approximation is

$$\sum_{r=0}^{x} p(r;np) + \frac{p}{2}(x-np)\, p(x;np) \qquad (10)$$

where $p(r;np)$ is the r th Poisson probability with parameter np. The cumulative binomial probability, approximated by (10), is

$$\sum_{r=0}^{x} \binom{n}{r} p^r (1-p)^{n-r}$$

Of the simple approximations the Poisson is best for $p < 0.075$ and the arcsine, for $p > 0.075$. Of the "advanced" approximations the Poisson Gram - Charlier is best for $p < 0.075$, while for larger p, the Camp - Paulson is best. It is interesting to note that both the Poisson and Poisson Gram - Charlier approximations are practically independent of the value of n.

2.6 The Poisson Process

The Poisson distribution may be derived from the concept of a Poisson process. Let $(0, t)$ and $(t, t+h)$ be two contiguous intervals of time, with h considered small. Then let $P_n(t)$ be the probability that exactly n changes of some physical nature occur during time interval t. The Poisson process is characterized by the postulates stated in Feller [25] : "Whatever the number of changes during $(0, t)$, the probability that during $(t, t+h)$ a change occurs is $\lambda h + o(h)$, and the probability that more than one change occurs is $o(h)$." Here λ is a positive constant. This completes the formal concept of a Poisson process. In language less mathematical, the probability of an event (change) depends only upon the length of the time interval, and not upon either the position of the interval in the range of time, or the past history of the preceding intervals. Now, the event (n changes occur during $(0, h + t)$) has probability $P_n(h+t)$ of occuring and can be realized in three mutually exclusive ways, namely,

1) n changes occur in $(0, t)$ while no changes occur in $(t, t+h)$

2) $n - 1$ changes occur in $(0, t)$ while 1 change occurs in $(t, t+h)$

and 3) $n - x$ changes occur in $(0, t)$ while $x \geq 2$ changes occur in $(t, t+h)$

The respective probabilities of these "ways" or events are

1) $P_n(t)\,(1 - \lambda h - o(h))$

2) $P_{n-1}(t)\,\lambda h$

and 3) $P_{n-x}(t)\; o(h)$

Thus

$$P_n(t+h) = P_n(t)\left(1 - \lambda h + o(h)\right) + P_{n-1}(t)\,\lambda h + P_{n-x}(t)\,o(h)$$

so that

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}$$

and in the limit as $h \to 0$ we obtain the differential equation

$$\frac{d\,P_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t)$$

Solving this equation for $n = 0$ and 1, and then using mathematical induction, leads to the solution

$$P_n(t) = e^{-\lambda t}\,\frac{(\lambda t)^n}{n!} \qquad (11)$$

Thus, for a fixed time interval t, the random variable n giving the number of changes or events during t, is a Poisson random variable with parameter λt. The constant λ is the mean number of changes or events per unit time. The number of automobile accidents at a certain corner, during a certain hour on a specified day of the week, and the number of chromosome interchanges induced by X-rays during a fixed time, are two examples of the Poisson random variable and the Poisson process. It is important to mention that although we have only referred to random events in time, this has been for the sake of simplicity. The same arguments apply for random events in space of one, two or three dimensions. Thus, for example, the number of blades of grass in a square foot of lawn, the number of stars in a large volume of space and the number of dents per foot on a very long thin rod, may be assumed to be Poisson random variables, and the process involved, a Poisson process. We will refer to the Poisson process discussed here as a " - Poisson process" and shall adopt Birnbaum [5]'s phrasing by referring to the "amount of time", or "the amount of space", collectively, as "the amount of observation".

Suppose that T is the amount of observation required for n events to occur in a λ - Poisson process. Then we shall show that $2\lambda T$ is a chi-square random variable with 2n degrees of freedom, denoted by χ^2_{2m}. Let U be a random variable defined as the "time" or observation required for two successive events, or, using a different wording, the "waiting time" between successive events, in a λ - Poisson process. Then U has the cumulative distribution function (c. d. f.)

$$P(U < u) = P(\text{"waiting time" is less than } u)$$

$$= P\begin{pmatrix}\text{number of events during the} \\ \text{observation } u \text{ is greater than zero}\end{pmatrix}$$

$$= \sum_{x=1}^{\infty} e^{-\lambda u} \frac{(\lambda u)^x}{x!}$$

$$= 1 - e^{-\lambda u}$$

Thus the probability density function (p. d. f.) of U is

$$g(u) = \lambda e^{-\lambda u} \qquad u \geq 0 \qquad (12)$$

The mean of U is $E\{U\} = 1/\lambda$ and shall be denoted as θ, so that $\theta = 1/\lambda$. Problems in statistical inference involving the parameter λ are simplified by the fact that λ occurs as a parameter in both the Poisson distribution (11) and the exponential distribution (12). The characteristic function of U is

$$E\left\{e^{itU}\right\} = \int_0^{\infty} \lambda e^{-\lambda u + itu} \, du = \frac{\lambda}{\lambda - it} \qquad (13)$$

where $\lambda > t > 0$. If t is to be the amount of observation for n successive events to occur, then T is the sum of n independent random variables each having p. d. f. (12), and so, characteristic function (13). Thus T has characteristic function $\left(\frac{\lambda}{\lambda - it}\right)^n$ and the random variable, $2\lambda T$, has characteristic function $\frac{1}{(1 - 2it)^n}$. Thus $2\lambda T$ is a chi-square random variable with $2n$ degrees of freedom. This fact is used by Birnbaum [5] to obtain confidence intervals for the parameter λ as well as to test various hypotheses involving Poisson processes.

2.7 Interesting Examples of the Poisson as a Limiting Distribution

Both the examples offered are directly from Feller [25]. Consider a Markov chain with states E_0, E_1, ... having transition probabilities

$$p_{rk} = \begin{cases} e^{-\lambda} \sum_{x=0}^{r} \binom{r}{x} p^x q^{r-x} \frac{\lambda^{k-x}}{(k-x)!} & x \leq k \\ 0 & x > k \end{cases}$$

where p and q are constant probabilities of "success" and "failure", respectively, with $p + q = 1$, and $\lambda > 0$. Define the higher transition probability $p_{rk}^{(n)}$ as

$$p_{rk}^{(n)} = P\left\{ \begin{array}{l} \text{the system at time } r+n \text{ is in the state } E_k \mid \text{at} \\ \text{time } r \text{ the system was in the state } E_r \end{array} \right\}$$

Then it can be shown that the "stationary" probability that the system is in the state E_k is

$$\lim_{n\to\infty} p_{rk}^{(n)} = e^{-\lambda/q} \frac{(\lambda/q)^k}{k!}$$

Thus k is a Poisson random variable with parameter λ/q.

Consider the problem of placing r balls in n cells with each arrangement equiprobable. We seek the probability that exactly x cells are empty. Feller [25] shows that

$$P\{\text{exactly } x \text{ cells are empty}\} = \binom{n}{x} \sum_{j=0}^{n-x} (-1)^j \binom{n-x}{j} \left(1 - \frac{x+j}{n}\right)^r$$

However, this result is only wieldy if n and r are small, so that for large r and n, approximations are in order. Let r and n both approach infinity such that $\lambda = ne^{-r/n}$ remains bounded. In this case

$$\lim_{\lambda, n \to \infty} P\{\cdots\} = e^{-\lambda} \frac{\lambda^x}{x!}$$

Thus for large r and n, the exact number of empty cells is approximately a Poisson random variable with parameter $\lambda = ne^{-r/n}$

In chapter three four discrete distributions closely related to the Poisson distribution are discussed. Each one, under a limiting condition, approaches the Poisson distribution.

2.8 Approximations Involving the Poisson Distribution

If X is a Poisson random variable having parameter λ we may consider, as a rough approximation, that X is a normal random variable with mean λ and variance λ. However, even with a correction for continuity, the approximation is too rough for most purposes.

If $X_1, \ldots, X_n$ are n independent Poisson random variables having parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively, then $Z = \sum_{j=1}^{n} X_j/n$ has characteristic function

$$
\begin{aligned}
E\left\{e^{itZ}\right\} &= E\left\{e^{it\sum_{j=1}^{n} X_j/n}\right\} \\
&= \prod_{j=1}^{n} e^{\lambda_j (e^{it/n}-1)} \\
&= e^{\sum_{j=1}^{n}\lambda_j (e^{it/n}-1)} \\
&= e^{\lambda_0 (e^{it/n}-1)} \qquad (14)
\end{aligned}
$$

where $\lambda_0 = \sum_{j=1}^{n} \lambda_j$. By differentiating (14) twice with respect to t we obtain λ_0/n and λ_0/n^2 for the mean and variance of Z, respectively. Now let us attempt to find an approximation for (14). Expanding the power of e in (14) we have

$$
\begin{aligned}
e^{\lambda_0\left[e^{it/n}-1\right]} &= e^{\lambda_0\left[\sum_{r=0}^{\infty} \frac{(it)^r}{r!\,n^r} - 1\right]} \\
&= e^{\lambda_0\left[\sum_{r=1}^{\infty} \frac{(it)^r}{r!\,n^r}\right]} \qquad (15)
\end{aligned}
$$

If we neglect all terms of order $r > 2$ in the power of e in (15) we obtain

$$
e^{\lambda_0\left[\frac{it}{n} - \frac{t^2}{2n^2}\right]}
$$

which is the characteristic function of a normal random variable having mean λ_0/n and variance λ_0/n^2. Thus we see that for large n the random variable Z is approximately a normal random variable. In most instances we are interested in the special case where all the parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ are equal, say of value λ. Then $\lambda_o = n\lambda$ and Z, now denoted by $\overline{X}$, has the approximate normal distribution with mean λ and variance λ/n. Generally speaking, this is a very useful approximation. It may be used to obtain confidence intervals for λ, and in hypotheses testing, as shall be shown in chapters 11 and 12.

We shall now show that the statistic $2\sqrt{\overline{X}}$ has an asymptotic normal distribution with mean $2\sqrt{\lambda}$ and variance $1/n$. The proof given here is patterned after a more general proof found in Wilks [85], P. 259-260. For convenience let $g(\overline{x}) = 2\sqrt{\overline{x}}$. Since $\lambda > 0$, there exists an interval containing λ, say I, such that for all $\overline{x} \in I$, the first derivative, $g'(\overline{x})$, exists. From the law of large numbers (Wilks [85], P. 108) we have that if an arbitrary $\epsilon > 0$ is given, there exists an n, say n_ϵ, such that for all $n > n_\epsilon$

$$P\left(\overline{x} \in I \text{ for all } n > n_\epsilon\right) > 1 - \epsilon \qquad (16)$$

For any $\overline{x} \in I$ we have, from a well-known theorem in calculus,

$$g(\overline{x}) = g(\lambda) + g'(x^*)(\overline{x} - \lambda)$$

where $|x^* - \lambda| \leq |\overline{x} - \lambda|$. More explicitly we have

$$2\sqrt{\overline{x}} = 2\sqrt{\lambda} + \frac{1}{\sqrt{x^*}}(\overline{x} - \lambda) \qquad (17)$$

Rewriting (17) gives

$$2\sqrt{n}\left(\sqrt{\bar{x}}-\sqrt{\lambda}\right)=\sqrt{\frac{n}{x^*}}\left(\bar{x}-\lambda\right) \tag{18}$$

From (16) we have that (18) holds with probability greater than $1-\epsilon$ for all $n > n_\epsilon$. Thus in the limit as $n \to \infty$ (18) holds with probability one. Now, the statistic $\sqrt{n}\,(\bar{X}-\lambda)$ converges in distribution to a normal random variable, say S, having mean 0 and variance λ. Also, from Wilks [85], $1/\sqrt{X^*}$ converges in probability to $1/\sqrt{\lambda}$. Thus

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}}{\sqrt{X^*}}(\bar{X}-\lambda) < \omega\right) = P\left(\frac{S}{\sqrt{\lambda}} < \omega\right)$$

so that

$$\lim_{n\to\infty} P\left(2\sqrt{n}\left(\sqrt{\bar{X}}-\sqrt{\lambda}\right) < \omega\right) = P\left(T < \omega\right)$$

where $T = S/\sqrt{\lambda}$ has a standard normal distribution. It now follows that $2\sqrt{n}\,(\sqrt{\bar{X}})$ has a limiting normal distribution with mean $2\sqrt{n\lambda}$ and variance 1. Thus $2\sqrt{\bar{X}}$ has a limiting normal distribution with mean $2\sqrt{\lambda}$ and variance $1/n$. The confidence intervals for λ obtained using this approximation are almost identical to those using the immediately preceding approximation, a not too surprising result.

Chapter Three

Generalized and Compound Poisson Distributions

3.1 Introduction

When the simple discrete distributions fail to fit biological data adequately, more complicated distributions, such as the generalized and compound Poisson distributions, need to be investigated. In this chapter four prominent distributions, the Poisson $\vee$ Binomial, the Poisson $\vee$ Pascal, the negative - binomial and the Neyman Type A, will be unveiled. Their definitions and a few basic properties will be given. This chapter is preparation for chapters 7, 8, 9 and 10 which deal with the estimation of the parameters.

3.2 Generalized and Compound Poisson Distributions

Let X_1 be a random variable having c. d. f. $F_1(x_1 | \lambda)$ where the parameter λ may be regarded as a random variable X_2 having c. d. f. $F_2(x_2)$. Then we let $X_1 \wedge X_2$ denote the random variable having c. d. f.

$$\int_{-\infty}^{+\infty} F(x_1 | c x_2)\, dF_2(x_2)$$

where c is a suitable arbitrary constant, and call it the "compound X_1 variable with respect to the compounder X_2 ".

Let X_1 and X_2 be random variables having p. g. f.'s $g_1(z)$ and $g_2(z)$ respectively. Then let $X_1 \vee X_2$ denote a random variable having p. g. f. given by $g_1\left\{g_2(z)\right\}$ and call it the "generalized X_1 variable with respect to the generalizer X_2".

Let X_1 and X_2 have c. d. f.'s $F_1(x_1|\alpha)$ and $F_2(x_2|\beta)$, respectively, where α and β are parameters. If for each α there exists a β , and for each β there exists an α , such that

$$F_1(x|\alpha) = F_2(x|\beta)$$

whatever the value of x, the random variables X_1 and X_2 are said to be "equivalent", and we write $X_1 \backsim X_2$. Gurland [33] proves the following theorem. If X_1 is a random variable with p. g. f. of the form $\left[h(z)\right]^{\lambda}$, and the parameter λ is regarded as a random variable X_2 having c. d. f. $F_2(x)$ and p. g. f. $g_2(z)$, then whatever be the random variable X_2, we have that $X_1 \wedge X_2 \backsim X_2 \vee X_1$. The proof is simple. From the definitions of compound and generalized distributions we have that the p. g. f. of $X_1 \wedge X_2$ for suitable c is

$$\int_{-\infty}^{\infty} \left[h(z)\right]^{cx} dF_2(x)$$

and that the p. g. f. of $X_2 \quad X_1$ is

$$\begin{aligned} g_2\left\{g_1(z)\right\} &= g_2\left\{\left[h(z)\right]^{\lambda}\right\} \\ &= \int_{-\infty}^{+\infty} \left[h(z)\right]^{\lambda x} dF_2(x) \end{aligned}$$

Thus the random variables $X_1 \wedge X_2$ and $X_2 \vee X_1$ are equivalent, and are equal when $c = \lambda$. Since the Poisson random variable has p. g. f. of the form required by the theorem, every compound Poisson distribution may be considered to be some generalized distribution.

Applying the definition of a generalized distribution to the Poisson distribution, we have that the generalized Poisson distribution has p. g. f. of the form

$$g(z) = e^{\lambda[h(z)-1]} \tag{1}$$

where $\lambda > 0$ and $h(z)$ is an arbitrary p. g. f. .
Suppose that

$$g(z) = \sum_{x=0}^{\infty} P_x z^x \tag{2}$$

and

$$h(z) = \sum_{x=0}^{\infty} \pi_x z^x \tag{3}$$

where P_x and π_x are the generalized Poisson and generalizer probabilities respectively. We would now like to establish the following important recurrence relation

$$P_{x+1} = \frac{\lambda}{(x+1)} \sum_{j=0}^{x} (x-j+1)\, \pi_{x-j+1}\, P_j \tag{4}$$

Differentiate (1) with respect to z to obtain

$$g^{(1)}(z) = \lambda\, g(z)\, h^{(1)}(z)$$

Then, using Leibnitz's formula we have

$$g^{(r+1)}(z) = \lambda \sum_{j=0}^{r} \binom{r}{j} h^{(r-j+1)}(z)\, g^{(j)}(z) \tag{5}$$

Detecting from (2) and (3) that

$$\frac{1}{(r+1)!}\, g^{(r+1)}(z)\Big|_{z=0} = P_{r+1} \tag{6}$$

$$\frac{1}{j!}\, g^{(j)}(z)\Big|_{z=0} = P_j \qquad (7)$$

and $$\frac{1}{(r-j+1)!}\, h^{(r-j+1)}(z)\Big|_{z=0} = \Pi_{r-j+1} \qquad (8)$$

setting $z = 0$, and substituting the results (6), (7) and (8) into equation (5), gives the desired result (4).

3.3 The Pascal Distribution

The Pascal distribution, better known as the negative binomial distribution, may be obtained by assuming that the parameter λ of a Poisson distribution has the gamma distribution represented by the p. d. f.

$$\frac{\beta^k \lambda^{k-1}}{\Gamma(k)} e^{-\beta\lambda}$$

Thus the Pascal distribution is a compound Poisson distribution and has p. d. f.

$$\Pi_x = \int_0^\infty e^{-\lambda c} \frac{(\lambda c)^x}{x!} \beta^k \frac{\lambda^{k-1} e^{-\beta\lambda}}{\Gamma(k)}\, d\lambda$$

$$= \frac{\beta^k c^x}{\Gamma(k)\, x!} \int_0^\infty e^{-\lambda(\beta+c)} \lambda^{x+k-1}\, d\lambda$$

$$= \frac{\beta^k c^x}{\Gamma(k)\, x!} \; \frac{\Gamma(x+k)}{(\beta+c)^{x+k}}$$

$$= \binom{x+k-1}{x} p_1^k q_1^x \quad \text{(for integral } k\text{)} \qquad (9)$$

where $p_1 = \dfrac{\beta}{\beta + c}$ and $q_1 = \dfrac{c}{\beta + c}$ so that $p_1 + q_1 = 1$

and c is some suitable constant.

The p. d. f. (9) may also be obtained by a simple combinatorial argument. Consider n Bernoulli trials, and let k be the number of successes and x the number of failures, where p_1 is the constant probability of success associated with each trial, and q_1 is the probability of failure, so that $p_1 + q_1 = 1$. Keep k fixed, and consider the random variable x. It is easy to show that x has p. d. f. given by (9). The p. g. f. of x is

$$g(z) = \sum_{x=0}^{\infty} z^x \binom{x+k-1}{x} p_1^k q_1^x = p_1^k (1-q_1 z)^{-k}$$

$$= (q-pz)^{-k} \qquad (10)$$

where we have put $p = \dfrac{q_1}{p_1}$ and $q = \dfrac{1}{p_1}$, so that $q - p = 1$.

Suppose we define the random variable n to be the number of trials required to obtain exactly k successes. Then $n = k + x$, so that n has p. d. f.

$$\binom{n-1}{k-1} p_1^k q_1^{n-k}$$

and probability generating function (p. g. f.)

$$\sum_{n=k}^{\infty} z^n \binom{n-1}{k-1} p_1^k q_1^{n-k} = (p_1 z)^k (1-q_1 z)^{-k}$$

We shall be concerned with the random variable, x, rather than the closely related random variable, n, which was introduced for the sake of completeness. According to Feller [25], n has a Pascal or negative binomial distribution, while the distribution of x is not named. We shall adopt Katti and Gurland [36]'s terminology and refer to the distribution of x as the Pascal or negative binomial distribution.

If we denote the Pascal probabilities by π_x, as in (9), we may use (9) to prove the recurrence relation

$$\pi_x = \frac{q}{kp}\left[(x+1)\pi_{x+1} - \frac{p}{q} x \pi_x \right] \tag{11}$$

The moments about the origin μ'_r may be obtained as follows. First, let $z = e^t$, then it can be shown that

$$\frac{d^r g(z)}{dt^r} = \sum_{j=1}^{r} a_j(r) \frac{d^j g(z)}{dz^j} \tag{12}$$

where the $a_j(r)$ satisfy the recurrence relation

$$a_j(r) = (r-j+1)\, a_{j-1}(r-1) + a_j(r-1) \tag{13}$$

which has boundary conditions

$$a_j(r) = 0 \quad \begin{cases} j > r \\ j = 0 \text{ and any } r \geq 0 \end{cases}$$

Using (13) it is easy to form a table of values of $a_j(r)$. From (10) we have

$$\frac{d^j g(z)}{dz^j} = p^j k(k+1) \dots (k+j-1)(q-pz)^{-k-j}$$

Thus

$$\mu_r' = \frac{d^r g_r}{dt^r}\bigg|_{t=0} = \sum_{j=1}^{r} a_j(r)\,\frac{d^j g}{dz^j}\bigg|_{z=1}$$

$$= \sum_{j=1}^{r} a_j(r)\; p^j\, k(k+1)\cdots(k+j-1) \qquad (14)$$

In particular, we have from (14) that

$$\mu_1' = kp$$

$$\mu_2' = kp\left[(k+1)p + 1\right]$$

and

$$\mu_3' = kp\left[(k+1)(k+2)p^2 + 3(k+1)p + 1\right]$$

To conclude this section we show that the Pascal distribution approaches the Poisson distribution in the limit. Consider the limit of (10) as $k \to \infty$ and $p \to 0$ such that $kp = \lambda$, a constant greater than zero. Then

$$\lim\,(q - pz)^{-k} = \lim\, q^{-k}\left(1 - \frac{pz}{q}\right)^{-k}$$

$$= \lim\left(1 + \frac{\lambda}{k}\right)^{-k}\left(1 - \frac{\lambda z}{k(1+p)}\right)^{-k}$$

$$= e^{\lambda(z-1)}$$

which is the p. g. f. of a Poisson random variable with parameter λ .

3.4 The Neyman Type A Distribution

Suppose that $h(z)$ in (1) is the p. g. f. of a Poisson random variable with parameter λ_1, then we obtain the Poisson Poisson distribution or Neyman Type A distribution, having p. g. f.

$$g(z) = e^{\lambda\left[e^{\lambda_1(z-1)} - 1\right]} \tag{15}$$

where $\lambda_1 > 0$ and $\lambda > 0$. Expanding $g(z)$ gives

$$g(z) = \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} e^{\lambda_1 j(z-1)}$$

$$= \sum_{j=0}^{\infty}\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} e^{-\lambda_1 j}(\lambda_1 j)^x \frac{z^x}{x!}$$

so that the x th probability P_x is given by

$$P_x = \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} e^{-\lambda_1 j}\frac{(\lambda_1 j)^x}{x!} \tag{16}$$

Using (16) we shall derive the recurrence relation

$$P_{x+1} = \frac{\lambda\lambda_1 e^{-\lambda_1}}{(x+1)}\sum_{k=0}^{x}\frac{\lambda_1^k}{k!}P_{x-k} \tag{17}$$

For

$$P_{x+1} = \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} e^{-\lambda_1 j}\frac{(\lambda_1 j)^{x+1}}{(x+1)!}$$

$$= \frac{\lambda_1}{(x+1)}\sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{(j-1)!} e^{-\lambda_1 j}\frac{(\lambda_1 j)^x}{x!}$$

$$= \frac{\lambda\lambda_1 e^{-\lambda_1}}{(x+1)}\sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} e^{-\lambda_1 j}\frac{\lambda_1^x}{x!}\sum_{k=0}^{x}\binom{x}{k} j^{x-k}$$

$$= \frac{\lambda\lambda_1 e^{-\lambda_1}}{(x+1)}\sum_{k=0}^{x}\frac{\lambda_1^k}{k!}\sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} e^{-\lambda_1 j}\frac{\lambda_1^{x-k} j^{x-k}}{(x-k)!}$$

$$= \frac{\lambda\lambda_1 e^{-\lambda_1}}{(x+1)}\sum_{k=0}^{x}\frac{\lambda_1^k}{k!}P_{x-k}$$

To obtain the moments about the origin, substitute $z = e^t$ in (15) and then differentiate with respect to t. The first three moments are

$$\mu_1' = \frac{dg}{dt}\Big|_{t=0} = \lambda\lambda_1$$

$$\mu_2' = \frac{d^2g}{dt^2}\Big|_{t=0} = \lambda\lambda_1\left[\lambda\lambda_1 + \lambda_1 + 1\right]$$

and

$$\mu_3' = \frac{d^3g}{dt^3}\Big|_{t=0} = \lambda\lambda_1\left[\lambda_1^2 + 2\lambda^2\lambda_1^2 + 3\lambda\lambda_1^2 + 3\lambda_1 + 3\lambda\lambda_1 + 4\right]$$

The Neyman Type A distribution approaches the Poisson distribution in the limit. Let $\lambda_1 \to 0$ and $\lambda \to \infty$ such that $\lambda\lambda_1 = \alpha$, a constant greater than zero. Then from (15) we have

$$\lim e^{\lambda\left[e^{\lambda_1(z-1)} - 1\right]} = \lim e^{\lambda\left[\ln e^{\lambda_1(z-1)}\right]}$$

$$= \lim e^{\lambda\lambda_1(z-1)}$$

$$= e^{\alpha(z-1)}$$

Thus the limiting distribution is a Poisson with parameter $\alpha = \lambda\lambda_1$.

3.5 The Poisson v Binomial Distribution

Suppose that $h(z)$ in (1) is the p. g. f. of a binomial random variable, then we obtain the generalized Poisson v Binomial distribution having p. g. f.

$$g(z) = e^{\lambda\left[(q+pz)^n - 1\right]} \tag{18}$$

where $p > 0$, $q > 0$ and $p + q = 1$, and $\lambda > 0$, and n is a positive integer. Expanding (18) we have

$$g(z) = \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} (q + pz)^{nj}$$

$$= \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \sum_{x=0}^{nj} \binom{nj}{x} q^{nj-x} (pz)^x$$

so that the x-th probability is given by

$$P_x = \sum_{j=0}^{\infty} \binom{nj}{x} p^x q^{nj-x} e^{-\lambda} \frac{\lambda^j}{j!} \tag{19}$$

This, we may note, is also the p. d. f. of a compound Binomial $\wedge$ Poisson distribution, and serves as an illustration of the theorem due to Gurland [33] in section 3.2.

If in the recurrence relation (4) we let

$$\pi_{x-j+1} = \binom{n}{x-j+1} p^{x-j+1} q^{n-x+j-1}$$

and denote the Poisson $\vee$ Binomial probabilities by P_x, we obtain the following recurrence relation

$$P_{x+1} = \frac{\lambda n p}{(x+1)} \sum_{j=0}^{x} \binom{n-1}{j} p^j q^{n-j+1} P_{x-j} \tag{20}$$

This equation is used by Sprott [73] in his method of determining the maximum likelihood estimates of the parameters of the Poisson $\vee$ Binomial distribution.

By substituting $z = e^t$ in (18) and differentiating with respect to t, we obtain the first three moments about the origin,

$$\mu_1' = \left.\frac{dg}{dt}\right|_{t=0} = \lambda n p$$

$$\mu_2' = \left.\frac{d^2 g}{dt^2}\right|_{t=0} = \lambda n p^2 (\lambda n + n - 1) + \lambda n p$$

and $\mu_3' = \dfrac{d^3 g}{dt^3}\Big|_{t=0} = \lambda n p^3 \left[\lambda^2 n^2 + 3\lambda n(n-1) + (n-1)(n-2)\right] + 3\lambda n p^2(\lambda n + n - 1) + \lambda n p$

To conclude our discussion of the Poisson v Binomial distribution we give two limiting distributions. Suppose that $p \to 0$ and $n \to \infty$ such that $np = \alpha$, a constant greater than zero. Then considering (18) we have

$$\lim e^{\lambda\left[(q+pz)^n - 1\right]} = \lim e^{\lambda\left[(1-p)^n\left(1+\frac{pz}{1-p}\right)^n - 1\right]}$$

$$= \lim e^{\lambda\left[\left(1-\frac{\alpha}{n}\right)^n\left(1+\frac{\alpha z}{n(1-p)}\right)^n - 1\right]}$$

$$= e^{\lambda\left[e^{-\alpha}e^{\alpha z} - 1\right]} = e^{\lambda\left[e^{\alpha(z-1)} - 1\right]}$$

which is the p. d. f. of the Neyman Type A distribution having parameters λ and $\alpha = np$. This result is not too surprising as the binomial distribution, under these limiting conditions, approaches the Poisson distribution. Suppose now that $p \to 0$ and $\lambda \to \infty$ such that $\lambda p = \alpha$, a constant greater than zero. Then from (18) we have that

$$\lim e^{\lambda\left[(q+pz)^n - 1\right]} = \lim e^{\lambda\left[\ln(q+pz)^n\right]}$$

$$= \lim (q+pz)^{n\lambda}$$

$$= \lim (1-p)^{n\lambda}\left(1+\frac{pz}{1-p}\right)^{n\lambda}$$

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

$$= e^{n\alpha(z-1)}$$

which is the p. g. f. of a Poisson distribution having parameter $n\alpha$.

3.6 The Poisson $\vee$ Pascal Distribution

If we select h(z) of (1) to be the p. g. f. (10) of the Pascal distribution we obtain the generalized Poisson $\vee$ Pascal distribution, sometimes referred to as the generalized Polya - Aeppli distribution, having p. g. f. given by

$$g(z) = e^{\lambda\left[(q-pz)^{-k}-1\right]} \tag{21}$$

where $p > 0$, $q - p = 1$ and $k > 0$. Expanding (21) we have

$$\begin{aligned} g(z) &= \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}(q-pz)^{-kj} \\ &= \sum_{j=0}^{\infty}\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}\binom{-kj}{x} q^{-kj-x}(-pz)^x \\ &= \sum_{j=0}^{\infty}\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}\binom{kj+x-1}{x} q^{-kj-x} p^x z^x. \end{aligned}$$

so that the Poisson $\vee$ Pascal probabilities P_x are given by

$$P_x = \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}\binom{kj+x-1}{x} q^{-kj-x} p^x \tag{22}$$

We may note that P_x can also be obtained by compounding a Pascal distribution with a Poisson distribution.

If in the recurrence relation (4) we let

$$\pi_{x-j+1} = \binom{k+x-j}{x-j+1} p^{x-j+1} q^{-k-x+j-1} \tag{23}$$

the Pascal probability obtained from either (9) or (10), we obtain the following recurrence relation for the Poisson ∨ Pascal probabilities P_x,

$$P_{x+1} = \frac{\lambda k p}{(x+1)} \sum_{j=0}^{x} \binom{k+j}{j} p^j q^{-k-j-1} P_{x-j} \tag{24}$$

From (21) we may obtain the moments about the origin by first substituting $z = e^t$, and then differentiating with respect to t. The first three moments are

$$\mu_1' = \left.\frac{dg}{dt}\right|_{t=0} = \lambda k p$$

$$\mu_2' = \left.\frac{d^2 g}{dt^2}\right|_{t=0} = \lambda k p^2 (\lambda k + k + 1) + \lambda k p$$

and

$$\mu_3' = \lambda k p^3 \left[\lambda^2 k^2 + 3\lambda k + (k+1)(k+2)\right] + 3\lambda k p^2 (\lambda k + k + 1) + \lambda k p$$

Under three different limiting conditions the Poisson ∨ Pascal distribution approaches the Poisson, the Neyman Type A and the Pascal distribution. First, suppose that $p \to 0$ and $\lambda \to \infty$ such that $\lambda p = \alpha$, a constant greater than zero. Then from (21) we have

$$\lim e^{\lambda\left[(q-pz)^{-k} - 1\right]} = \lim e^{\lambda k p \left[\frac{(q-pz)^{-k} - 1}{kp}\right]}$$

$$= \lim e^{k\alpha \left[\frac{\ln (q-pz)^{-k}}{kp}\right]}$$

$$= \lim e^{k\alpha \left[-\frac{\ln q}{p} - \frac{\ln\left(1 - \frac{pz}{(1+p)}\right)}{p}\right]} = e^{k\alpha(z-1)}$$

which is the p. g. f. of a Poisson random variable with parameter k. Secondly, suppose that $p \to 0$ and $k \to \infty$ such that $kp = \lambda_1$, a constant greater than zero. Then consider the limit of (21)

$$\lim e^{\lambda\left[(q-pz)^{-k}-1\right]} = \lim e^{\lambda\left[(1+p)^{-k}\left(1-\frac{pz}{(1+p)}\right)^{-k}-1\right]}$$

$$= \lim e^{\lambda\left[\left(1+\frac{\lambda_1}{k}\right)^{-k}\left(1-\frac{\lambda_1 z}{k(1+p)}\right)^{-k}-1\right]}$$

$$= e^{\lambda\left[e^{\lambda_1(z-1)}-1\right]}$$

This is the p. g. f. of a Neyman Type A distribution having parameters λ and $\lambda_1 = kp$. Finally consider the limit of (21) as $k \to 0$ and $\lambda \to \infty$ such that $\lambda k = k_1$, a positive constant. Then

$$\lim e^{\lambda\left[(q-pz)^{-k}-1\right]} = \lim e^{\lambda\left[\ln(q-pz)^{-k}\right]}$$

$$= \lim (q-pz)^{-k\lambda} = (q-pz)^{-k_1}$$

and this is the p. g. f. of a Pascal distribution.

Chapter Four

The Complete Poisson Distribution

4.1 Introduction

The complete Poisson distribution has universal application so that fitting it to observed data is an important problem. Some function of the observations must be selected so as to provide as "good" an estimate of the parameter as possible. Because of the simple form of the complete Poisson distribution, the problem could be considered to be solved. In this chapter we consider point estimation of λ, of integral powers of λ, and of real - valued function of λ. A discussion of point estimation of the parameter of a Poisson process concludes the chapter.

4.2 Simple Point Estimation

Let X be a Poisson random variable with parameter λ, then $E\{X\} = \lambda$. Thus, if an observation, x, is taken, the "expected" or "anticipated" value of x is λ. On these grounds we may select the single observation x as an estimate of λ. Again, $E\{X^2\} = \lambda^2 + \lambda$. Thus the "expected" value of the square of the observation is $\lambda^2 + \lambda$, so that the solution

$$\lambda_x = \frac{-1 + \sqrt{1+4x^2}}{2}$$

of the quadratic equation

$$\lambda^2 + \lambda - x^2 = 0$$

may be taken as an estimate of λ. These two estimates are based on a single observation and are very rough, yet, in a sense, they reflect the nature of the well-known "method of moments".

The above arguments may be applied to a random sample, $x_1, \ldots, x_n$, of size n from the Poisson distribution, to obtain more accurate estimates of λ. From section 2.4 we know that the random variable $\sum_{i=1}^{n} X_i$ has a Poisson distribution with parameter $n\lambda$. Let $T = \sum_{i=1}^{n} X_i$. Then $E\{T\} = n\lambda$. Thus the observed value for T, say t, may be taken as an estimate of $n\lambda$, so that t/n, is an estimate of λ. Again, $E\{T^2\} = n^2\lambda^2 + n\lambda$, so that the solution

$$\lambda_t = \frac{-1 + \sqrt{1+4t^2}}{2n}$$

of the quadratic equation

$$n^2\lambda^2 + n\lambda - t^2 = 0$$

may be taken as an estimate of λ. These two estimates are based on n observations of the random variable X, and, of course, a single observation of the random variable T.

The estimates λ_x and λ_t are biased, while estimates x and t/n are unbiased. Later we shall show that t/n is an "efficient" or "best" estimate of λ, having minimum variance among all unbiased estimates of λ.

Let $p(x;\lambda)$ be the Poisson p. d. f. or frequency function, and let n_x be the frequency of the value x in a random sample. Then for any two selected values of x, say k_1 and k_2, we have

$$\frac{p(k_1;\lambda)}{p(k_2;\lambda)} = \lambda^{k_1-k_2}\,\frac{k_2!}{k_1!}$$

Thus the ratio of the theoretical frequencies may be replaced by the ratio of the observed frequencies to obtain an estimate for λ, namely,

$$\lambda_R = \left(\frac{k_1!}{k_2!}\;\frac{n_{k_1}}{n_{k_2}}\right)^{\frac{1}{k_1-k_2}}$$

The best values for k_1 and k_2 are those having largest observed frequencies.

4.3 The Maximum Likelihood Estimate

Let $L(x_1, \ldots, x_n; \lambda)$ be the likelihood function of a random sample, $x_1, \ldots, x_n$, of size n, taken from the Poisson distribution having parameter λ. Then

$$L(x_1, \cdots, x_n; \lambda) = e^{-n\lambda}\,\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \tag{1}$$

Take the natural logarithm of (1), differentiate with respect to λ and equate to zero. The solution for λ is the maximum likelihood estimate $\hat{\lambda}$. Thus

$$\hat{\lambda} = \sum_{i=1}^{n} x_i \Big/ n \tag{2}$$

The estimate $\hat{\lambda}$ is the estimate t/n obtained in the preceding section. From section 2.8 we have that the mean of $\hat{\lambda}$ is λ and the variance is λ/n. Thus $\hat{\lambda}$ is an unbiased estimate of λ. Wilks [85] demonstrates that the variance, $\sigma^2(\hat{\theta})$, of any unbiased estimate $\hat{\theta}(x_1, \ldots, x_n)$ of a parameter θ, under a regularity condition, obeys the inequality

$$\sigma^2(\hat{\theta}) \geq \frac{1}{E\left\{S_n^2\right\}}$$

where $S_n = S_n(x_1, \ldots, x_n; \theta) = \dfrac{\partial \ln L}{\partial \theta}$ where $L = L(x_1, \ldots, x_n; \theta)$ is the likelihood function of the random sample. In other words, the variance of unbiased estimators for θ have lower bound

$$1 \Big/ E\left\{S_n^2\right\}$$

Applying Wilks' result to our case where $\theta = \lambda$, gives

$$S_n = \frac{\partial \ln}{\partial \lambda}\left(e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\right)$$

so that

$$S_n^2 = \left(\frac{\sum_{i=1}^{n} x_i}{\lambda} - n\right)^2$$

and $E\left\{S_n^2\right\} = n/\lambda$. Thus $\hat{\lambda} = \sum_{i=1}^{n} x_i/n$ is an "efficient" or "best" estimate as its variance takes on the lower bound value λ/n. The result $\sum_{i=1}^{n} x_i/n$ as an estimate of λ is also obtained by Roy and Mitra [70]'s "ratio method", which is discussed in section 4.4 .

4.4 The U. M. V. U. Estimate of λ^r

Noak [56] has defined the power series distribution (p. s. d.) as the distribution with probability function

$$P\{X=x\} = a(x)\frac{\theta^x}{f(\theta)} \qquad x=0,1,2,\cdots \qquad (3)$$

where $\theta > 0$ is an unknown parameter, $a(x) > 0$ and $a(0) = 1$, and $f(\theta) = \sum_{x=0}^{\infty} a(x)\theta^x$. The Poisson p. d. f. has the form (3) if we let $\theta = \lambda$ and $a(x) = 1/x!$, so that $f(\theta) = e^{\lambda}$. Now, Roy and Mitra [70] have derived the unique uniformly minimum variance unbiased (U. M. V. U.) estimate for θ^r, where r is a given positive integer. Consider a random sample $x_1, \ldots, x_n$ of size n from (3) and define

$$t_r(x) = \begin{cases} 0 & x < r \\ \dfrac{a(x-r)}{a(x)} & x \geq r \end{cases}$$

Then
$$E\{t_r(X)\} = \sum_{x=0}^{\infty} t_r(x)\, a(x)\frac{\theta^x}{f(\theta)} = \theta^r \qquad (4)$$

Put $T = \sum_{i=1}^{n} X_i$. Then T can easily be shown to be a complete sufficient statistic for θ in the sense of Lehmann and Scheffe [48]. The p. d. f. of the statistic T is

$$P\{T=t\} = \frac{\theta^t\, C(t,n)}{[f(\theta)]^n} \qquad t=0,1,2,\cdots \qquad (5)$$

where $C(t, n) = \sum \prod_{i=1}^{n} a(x_i)$ and the summation is over non-negative integral values of $x_1, \ldots, x_n$ such that

$\sum_{i=1}^{n} x_i = t$. It should be noted that the p. d. f. (5) has the form (3). Now define

$$u_r(t) = \begin{cases} 0 & t < r \\ \dfrac{C(t-r,n)}{C(t,n)} & t \geq r \end{cases} \tag{6}$$

Because of (4) we have that

$$E\left\{u_r(T)\right\} = \theta^r \tag{7}$$

Because $u_r(T)$ is an unbiased estimate for θ^r and the statistic T is complete and sifficient for θ we have by the Rao - Blackwell Theorem [64] and [6], that $u_r(T)$ is the unique U. M. V. U. estimate of θ^r. Applying this result to the complete Poisson distribution with parameter λ gives,

$$C(t,n) = \frac{n^t}{t!}$$

so that $u_r(t)$ becomes

$$u_r(t) = \begin{cases} 0 & t < r \\ \dfrac{t^{[r]}}{n^r} & t \geq r \end{cases}$$

where $t^{[r]} = t(t-1)\ldots(t-r+1)$. Thus the unique U. M. V. U. estimate $u_r(T)$ of λ^r is $t^{[r]}/n^r$. If we put $r = 1$, we obtain the estimate t/n of λ, as before.

Let us again consider the general case represented by (3). Put $r = 1$, then for the variance of $u_1(T)$ we have

$$V\left\{u_1(T)\right\} = E\left\{u_1(T)\right\}^2 - \theta^2 \tag{8}$$

and the U. M. V. U. estimate of $V\left\{u_1(T)\right\}$ is

$$v(t) = \left\{u_1(t)\right\}^2 - u_2(t) \tag{9}$$

Applying (8) and (9) to the Poisson distribution, we have

$$V\left\{u_1(T)\right\} = E\left\{\frac{t^2}{n^2}\right\} - \lambda^2 = \frac{\lambda}{n}$$

and

$$v(t) = \frac{t^2}{n^2} - \frac{t(t-1)}{n^2} = \frac{t}{n^2}$$

Thus t/n^2 is a U. M. V. U. estimate of λ/n. In chapter 5 we continue Roy and Mitra's argument, and consider the problem of obtaining the U. M. V. U. estimate of θ^r, when sampling from a distribution of the same generality as (3), but being truncated on the left at a fixed point c.

4.5 The U. M. V. U. Estimate of $g(\lambda)$

Suppose that X is a Poisson random variable with parameter λ, and that $g(\lambda) = e^{-\lambda}(1 + \lambda)$, that is, is the probability that the random variable takes on value 0 or 1. We may wish to obtain the U. M. V. U. estimate of $g(\lambda)$. The following theorem due to Guttman [39] allows us to obtain this estimate. Let X be a discrete random variable, and $t(X)$, a sufficient statistic for the parameter λ. Suppose that $t(X)$ assumes only positive integer values with probabilities

$$p_\lambda(t) = m(\lambda)\, k_t\, \lambda^t \qquad t = 0,1,2,\cdots \tag{10}$$

Let $g(\lambda)$ be a real - valued function of the parameter λ which takes on values in an interval containing the origin. Then there exists an essentially unique

U. M. V. U. estimate of $g(\lambda)$, if and only if,

$$G(\lambda) = \frac{g(\lambda)}{m(\lambda)}$$

is analytic at $\lambda = 0$, with power series expansion

$$G(\lambda) = \sum_{t=0}^{\infty} a_t \lambda^t$$

such that $a_t = 0$ for all t for which $k_t = 0$.

The sufficiency part of the proof follows. Define

$$f_t = \begin{cases} \dfrac{a_t}{k_t} & k_t \neq 0 \\ \text{arbitrary} & k_t = 0 \end{cases} \tag{11}$$

then

$$\begin{aligned} E\{f_t\} &= \sum_{t=0}^{\infty} f_t\, m(\lambda)\, k_t\, \lambda^t \\ &= \sum_{t=0}^{\infty} a_t\, m(\lambda)\, \lambda^t \\ &= m(\lambda) \sum_{t=0}^{\infty} a_t \lambda^t \\ &= m(\lambda)\, G(\lambda) \quad = g(\lambda) \end{aligned}$$

Thus f_t is an unbiased estimate for $g(\lambda)$ and it is uniquely defined for points of non-zero probability. Since, putting $a_t = 0$ in (11) implies that $f_t = 0$ is the only unbiased estimate of zero, we have that f_t is complete. The statistic $t(X)$ is given as being sufficient for λ. Thus by the Rao - Blackwell theorem [64] and [6], f_t is the U. M. V. U. estimate of $g(\lambda)$. For the necessity part of the proof assume that f_t is an unbiased estimate of $g(\lambda)$. Then

$$\sum_{t=0}^{\infty} f_t\, m(\lambda)\, k_t\, \lambda^t = g(\lambda)$$

Thus
$$\sum_{t=0}^{\infty} a_t \lambda^t = \frac{g(\lambda)}{m(\lambda)} = G(\lambda)$$

where $a_t = f_t\, k_t$. Since λ includes the origin, $G(\lambda)$ is analytic at $\lambda = 0$. The remainder of the proof follows as in the sufficiency part.

Let us apply the theorem to obtain the U. M. V. U. estimate of $g(\lambda) = e^{-\lambda}(1+\lambda)$ where λ is the parameter of a Poisson distribution. Now, in section 2.4, we have shown that $\sum_{i=1}^{n} X_i$ has a Poisson distribution with parameter $n\lambda$. Also it is easy to show that $\sum_{i=1}^{n} X_i$ is a sufficient statistic for λ. Then, let $t = \sum_{i=1}^{n} X_i$. From (10) we have that $m(\lambda) = e^{-n\lambda}$ and $k_t = n^t/t!$. Thus, we may obtain a_t from the identity

$$\sum_{t=0}^{\infty} a_t \lambda^t = \frac{g(\lambda)}{m(\lambda)} = e^{\lambda(n-1)}$$

$$= \sum_{t=0}^{\infty} \frac{\lambda^t}{t!} (n-1)^t$$

Comparing the coefficients of λ^t gives

$$a_t = \frac{(n-1)^{t-1}}{(t-1)!}\left(\frac{n-1}{t} + 1\right)$$

so that $f_t = \left(\frac{n-1}{n}\right)^t \left(1 + \frac{t}{n-1}\right)$ is the U. M. V. U. estimate of the probability that the Poisson random variable takes on value 0 or 1.

4.6 The Poisson Process

In section 2.6 it was shown that $2\lambda T$ has a χ^2_{2n} distribution where T is the amount of observation required for exactly n events to occur in a λ-Poisson process. Using this fact we may obtain a confidence interval for λ having confidence coefficient $1-\epsilon$. From tables of the χ^2_{2n} distribution obtain C and D such that

$$P\left\{ C \leq 2\lambda T \leq D \right\} = 1-\epsilon$$

Now, given the event

$$\frac{C}{2T} < \lambda < \frac{D}{2T}$$

and a point estimate of λ, say λ', we can minimize the maximum percentage deviation of λ' from λ by selecting

$$\lambda' = \frac{C+D}{4T}$$

In this case the maximum percentage deviation of λ' from λ, with respect to λ', occurs when

$$\lambda = D/2T$$

and is

$$\tau = \frac{(D-C)}{(C+D)} 100\ \%$$

As the number n of events observed increases, τ decreases. Thus, from χ^2_{2n} tables a value for n can be obtained so as to provide an estimate λ' having maximum percentage deviation τ less than some given positive value, all at a confidence level $1-\epsilon$.

As a second problem, suppose that an estimate λ' of λ is desired such that for given $\epsilon > 0$ and $\beta > 0$

$$P\left\{ |\lambda' - \lambda| \leq \beta \right\} \geq 1 - \epsilon$$

Let T_n be the "time" required for n events to occur. Put

$$c = \frac{\epsilon \beta^2}{2n}$$

Observe the Poisson process for $\frac{1}{2cT_n}$ units of "time", and define the random variable X as the number of events occuring. Then

$$\lambda' = 2c\,T_n\,X$$

is the desired point estimate for λ. For

$$E\left\{\lambda'\right\} = 2cT_n\,E\left\{X\right\} = \lambda$$

and

$$V\left\{\lambda'\right\} = E\left\{\lambda'\right\}^2 - \lambda^2 = 4c^2 T_n^2\,E\left\{X\right\}^2 - \lambda^2$$

$$= 2\,cT_n\,\lambda = 2cn$$

From Tchebycheff's inequality we have, finally, that

$$P\left\{|\lambda' - \lambda| \leq \beta\right\} \geq 1 - \frac{2nc}{\beta^2} = 1 - \epsilon$$

The results given in this section were developed by Birnbaum [5].

Chapter Five

The Truncated Poisson Distribution

5.1 Introduction

This chapter deals with the problem of point estimating the parameter of a truncated Poisson distribution. The emphasis is placed on the simplest and most useful case of truncation, that is, truncation of the zero class. To begin, a very general procedure for obtaining the maximum likelihood estimate of a truncated distribution is presented. Then, a discussion of the case of truncation on the right is given, a case where no unbiased estimates of the parameter exist. The final section offers a rather extensive discussion of the case of truncation on the left.

5.2 The Iterative Maximum Likelihood Procedure

A "truncated" Poisson distribution is one in which a certain subset A of the sample space, $0, 1, 2, \ldots$ of the complete distribution, is missing. The probabilities of the remaining values x are

$$\frac{e^{-\lambda}\lambda^{x}}{\left(1-\sum_{r\in A} e^{-\lambda}\frac{\lambda^{r}}{r!}\right)x!} \tag{1}$$

Thus, a random sample of size n, drawn from the truncated Poisson distribution, will not contain values from A.

Hartley [40] provides an iterative approach to maximum likelihood estimation from incomplete data that is applicable to any discrete distribution for which a maximum likelihood procedure for the complete distribution is available. Although in many cases special methods are simpler (for example, when tables are available), Hartley's [40] method applies to the many cases where no such special methods exist.

First of all, we shall introduce notation, and then, outline the procedure for obtaining the maximum likelihood estimate of λ, the parameter of a suitable, yet unspecified, discrete distribution having p. d. f. $f(x;\lambda)$. Next we shall give Hartley's proof that the procedure does provide the maximum likelihood estimate. Now, let:

A^* - be the set of permissible values, labelled by i, and A - be the set of missing values, labelled by j. Thus the union of A^* and A is the sample space of the complete p. d. f., $f(x;\lambda)$. Let:

n_i - be the observed frequency of value i, $i \in A^*$,

n_j - be the unknown, unobserved frequency of value j, $j \in A$,

${}_cn_j$ - be the c-th estimate of n_j,

n - be the observed sample size,

${}_cn'$ - be the c-th estimate of the total number of missing frequencies, that is ${}_cn' = \sum_{j \in A} {}_cn_j$,

$f(i;\lambda)$ - be the probability that $x = i$, $i \in A^*$,

$f(j;\lambda)$ - be the probability that $x = j$, $j \in A$,

and finally, let

$$f(\lambda) = \sum_{j \in A} f(j;\lambda)$$

Then the procedure for obtaining the maximum likelihood estimate, $\hat{\lambda}$, is:

1. For all $j \in A$, make initial, probably rough, estimates of the missing frequencies n_j, and denote these estimates by ${}_0n_j$. Compute ${}_0n' = \sum_{j \in A} {}_0n_j$.
2. Using the given n_i and the ${}_0n_j$ compute an initial estimate of λ, denoted by ${}_1\lambda$, from the expression for $\hat{\lambda}$ in the complete case. For the purposes of illustration, when the maximum likelihood estimate of λ in the complete case is the mean, we have,

$$ {}_1\lambda = \frac{\sum_{i \in A^*} i\, n_i + \sum_{j \in A} j\, {}_0n_j}{n + {}_0n'} \tag{2}$$

3. Using ${}_1\lambda$ compute "improved" estimates, ${}_1n_j$, of the missing frequencies n_j from

$$ {}_1n_j = n \frac{f(j;{}_1\lambda)}{1 - f(\lambda)}$$

and then obtain ${}_1n' = \sum_{j \in A} {}_1n_j$.

4. With the n_i and the "improved" ${}_1n_j$, compute an "improved" estimate ${}_2\lambda$, of λ, from

$$ {}_2\lambda = \frac{\sum_{i \in A^*} i\, n_i + \sum_{j \in A} j\, {}_1n_j}{n + {}_1n'} \tag{3}$$

5. Continue this procedure, obtaining "improved" estimates ${}_1\lambda$, ${}_2\lambda$, ..., ${}_c\lambda$, until little change occurs in the estimates of λ. Then the final estimate ${}_c\lambda$ is an approximation to the maximum likelihood estimate of λ, and the approximation may be made to any desired number of decimal places. This completes the procedure.

We now attempt to establish the validity of the procedure. To distinguish the proof from the procedure given, we shall use the symbol θ to refer to the parameter of the discrete random variable x, rather than λ. Let x take on values belonging to two mutually exclusive and exhaustive sets, denoted by A^* and A. The probabilities for $i \in A^*$ and $j \in A$ are denoted by $f(i;\theta)$ and $f(j;\theta)$, respectively. Clearly,

$$\sum_{i \in A^*} f(i;\theta) + \sum_{j \in A} f(j;\theta) = 1$$

Let $f(\theta) = \sum_{j \in A} f(j;\theta)$. Take a random sample from the population of x, and let n_i denote the number of observed values i. The truncation of values $j \in A$ accounts for the fact that no values j occur. Now, the maximum likelihood equation for the estimation of the parameter in the case of truncation is

$$\sum_{i \in A^*} n_i \left[\frac{f^{(1)}(i;\theta)}{f(i;\theta)} + \frac{f^{(1)}(\theta)}{1 - f(\theta)} \right] = 0 \qquad (4)$$

where $f^{(1)}(i;\theta) = \dfrac{\partial f(i;\theta)}{\partial \theta}$ and $f^{(1)}(\theta) = \dfrac{\partial f(\theta)}{\partial \theta}$

Define auxiliary variables n_j, for each $j \in A$, by

$$n_j = n \frac{f(j;\theta)}{1-f(\theta)} \tag{5}$$

Then, (4) becomes

$$\sum_{i \in A^*} n_i \frac{f^{(1)}(i;\theta)}{f(i;\theta)} + \sum_{j \in A} n_j \frac{f^{(1)}(j;\theta)}{f(j;\theta)} = 0 \tag{6}$$

Note that equation (6) is the maximum likelihood equation for a complete sample of size $n + n'$, with observed cell frequencies n_i and n_j. Then the iterative procedure yields solutions

$$\hat{\theta} = \lim_{c \to \infty} {}_c\theta$$

and

$$\hat{n}_j = \lim_{c \to \infty} {}_c n_j \qquad j \in A$$

of equations (5) and (6), and so (4), all, of course, on the assumption that the iterative procedure converges. Hartley [40] maintains that the convergence is extremely rapid in 30 examples that he has worked out.

To obtain a value for the variance of the maximum likelihood estimate, $\hat{\theta}$, define $L(\theta)$ to be the likelihood function, and

$$L_\theta(\theta) = \frac{\partial L(\theta)}{\partial \theta} \quad \text{and} \quad L_{\theta\theta}(\theta) = \frac{\partial^2 L(\theta)}{\partial \theta^2}$$

Recalling that the c-th estimate of $\hat{\theta}$ is ${}_c\theta$ and the fact that $L_\theta(\hat{\theta}) = 0$, we have an estimate of $L_{\theta\theta}(\hat{\theta})$, namely,

$$\hat{L}_{\theta\theta}(\hat{\theta}) = \frac{L_\theta({}_c\theta) - L_\theta(\hat{\theta})}{{}_c\theta - \hat{\theta}}$$

$$= \frac{L_\theta({}_c\theta)}{{}_c\theta - \hat{\theta}}$$

The quantity $\frac{L_\theta(c\theta)}{c\theta-\hat{\theta}}$ is called the "rate of change of score" by R.A. Fisher. Thus, an estimate of the variance of $\hat{\theta}$ is:

$$-\frac{1}{\hat{L}_{\theta\theta}}$$

Hartley's very general method is illustrated in his paper with the Poisson distribution. The method has the advantage that it applies to all cases of truncation. However, iteration involves much time and computation, so that for those distributions having widespread application, such as the Poisson, a more efficient procedure is sought for.

5.3 Truncation on the Right

Let the set A, referred to in section 5.2, consist of the values d+1, d+2, ... where $d \geq 0$ is a fixed integer called "the truncation point". "Permissible" values are then 0, 1, 2, ..., d. Let X be a truncated Poisson random variable with parameter λ, and taking on values $x = 0, 1, 2, \ldots, d$ with probability

$$p(x;\lambda,d) = \frac{e^{-\lambda}\lambda^{x}}{F(d)\;x!} \qquad \begin{matrix} x=0,1,\cdots,d \\ \lambda>0 \end{matrix} \qquad (7)$$

where

$$F(d) = \sum_{r=0}^{d} e^{-\lambda}\frac{\lambda^{r}}{r!} \qquad (8)$$

We shall now prove that an unbiased estimate for the parameter λ does not exist in this special case of truncation on the right. For, assume that an unbiased estimate for λ does exist, and is denoted by $\lambda(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ is a random sample of size n taken from the truncated Poisson distribution (7). Then the joint p. d. f. of $x_1, \ldots, x_n$ is

$$e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \Big/ \left[\sum_{r=0}^{d} e^{-\lambda} \frac{\lambda^r}{r!} \right]^n \tag{9}$$

so that

$$\sum_{x_1=0}^{d} \cdots \sum_{x_n=0}^{d} \lambda(x_1, \ldots, x_n) \frac{e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}}{\left[\sum_{r=0}^{d} e^{-\lambda} \frac{\lambda^r}{r!} \right]^n} = \lambda$$

Simplifying, we have

$$\sum_{x_1=0}^{d} \cdots \sum_{x_n=0}^{d} \lambda(x_1, \ldots, x_n) \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} = \lambda \left[\sum_{r=0}^{d} \frac{\lambda^r}{r!} \right]^n \tag{10}$$

The left side of (10) is a polynomial of degree nd in λ, while the right side of (10) is a polynomial of degree $(nd + 1)$ in λ. This implies that the coefficient of λ^{nd+1} be zero, that is, that

$$\left(1/d! \right)^n$$

be zero, which is of course impossible for finite d and n. Thus no unbiased estimate exists.

Moore [52] has suggested a simple estimate for the parameter λ. From (7) we note that

$$\frac{\sum_{x=0}^{d} x e^{-\lambda} \frac{\lambda^x}{x!}}{\sum_{x=0}^{d-1} e^{-\lambda} \frac{\lambda^x}{x!}} = \lambda$$

If, from a set of observations from (7), we have n_x as the number of x values observed, and n, the total number of observations, then we may select, as an estimate of λ, the quantity

$$\lambda_M = \sum_{x=0}^{d} x n_x \Big/ \sum_{x=0}^{d-1} n_x$$

Moore [52] has obtained an expression for the mean and variance of λ_m, namely,

$$E\{\lambda_M\} = \lambda - \frac{\sum_{x=0}^{d-1} x p(x;\lambda) - \left(\sum_{x=0}^{d-1} p(x;\lambda)\right)\left(\sum_{x=0}^{d} x p(x;\lambda)\right)}{n\left(\sum_{x=0}^{d-1} p(x;\lambda)\right)^2} \tag{11}$$

and

$$V\{\lambda_M\} = \frac{\sum_{x=0}^{d} x^2 p(x;\lambda) - \left(\sum_{x=0}^{d} x p(x;\lambda)\right)^2}{n\left(\sum_{x=0}^{d-1} p(x;\lambda)\right)^2}$$

where we have defined $p(x;\lambda)$ in section 2.2 to be

$$p(x;\lambda) = e^{-\lambda}\lambda^x / x!$$

It may be noted from (11) that λ_m has a bias which decreases as n increases. An estimate for the variance of λ_m, $V\{\lambda_m\}$ is

$$s_x^2 = \frac{\sum_{x=0}^{d} x^2 n_x - \left(\sum_{x=0}^{d} x n_x\right)^2 \Big/ n}{\left(\sum_{x=0}^{d-1} n_x\right)^2}$$

A maximum likelihood procedure for estimating λ is given by Cohen [13]. From (9) we have the likelihood equation, which in slightly different form, is

$$L = L(x_1, \cdots, x_n; \lambda, d) = \prod_{x=0}^{d} \left[\frac{e^{-\lambda} \lambda^x}{F(d)\ x!} \right]^{n_x}$$

Thus

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{x=0}^{d} n_x \left[\frac{x}{\lambda} - \frac{F(d-1)}{F(d)} \right] \tag{12}$$

Setting (12) equal to zero and solving for λ gives the maximum likelihood estimate, $\hat{\lambda}$, that is,

$$\bar{x} = \hat{\lambda} \frac{F(d-1)}{F(d)}$$

where $\bar{x}$ is the sample mean, $\sum_{x=0}^{d} x\, n_x / n$ and $F(d)$ is defined in (8). Cohen [13] provides tables of $\bar{x}$ as a function of λ and d, for λ from 0.005 to 14.5 and d from 1 to 34, with λ given to 5 decimal places. When $\bar{x}$ and d are given, $\hat{\lambda}$ is obtained accurately from the tables by inverse linear interpolation.

Substitute (8) into (13) and obtain

$$\sum_{x=0}^{d} \lambda^x \left[\frac{d!}{(x-1)!} - \bar{x} \frac{d!}{x!} \right] = 0 \tag{14}$$

a polynomial equation of degree d in λ having exactly one positive root. If a more accurate value for $\hat{\lambda}$ is desired, (14) may be solved by one of the several standard, iterative methods of determining roots of polynomial equations.

The variance of $\hat{\lambda}$, say $V\{\hat{\lambda}\}$, is found to be

$$V\{\hat{\lambda}\} \simeq \frac{\lambda}{\mathcal{N}} \chi_d(\lambda)$$

where

$$\chi_d(\lambda) = \frac{\left(F(d)\right)^2}{F(d-1)\left[F(d) + \lambda\, p(d;\lambda)\right] - \lambda F(d)\, p(d-1;\lambda)}$$

where $p(d;\lambda) = e^{-\lambda} \frac{\lambda^d}{d!}$. Cohen [13] provides a table of the variance function $\chi_d(\lambda)$, as a function of λ and d, for d = 2 (1) 14 and λ going from 0.001 to 15.

Murakami, Asai and Kawamura [53] have also obtained the maximum likelihood estimate $\hat{\lambda}$ for this particular case of the truncated Poisson distribution. They suggest the use of nomograms for obtaining $\hat{\lambda}$, given $\bar{x}$ and d. However, their approach does not offer as accurate values for $\hat{\lambda}$ as that of Cohen [13]. Murakami, et al [53] examine the relative efficiency of Moore's estimate, λ_m, compared to the maximum likelihood estimate $\hat{\lambda}$. They determine an approximate expression for $V(\lambda_m)$, and using it, determine the ratio

$$V\{\hat{\lambda}\} \big/ V\{\lambda_M\}$$

For given r = 1 (1) 10 this ratio is plotted as a function of λ, with λ ranging from 0 to 10. They conclude that, although the variance of the maximum likelihood estimate is smaller for all r than the variance of Moore's estimate, the ease in obtaining the latter outweighs the advantages of better efficiency.

However, the tables provided by Cohen [13] permit one to determine $\hat{\lambda}$ quickly and accurately, and so, the consistent and asymptotically efficient maximum likelihood estimate, $\hat{\lambda}$, seems to be preferable in this particular case of the truncated Poisson distribution.

5.4 Truncation on the Left

Let the set A, referred to in section 5.2, consist of the values 0, 1, 2, ..., c where $c \geq 0$. is a fixed integer called "the truncation point". "Permissible" values are then c+1, c+2, ... Let X be a random variable having a truncated Poisson distribution with parameter λ and p. d. f.

$$p_*(x;\lambda,c) = \frac{e^{-\lambda}\lambda^x}{[1-F(c)]\,x!} \qquad x = c+1, c+2, \cdots \quad \lambda > 0 \tag{15}$$

where

$$F(c) = \sum_{r=0}^{c} e^{-\lambda}\frac{\lambda^r}{r!}$$

A simple procedure for estimating λ is given by Rider [68]. Let a random sample of size n be taken from (15) and let n_x be the numbers of x values observed, x = c+1, c+2, ... Let N be the total number of the observations, assuming that we have sampled from a complete distribution, and that the observations actually observed form only a part of the "complete" sample. Define

$$T_0 = \sum_{x=c+1}^{\infty} n_x \tag{16}$$

$$T_1 = \sum_{x=c+1}^{\infty} x\, n_x \tag{17}$$

and
$$T_2 = \sum_{x=c+1}^{\infty} x^2 n_x \tag{18}$$

Also let
$$T_0' = N \sum_{r=0}^{c} p(r;\lambda) + T_0 \tag{19}$$

$$T_1' = N \sum_{r=0}^{c} r\, p(r;\lambda) + T_1 \tag{20}$$

and
$$T_2' = N \sum_{r=0}^{c} r^2 p(r;\lambda) + T_2 \tag{21}$$

where $p(r;\lambda) = e^{-\lambda}\lambda^r/r!$. Then we may take T_1'/T_0' , and T_2'/T_0', to be estimates of λ and $\lambda + \lambda^2$, respectively. Thus
$$T_1' = \lambda T_0' \tag{22}$$

and
$$T_2' = (\lambda + \lambda^2) T_0' \tag{23}$$

Substituting (19), (20) and (21) into (22) and (23) gives
$$T_1 - \lambda T_0 = N e^{-\lambda} \frac{\lambda^{c+1}}{(c+1)!} \tag{24}$$

$$T_2 - (\lambda + \lambda^2) T_0 = N e^{-\lambda} \frac{\lambda^{c+1}}{c!} (\lambda + c + 1) \tag{25}$$

Solving (24) and (25) for λ gives us Rider's [68] estimate
$$\lambda_R = \frac{T_2 - (c+1) T_1}{T_1 - c T_0}$$

This estimate has the advantage of being easy to obtain; however, it has the disadvantages of being biased and inefficient.

Tate and Goen [77] mention a simple estimate for λ. Let

$$U_c(x) = \begin{cases} 0 & x = c + 1 \\ x & x \geq c + 2 \end{cases}$$

Then an estimate for λ is $V_c(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} U_c(x_i)$.

For the special case when $c = 0$, $V_0(x_1, \ldots, x_n)$ is Plackett's estimate, λ_P, which shall be considered shortly. The estimate $V_c(x_1, \ldots, x_n)$ is superior to Rider's estimate, λ_R, for two reasons. It is unbiased and simpler to compute.

To obtain the maximum likelihood estimate of λ, say $\hat{\lambda}_c$, we note from (15) that the likelihood function is

$$L = L(x_1, \cdots, x_n; \lambda, c) = \prod_{x=c+1}^{\infty} \left[\frac{e^{-\lambda} \lambda^x}{[1-F(c)]\, x!} \right]^{n_x} \tag{26}$$

so that

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{x=c+1}^{\infty} \left[\frac{x}{\lambda} - \frac{1-F(c-1)}{1-F(c)} \right] \tag{27}$$

Equating (27) to zero, we arrive at

$$\bar{x} = \hat{\lambda}_c \left(1-F(c-1)\right) / \left(1-F(c)\right) \tag{28}$$

where

$$\bar{x} = \sum_{x=c+1}^{\infty} x\, n_x \Big/ n$$

The case of truncation at zero, that is, where $c = 0$, is especially important. Putting $c = 0$ in (28) gives

$$\bar{x} = \frac{\hat{\lambda}_0}{1 - e^{-\hat{\lambda}_0}} \tag{29}$$

David and Johnson [21] derive (29) and maintain that " it does not seem possible to obtain an explicit expression for $\hat{\lambda}_0$ ", and imply that tables would be helpful.

Irwin [43] derives an explicit expression for $\hat{\lambda}_0$ from (29) in the form of a Lagrange series. First

$$\hat{\lambda}_0 = \bar{x} - \bar{x}e^{-\lambda}$$

and by Lagrange's expansion,

$$\begin{aligned}\hat{\lambda}_0 &= \bar{x} - \bar{x}e^{-\bar{x}} + \frac{\bar{x}^2}{2!}\frac{de^{-2\bar{x}}}{d\bar{x}} - \cdots + (-1)^r\frac{\bar{x}^r}{r!}\frac{d^{r-1}e^{-r\bar{x}}}{d\bar{x}^{r-1}} + \cdots \\ &= \bar{x} - \sum_{r=1}^{\infty}\frac{r^{r-1}}{r!}\left(\bar{x}e^{-\bar{x}}\right)^r\end{aligned}$$

Using Stirling's Theorem Irwin [43] demonstrates that the series is convergent for $\bar{x} \geq 1$ but only satisfactorily for $\bar{x} \geq 2$. He illustrates the applicability of this expansion by obtaining $\hat{\lambda}_0$ from data given in Finney and Varley [26] on the distribution of eggs laid in unopened flower heads of the black knapweed by the Knapweed gall - fly.

Finney and Varley [26] maintain that (29) can be solved rapidly by iterative or interpolatory processes and a table for direct reading of $\hat{\lambda}_0$ and N V($\hat{\lambda}_0$) as functions of $\bar{x}$ could easily be constructed where N is the total number of observations and V($\hat{\lambda}_0$) is the variance of $\hat{\lambda}_0$.

David and Johnson [21] and Rider [68] have provided tables for solution of (29) but these tables are inadequate. Cohen [14] provides adequate tables for solution of (29) for

$$\bar{x} = 1.0005\ (0.0005)\ 1.005\ (0.005)\ 1.250$$
$$(0.01)\ 1.75\ (0.05)\ 5.00\ (0.1)\ 11.0$$
$$\text{and}\ \ 11.0\ \text{ to }\ 12.5\ .$$

Linear interpolation using the tables offers accuracy to at least 3 decimal places and usually 4. A folded scale chart of $\delta = \bar{x} - \lambda$ as a function of $\bar{x}$ is given when a quick solution of (29) is desired. By putting $c = 0$ in (26) we obtain

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -n\left[\frac{\bar{x}^2}{\lambda^2} - \frac{e^{-2\lambda}}{(1-e^{-\lambda})^2}\right]$$

Thus, an asymptotic expression for the variance of $\hat{\lambda}_0$, $V(\hat{\lambda}_0)$ is

$$V(\hat{\lambda}_0) = -\frac{1}{E\left\{\frac{\partial^2 \ln L}{\partial \lambda^2}\right\}} \backsimeq \frac{\lambda}{n}\,\chi(\lambda)$$

where

$$\chi(\lambda) = \frac{(1-e^{-\lambda})^2}{1-(\lambda+1)e^{-\lambda}}$$

Also, it is easy to see that

$$\frac{\lambda}{n} \leq V(\hat{\lambda}_0) \leq \frac{2\lambda}{n}$$

Cohen [14] gives a table having $\chi(\lambda)$ tabulated for λ ranging from 0.001 to 14.5, and a graph of $\chi(\lambda)$

plotted against λ.

Two relatively simple methods for obtaining an estimate of λ for the case of truncation on the left with $c = 0$, are due to Plackett [60] and David and Johnson [21]. Plackett suggests using the unbiased estimate

$$\lambda_p = \sum_{x=2}^{\infty} x n_x \Big/ n \qquad (30)$$

The efficiency of λ_p is always greater than 95%, and the estimate is exceptionally easy to calculate. The variance of λ_p can be shown to be exactly

$$V\{\lambda_p\} = \frac{1}{n}\left(\lambda + \frac{\lambda^2}{e^{\lambda}-1}\right)$$

An unbiased estimate for $V(\lambda_p)$ is

$$\left(n\lambda_p + 2n_2\right) \Big/ n^2$$

David and Johnson [21] suggest the method of moments. Let M_1' and M_2' be the first and second population moments about the origin, respectively, of a Poisson distribution truncated at zero. Let m_1', m_2' be the corresponding sample moments. Then

$$M_1' = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})\,x!} = \frac{\lambda}{1-e^{-\lambda}} \qquad (31)$$

$$M_2' = \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})\,x!} = \frac{\lambda+\lambda^2}{(1-e^{-\lambda})} \qquad (32)$$

$$m_1' = \frac{\displaystyle\sum_{x=1}^{\infty} x n_x}{n}$$

and $$m_2' = \sum_{x=1}^{\infty} x^2 m_x \Big/ m$$

Solving (31) and (32) simultaneously we have

$$\lambda = \frac{M_2'}{M_1'} - 1$$

which suggests an estimate for λ, namely,

$$\overset{\sim}{\lambda} = \frac{m_2'}{m_1'} - 1$$

An approximation to the variance of $\overset{\sim}{\lambda} + 1$ is

$$V\{\overset{\sim}{\lambda}+1\} \simeq \frac{(\lambda+2)(1-e^{-\lambda})}{n}$$

An approximation is also made of the relative efficiency of the "crude" moments estimate $\overset{\sim}{\lambda}$ compared to the maximum likelihood estimate $\hat{\lambda}_0$. Their ratio is

$$\frac{V\{\hat{\lambda}_0\}}{V\{\overset{\sim}{\lambda}\}} \simeq \frac{\lambda\ (e^{\lambda}-1)}{(\lambda+2)(e^{\lambda}-\lambda-1)} \tag{33}$$

David and Johnson [21] evaluate (33) for a few values of λ and obtain values ranging from 0.72 to 0.87 . Of the two simple estimates for λ, λ_P and $\overset{\sim}{\lambda}$, Plackett's λ_P is superior.

The most desirable estimates for parameters are unbiased, and have minimum variance among all unbiased estimates. Tate and Goen [77] have derived the uniformly minimum variance unbiased (U. M. V. U.) estimate for λ, for the general case of the Poisson distribution truncated on the left at $c \geq 0$. However, they only provide adequate tables for the case $c = 0$. A limited solution to the case where $c = 1$ is also given. Their development follows.

Let X be the truncated Poisson random variable having p. d. f. given by (15) and characteristic function denoted by $\phi_c(\alpha)$. Let $X_1, \ldots, X_n$ be a random sample of size n from (15) and let

$$T_c = \sum_{i=1}^{n} X_i \tag{34}$$

It is easy to demonstrate the completeness of T_c as well as its sufficiency for the family $\{p(x;\lambda)\}$, defined in section 2.2 . According to Tukey [81], T_c is also sufficient for the family $\{p_*(x;\lambda,c)\}$.

Treat the case c = 0 first. The characteristic function of T_0, denoted by $\chi_0(\alpha)$ is

$$\begin{aligned}
\chi_0(\alpha) &= \left[\phi_0(\alpha)\right]^n \\
&= \left[\sum_{x=1}^{\infty} \frac{\lambda^x e^{i\alpha x-\lambda}}{(1-e^{-\lambda})\,x!}\right]^n \\
&= \left[\frac{e^{\lambda e^{i\alpha}}-1}{e^{\lambda}-1}\right]^n \\
&= \left[-\frac{1}{(e^{\lambda}-1)}\right]^n \sum_{j=0}^{n} \binom{n}{j} (-1)^j e^{j\lambda e^{i\alpha}}
\end{aligned}$$

Now, the p. d. f. of T_0, denoted by $p_0(t)$, can be obtained by the inversion formula for characteristic functions, as follows

$$p_0(t) = \frac{1}{2\pi} \int \chi_0(\alpha) e^{-i\alpha t}\, d\alpha$$

$$= \left(\frac{-1}{e^{\lambda}-1}\right)^{n} \sum_{j=0}^{n} \binom{n}{j} \frac{(-1)^{j}}{2\pi} \int e^{j\lambda e^{i\alpha} - i\alpha t} \, d\alpha$$

$$= \left(\frac{-1}{e^{\lambda}-1}\right)^{n} \sum_{j=0}^{n} \binom{n}{j} (-1)^{j} \frac{(j\lambda)^{t}}{t!}$$

$$= \frac{\lambda^{t} n!}{(e^{\lambda}-1)^{n}\, t!} \; S(t,n) \qquad t = n, n+1, \cdots \tag{36}$$

where

$$S(t,n) = \begin{cases} \dfrac{(-1)^{n}}{n!} \displaystyle\sum_{j=0}^{n} \binom{n}{j} (-1)^{j} j^{t} & t = n, n+1, \cdots \\ 0 & t < n \end{cases} \tag{37}$$

$S(t, n)$ defines a Stirling number of the second kind. Jordan [44] and Riordan [69] define Stirling numbers and also establish the two following relations

$$S(t,n) = S(t-1, n-1) + n\, S(t-1, n) \tag{38}$$

and

$$S(t+1, n+1) = \sum_{j=n}^{t} \binom{t}{j} S(j,n) \tag{39}$$

Now, suppose that $\hat{\lambda}_0(t)$ is an unbiased statistic for λ based on T_0, then since T_0 is a complete and sufficient statistic for the family $\left\{ p_*(x;\lambda,0) \right\}$, $\hat{\lambda}_0(t)$ is unique and has minimum variance (Lehmann and Scheffe [48]). Thus

$$\sum_{t=n}^{\infty} \hat{\lambda}_0(t) \frac{\lambda^{t} n!}{t!\,(e^{\lambda}-1)^{n}} \; S(t,n) = \lambda \tag{40}$$

But (36) implies that

$$(e^{\lambda}-1)^{n} = \sum_{t=n}^{\infty} \lambda^{t} \frac{n!}{t!} \; S(t,n)$$

so that (40) becomes

$$\sum_{t=n}^{\infty} \hat{\lambda}_0(t) \frac{\lambda^{t}}{t!} S(t,n) = \sum_{t=n}^{\infty} \frac{\lambda^{t+1}}{t!} S(t,n) \tag{41}$$

Equating the coefficients of λ^t in (41) and using (38) results in

$$\hat{\lambda}_0(t) = t\,\frac{S(t-1,n)}{S(t,n)} = \frac{t}{n}\left(1 - \frac{S(t-1,n)}{S(t,n)}\right)$$

$$= \frac{t}{n}\,C(n,t) \tag{42}$$

where $$C(n,t) = 1 - \frac{S(t-1,n-1)}{S(t,n)} \tag{43}$$

From (43) and Miksa [50]'s tables of $S(t, n)$ for $n = 1\ (1)\ t$ with $t = 1\ (1)\ 50$, Tate and Goen [77] offer a table of $C(n, t)$ for $n = 2\ (1)\ t - 1$ with $t = 3\ (1)\ 50$ correct to 5 decimal places. For large t, that is, $t >> n$, Jordan [44] and Riordan [69] show that $S(t, n)$ can be approximated by $n^t/n!$, making

$$\hat{\lambda}_0(t) \simeq \frac{t}{n}\left[1 - \frac{(n-1)^{t-1}}{n^{t-1}}\right] \tag{44}$$

The approximation (44) is satisfactory for $2 \leq n \leq 15$ and $t \geq 51$.

If the variance of $\hat{\lambda}_0(t)$ is denoted by $\sigma^2_{\hat{\lambda}_0}$, Tate and Goen [77] have proved that

$$\frac{\lambda(1-e^{-\lambda})^2}{n(1-e^{-\lambda}-\lambda e^{-\lambda})} < \sigma^2_{\hat{\lambda}_0} < \frac{1}{n}\left(\lambda + \frac{\lambda^2}{e^{\lambda}-1}\right)$$

We shall now continue Tate and Goen [77] 's development for the case when $c > 0$. The statistic, T_c, defined in (34), has characteristic function

$$\chi_c(\alpha) = \left(\sum_{x=c+1}^{\infty} \frac{e^{i\alpha x - \lambda} \lambda^x}{(1-F(c))\, x!} \right)^m$$

$$= \frac{e^{-m\lambda}}{(1-F(c))^m} \left(e^{\lambda e^{i\alpha}} - \sum_{x=0}^{c} e^{i\alpha x} \frac{\lambda^x}{x!} \right)^m$$

Using a similar method to that used in the case where $c = 0$, we obtain the p. d. f. of t, $p_c(t)$, to be

$$p_c(t) = \frac{m!\, \lambda^t\, S^c_{m,t}}{t! \left(e^{\lambda} - \sum_{\tau=0}^{c} \frac{\lambda^\tau}{\tau!} \right)^m} \qquad t = m(c+1), m(c+1)+1, \cdots \tag{45}$$

where

$$S^c_{m,t} = \frac{(-1)^m t!}{m!} \sum \frac{m!\; (-1)^{k_1}\; k_1^{\,t - \sum_{\tau=0}^{c} \tau k_{\tau+2}}}{k_1! \cdots k_{c+2}! \left(t - \sum_{\tau=0}^{c} \tau k_{\tau+2} \right)! \prod_{\tau=0}^{c} (\tau!)^{k_{\tau+2}}} \tag{46}$$

where $k_i = 0, 1, 2, \ldots, n$ with $i = 1, 2, \ldots, c+2$

$t = n(c+1),\ n(c+1)+1, \ldots$

and the summation is taken over all $k_1, \ldots, k_{c+2}$ such that $k_1 + \ldots + k_{c+2} = n$. Note that $S^0_{n,t} = S(t, n)$.

Along the lines used in the $c = 0$ case we obtain the following expression for the U. M. V. U. estimate of λ

$$\hat{\lambda}_c(t) = t\, \frac{S^c_{m,t-1}}{S^c_{m,t}} \tag{47}$$

Jordan [44] defines

$$\bar{C}_{p,i} = \sum_{j=p+1}^{2p-i} (-1)^{j+1} \binom{2p-1}{j} S(j, j-p) \tag{48}$$

and proves that

$$\bar{C}_{t-n,t-2n} = n\,\bar{C}_{t-n-1,t-2n-1} + (t-1)\,\bar{C}_{t-n-1,t-2n}$$

This recurrence relation permits one to tabulate the $\bar{c}_{p,i}$. From (46), (48) and (37) we have for $c = 1$,

$$S'_{n,t} = \bar{C}_{t-n,t-2n}$$

Thus for $c = 1$, (47) reduces to

$$\hat{\lambda}_1(t) = t\,\frac{\bar{C}_{t-n-1,t-2n-1}}{\bar{C}_{t-n,t-2n}}$$

A table in Jordan enables the estimation problem to be solved for $n = 1, 2, \ldots, 5$ with $2n + 1 \leq t \leq n + 6$. More extensive tables of $\bar{c}_{p,i}$ are required so as to permit evaluation of $\hat{\lambda}_1(t)$ over a greater region of n and t. No practical solution seems likely by this method for $c > 1$.

To conclude this section we present Roy and Mitra [70]'s derivation of the U. M. V. U. estimate of θ^2 when θ is the parameter of a p. s. d. truncated on the left (the complete p. s. d. is defined in section 4.4). To conform with their notation let the truncation point be $s - 1$ instead of c. Let X be the random variable having the power series distribution (p. s. d.) truncated on the left at $s - 1$. The p. d. f. of X is

$$P\{X = x\} = \frac{a(x)\,\theta^x}{f_s(\theta)} \qquad x = s, s+1, \cdots \tag{49}$$

where
$$f_s(\theta) = \sum_{x=s}^{\infty} a(x)\,\theta^x$$

We notice that the truncated Poisson distribution (15) has this form when $\theta = \lambda$, $s = c + 1$, $a(x) = 1/x!$ and $f_s(\theta) = e^{\lambda}\left[1 - F(c)\right]$. By similar arguments to those used in section 4.4 the U. M. V. U. estimate of a positive integral power of the parameter, θ^r is $u_{r,s}(T)$ where

$$u_{r,s}(t) = \begin{cases} \dfrac{C_s(t-r, n)}{C_s(t, n)} & t \geq ns+r \\ 0 & t < ns+r \end{cases} \tag{50}$$

where $T = \sum_{i=1}^{n} X_i$ and $C_s(t, n) = \sum_S \prod_{i=1}^{n} a(x_i)$ with the summation $\sum_S$ being over integral values $x_1, \ldots, x_n$ such that $x_1 + x_2 + \ldots + x_n = t$ and $x_i \geq s$. It can be shown that

$$C_s(t, n) = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} \left\{a(s-1)\right\}^j C_{s-1}\left(t - j(s-1), n-j\right) \tag{51}$$

where $C_0(t, n) = C(t, n)$. Thus (51) enables us to evaluate $C_s(t, n)$ for all s, t and n, although it certainly appears to be a tedious chore for large s.
For $s = 1$, that is truncation at zero, (51) reduces to

$$C_1(t, n) = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} C(t, n-j) \tag{52}$$

where $C(t, n)$ is defined in section 4.4. Applying the theory to the Poisson distribution truncated at zero,

(52) simplifies to

$$C_1(t,m) = \frac{1}{t!} \sum_{j=1}^{m} (-1)^{m-j} \binom{m}{j} j^t \tag{53}$$

Roy and Mitra [70] have tabulated $u_{1,1}(t)$ obtainable from (53) and (50), for n = 2 (1) 10 and t = 2 (1) 96. Note that $u_{1,1}(t)$ is the U. M. V. U. estimate of λ, and is identical to $\tilde{\lambda}_o(t)$, developed by Tate and Goen [77].

We are now prepared to suggest a procedure for estimating λ when the zero class is missing from the Poisson distribution. Use the U. M. V. U. estimates, $\tilde{\lambda}_o(t)$ and $u_{1,1}(t)$, whenever possible, that is, when

$$\left.\begin{array}{l} 1 \leq n \leq \quad \text{with } 1 \leq t \leq 50, \\ n = t \geq 51, \\ n = 1 \text{ with } t \geq 51, \end{array}\right\} \quad \text{Use } \tilde{\lambda}_o(t)$$

or

$$\left.\begin{array}{l} n = 2\ (1)\ 10 \\ t = 2\ (1)\ 96 \end{array}\right\} \quad \text{Use } u_{1,1}(t) \quad .$$

For the regions $2 \leq n \leq 15$ with $t \geq 51$ and $n \geq 16$ with $t >> n$ use either the approximation

$$\tilde{\lambda}_o(t) \simeq \frac{t}{m}\left[1 - \left(\frac{m-1}{m}\right)^{t-1}\right]$$

or the maximum likelihood estimate $\hat{\lambda}_o$, obtainable from Cohen [14]'s tables. Outside these regions, use $\hat{\lambda}_o$, if the tables permit, and Plackett's estimate, λ_P if they do not. If no tables are available, Plackett's estimate is recommended.

Chapter Six

Censored and Other Special Poisson Distributions

6.1 Introduction

Censored Poisson distributions have not been as thoroughly investigated by statisticians as the truncated Poisson. Among the main topics in this chapter is Hartley [40]'s iterative maximum likelihood procedure for estimating the parameter λ in a most general case of censoring. The maximum likelihood estimates of the parameters of two special cases of the truncated Poisson, as well as a modified Poisson distribution, are dealt with.

6.2 The Censored Poisson Distribution

Let n_x be the number of x values observed in a random sample of fixed size n from a Poisson population. Classically, this is a "censored" population if the numbers n_x are each known for $x \leq c$, and are unknown, except for their total number, $\sum_{x=c+1}^{\infty} n_x$, for $x > c$, where c is a positive integer called the "point of censorship". Hartley [40] deals with a more general situation which he terms "grouped frequencies" censoring. Here, the entire population is divided into mutually exclusive groups of frequencies. If a random sample is taken, only the total number of observations in each group is known.

We shall consider those discrete distributions having maximum likelihood solutions in the complete case. Since the complete Poisson distribution has a maximum likelihood solution the treatment given applies to the censored Poisson distribution. Let us introduce the notation. Suppose that the entire population of the discrete random variable x is divided into G groups, labelled by g, where $g = 1, 2, \ldots, G$, and that the values of x in each group are labelled by j, where $j = 1, 2, \ldots$. Let

$$f(j,g;\lambda) = P\{x = j^{th} \text{ integer in the } g^{th} \text{ group}\}$$

and

$$F(g;\lambda) = \sum_j f(j,g;\lambda) = P\{x \in g^{th} \text{ group}\} \tag{1}$$

where, λ is the parameter of the discrete random variable x, and $\sum_j$ is the summation over all j.

Let

N_g = total number of observations in g-th group,

${}_c n_{jg}$ = c-th estimate of the j-th frequency in the g-th group

and n = the total number of observations, so that

$$n = \sum_{g=1}^{G} N_g \ .$$

Hartley [40] suggests the following procedure for obtaining the maximum likelihood estimate of λ, say $\hat{\lambda}$.

1. Inspect the group frequencies N_g and estimate roughly the values ${}_0n_{1g}$, ${}_0n_{2g}$, ... for each g.
2. Using the ${}_0n_{jg}$ compute an initial estimate, ${}_1\lambda$, of the maximum likelihood estimate $\hat{\lambda}$, using the maximum likelihood solution for the complete case. For purposes of illustration, if the parameter λ is

estimated by the mean in the complete case, we have

$$_{1}\lambda = \frac{\sum_{j,g} x \; {}_{0}m_{jg}}{m} \qquad (2)$$

where the summation, $\sum_{j,g}$, is over all arrangements of j and g.

3. Using, $_{1}\lambda$, compute "improved" estimates of the individual frequencies, say, $_{1}n_{0g}$, $_{1}n_{1g}$, $_{1}n_{2g}$, ..., with

$$_{1}m_{jg} = N_g \frac{f(j,g;\, _{1}\lambda)}{F(g;\, _{1}\lambda)}$$

4. Using the $_{1}n_{jg}$ compute an "improved" estimate of λ, say $_{2}\lambda$ with

$$_{2}\lambda = \frac{\sum_{j,g} x \; {}_{1}m_{jg}}{m}$$

5. Repeat the procedure until there is little change in the $_{c}\lambda$. The final value for $_{c}\lambda$ is an approximation to the maximum likelihood estimate .

The proof that this procedure yields the maximum likelihood estimate is similar to the one given by Hartley [40] for truncated distributions described in section 5.2 . Again, we denote the parameter by θ to distinguish the proof from the procedure. The maximum likelihood equation for the "grouped frequencies" situation is

$$\sum_{g=1}^{G} \frac{N_g \, F^{(1)}(g;\theta)}{F(g;\theta)} = 0 \qquad (3)$$

where

$$F^{(1)}(g;\theta) = \frac{\partial F(g;\theta)}{\partial \theta}$$

Define auxiliary random variables by

$$n_{jg} = \frac{N_g f(j,g;\theta)}{F(g;\theta)} \tag{4}$$

Detecting that (1) implies

$$F^{(1)}(g;\theta) = \sum_j f^{(1)}(j,g;\theta)$$

we may rewrite (3) as

$$\sum_g \sum_j n_{jg} \frac{f^{(1)}(j,g;\theta)}{f(j,g;\theta)} = 0 \tag{5}$$

where

$$f^{(1)}(j,g;\theta) = \frac{\partial f(j,g;\theta)}{\partial \theta}$$

Now (5) is the maximum likelihood equations for a complete distribution having observed frequencies, $n_{j,g}$ with $n = \sum_j \sum_g n_{j,g}$. Thus, the procedure outlined, if it converges, will yield solutions to (4) and (5), and so, to (3).

Murakami, Asai, and Kawamura [53] examine maximum likelihood estimation in a classically censored Poisson population. Thus, the n_x are known for $x \le c$, and only $\sum_{x=c+1}^{\infty} n_x$ is known for $x > c$. Let the Poisson random variable x have parameter λ. The likelihood function for this censored case is

$$L = \binom{n}{n'} \left[\prod_{x=0}^{c} p(x;\lambda)^{n_x} \right] \left[\sum_{x=c+1}^{\infty} p(x;\lambda) \right]^{n-n'} \tag{6}$$

where $n' = \sum_{x=0}^{c} n_x$ and $n = \sum_{x=0}^{\infty} n_x$, and of course, $p(x;\lambda) = e^{-\lambda}\lambda^x/x!$. Take the natural logarithm of (6) and differentiate partially with respect to λ, to arrive at

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{x=0}^{c} \frac{x n_x}{\lambda} - n + (n-n') \frac{e^{-\lambda}\lambda^c/c!}{\sum_{x=c+1}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}} \tag{7}$$

Setting $\frac{\partial \ln L}{\partial \lambda} = 0$ gives us the maximum likelihood equation, which turns out to be quite complicated. To obtain a solution for λ, say $\hat{\lambda}$, nomograms are constructed as follows. Put

$$q = \frac{\sum_{x=0}^{c} x n_x}{n} = \hat{\lambda}\left(1 - p\,\chi(\hat{\lambda}, c)\right)$$

where
$$\chi(\hat{\lambda}, c) = \frac{e^{-\hat{\lambda}}\hat{\lambda}^c/c!}{\sum_{x=c+1}^{\infty} e^{-\hat{\lambda}}\frac{\hat{\lambda}^x}{x!}}$$

and
$$p = \frac{n-n'}{n}$$

Now, for a given fixed value c, and several different fixed values p, graphs of q vs $\hat{\lambda}$ may be constructed. This has been done by Murakami et al [53] for $c = 1\ (1)\ 10$.

Moore's estimate, λ_M, discussed in section 5.3, may also be used as an estimate of λ in the classical censored case. Murakami et al [53] obtain a slightly more accurate expression for the variance of λ_M, $V(\lambda_M)$. They plot the ratio $V(\hat{\lambda})/V(\lambda_M)$ against λ for $c = 1\ (1)\ 10$. For small λ the efficiency of λ_M is high, but for larger λ, λ_M is considerably inefficient.

6.3 Two Special Cases of the Truncated Poisson Distribution

Let X be a random variable having p. d. f.

$$g(x;\theta,\lambda) = \begin{cases} 1-\theta & x=0 \\ \dfrac{\theta e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda})\,x!} & x=1,2,\cdots \end{cases} \qquad (8)$$

where $\lambda > 0$ and $0 \leq \theta \leq 1$. For example, consider the distribution of biological organisms among colony sites, where no migration occurs between sites. Assume that that sites are distinct and countable, and that each has constant probability θ of being selected. Once a site is selected, assume that the number of organisms there is a truncated Poisson distribution with missing zero class and parameter λ. Then the random variable, defined as the number of organisms counted if a single site is selected at randon, has the distribution represented by (8). Note that when $\theta = 1$ we have the truncated Poisson distribution with missing zero class as a special case of (8).

Cohen [15] demonstrates a maximum likelihood approach to the problem of estimating the parameters λ and θ of (8). Take a random sample of size n from (8), and let n_o be the numbers of zeroes observed, and n^*, the number of observations greater than zero, so that $n_o + n^* = n$. Then the likelihood function is

$$L(x_1,\cdots,x_n;\theta,\lambda) = (1-\theta)^{n_o}\,\theta^{n^*}\prod_{i=1}^{n^*}\frac{e^{-\lambda}\lambda^{x_i}}{(1-e^{-\lambda})\,x_i!} \qquad (9)$$

Taking the natural logarithm of (9) and differentiating with respect to λ, and equating to zero, gives

$$\frac{\partial \ln L}{\partial \lambda} = -\left[n^* + n^* \frac{e^{-\lambda}}{1-e^{-\lambda}} \right] + \sum_{i=1}^{n^*} \frac{x_i}{\lambda} = 0 \qquad (10)$$

Taking the natural logarithm of (9) and differentiating with respect to θ, and equating to zero, gives

$$\frac{\partial \ln L}{\partial \theta} = \frac{n^*}{\theta} - \frac{n_0}{(1-\theta)} = 0 \qquad (11)$$

Solving (10) for the maximum likelihood estimate of λ, $\hat{\lambda}$ we obtain

$$\bar{x}^* = \frac{\lambda}{(1-e^{-\lambda})} \qquad (12)$$

where

$$\bar{x}^* = \frac{\sum_{i=1}^{n^*} x_i}{n^*}$$

Solving (11) for the maximum likelihood estimate of θ, $\hat{\theta}$ we obtain

$$\hat{\theta} = \frac{n^*}{n}$$

Equation (12) is identical to (29) in section 5.4 and can be solved for given $\bar{x}^*$ by consulting the tables in Cohen [14]. Cohen [15] fits (8) to data from Beall and Rescia [3] (the number of European corn - borers on small unit areas of a field as observed in 1937) and demonstrates the superiority of (8) over the complete Poisson distribution in this case.

Now, to compute the asymptotic variances of $\hat{\lambda}$ and $\hat{\theta}$, note that

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n^*}{\theta^2} - \frac{n_0}{(1-\theta)^2}$$

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -n^*\left[\frac{\bar{x}^*}{\lambda^2} - \frac{e^{-\lambda}}{(1-e^{-\lambda})^2}\right]$$

and $$\frac{\partial^2 \ln L}{\partial \theta\,\partial \lambda} = \frac{\partial^2 \ln L}{\partial \lambda\,\partial \theta} = 0 \qquad (13)$$

Then $$V\{\hat{\theta}\} \simeq \theta(1-\theta)/n$$

and $$V\{\hat{\lambda}\} \simeq \lambda\,\chi(\lambda)\,/\,E\{n^*\}$$

where $E\{n^*\} = n\theta$ and $\chi(\lambda) = \dfrac{(1-e^{-\lambda})^2}{1-(\lambda+1)\,e^{-\lambda}}$

From (13) we see that $\hat{\theta}$ and $\hat{\lambda}$ are asymptotically independent. It is also easy to see that

$$\frac{\lambda}{n\theta} \leq V\{\hat{\lambda}\} \leq \frac{2\lambda}{n\theta}$$

Let X be a random variable having p. d. f.

$$h(x;\lambda,\theta) = \begin{cases} 0 & x=0 \\ \dfrac{(1-\theta)\,e^{-\lambda}\lambda}{1-e^{-\lambda}(1+\theta\lambda)} & x=1 \\ \dfrac{e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda}-\theta\lambda e^{-\lambda})\,x!} & x=2,3,\cdots \end{cases} \qquad (14)$$

where $\lambda > 0$ and $0 \leq \theta \leq 1$. With $\theta = 1$, (14) is the p. d. f. of the truncated Poisson distribution with missing zero class. As an example, let a random variable be the number of insect eggs per nest where each nest must have at least one egg. Due to faulty observation, a proportion of ones, say θ, are overlooked or ignored. This random variable than has the distribution given by (14).

Cohen [16] obtains the maximum likelihood estimates for λ and θ, $\hat{\lambda}$ and $\hat{\theta}$. Take a random sample of size n from the population, and let n_1

be the number of ones in the sample. The likelihood function is

$$L(x_1,\cdots,x_n;\lambda,\theta) = \left[\frac{(1-\theta)e^{-\lambda}\lambda}{1-e^{-\lambda}(1+\theta\lambda)}\right]^{n_1} \prod_{*} \frac{e^{-\lambda}\lambda^{x_i}}{\left[1-e^{-\lambda}(1+\theta\lambda)\right] x_i!}$$

where $\prod_{*}$ is the product over all x_i's such that $x_i > 1$. By setting $\frac{\partial \ln L}{\partial \theta}$ and $\frac{\partial \ln L}{\partial \lambda}$ each equal to zero, we obtain the maximum likelihood equations for θ and λ

$$\frac{n_1}{n} = \frac{(1-\theta)e^{-\lambda}\lambda}{1-e^{-\lambda}(1+\theta\lambda)} \tag{15}$$

and

$$\frac{n\bar{x}}{n_1} = \frac{e^{\lambda}-\theta}{1-\theta} \tag{16}$$

where

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Let n^* be the number of observations greater than one, so that $n^* = n - n_1$. Solving (16) for θ gives us the maximum likelihood estimate, $\hat{\theta}$

$$\hat{\theta} = \frac{\left(n\bar{x} - n_1 e^{\hat{\lambda}}\right)}{n^* \ \bar{x}^*} \tag{17}$$

where

$$\bar{x}^* = \frac{\sum_{x=2}^{\infty} x n_x}{n^*}$$

where n_x is the number of observed x values. Substituting (17) into (15) results in

$$\bar{x}^* = \frac{(1-e^{-\hat{\lambda}})}{1-e^{-\hat{\lambda}}(1+\hat{\lambda})} = b(\hat{\lambda}) \tag{18}$$

It is interesting to note that $\hat{\lambda}$ is independent of the number of ones in the sample. Thus, if some ones are known to be missing, the maximum likelihood approach ignores the entire class of ones. To evaluate (18) for $\hat{\lambda}$, Cohen [16] has tabulated $b(\hat{\lambda})$ for $\hat{\lambda} = 0\ (0.10)\ 13.900$ with $b(\hat{\lambda})$ given to 4 decimal places. For quicker evaluation of $\hat{\lambda}$ a folded scale graph of $\delta = \overline{x}^* - \hat{\lambda}$ as a function of $\overline{x}^*$ is plotted. Only a slight sacrifice is made in accuracy. Now, $\hat{\theta}$ may be obtained from (17).

Using standard procedure, Cohen [16] obtains an expression for the asymptotic variance of $\hat{\lambda}$, namely

$$V\{\hat{\lambda}\} \simeq \frac{\lambda}{n^*}\,\chi(\lambda)$$

where

$$\chi(\lambda) = \frac{\left[1 - e^{-\lambda}(1+\lambda)\right]^2}{(1-e^{-\lambda})^2 - \lambda^2 e^{-\lambda}}$$

Cohen [16] tabulates $\chi(\lambda)$ for $\lambda = 0\ (0.1)\ 1.0\ (0.5)\ 5\ (1)\ 10$ and 15. The distribution (14) is fitted to data from Varley giving the number of gall - cells produced in the flower heads of the knapweed by larvae of the knapweed gall - fly in 1936, and a satisfactory fit results.

6.4 A Modified Poisson Distribution

Let X be a random variable having p. d. f.

$$\omega(x; \lambda, \theta) = \begin{cases} e^{-\lambda}(1+\theta\lambda) & x = 0 \\ (1-\theta)\lambda e^{-\lambda} & x = 1 \\ \dfrac{e^{-\lambda}\lambda^x}{x!} & x = 2, 3, \cdots \end{cases} \qquad (19)$$

where $\lambda > 0$ and $0 \leq \theta \leq 1$. As an example, let a random variable be the number of defects present per given unit, and let θ be the probability of misclassifying an item containing one defect by considering it as containing no defects. Then this random variable has distribution (19). For $\theta = 0$, (19) reduces to the complete Poisson p. d. f. .

Cohen [17] determines the maximum likelihood estimates of λ and θ, say $\hat{\lambda}$ and $\hat{\theta}$. Take a random **sample** of size n from the population and let n_0 be the number of zero observations, and n_1, the number of ones observed. The likelihood function is

$$L = \left[e^{-\lambda}(1+\theta\lambda)\right]^{n_0}\left[(1-\theta)\lambda e^{-\lambda}\right]^{n_1} \prod_{*} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}$$

where $\prod_{*}$ is the product over all x_i's that are greater than one. By setting $\frac{\partial \ln L}{\partial \lambda}$ and $\frac{\partial \ln L}{\partial \theta}$ each equal to zero we obtain the maximum likelihood equations

$$\frac{\sum_{i=1}^{n} x_i}{\lambda} + \frac{n_0\theta}{1+\theta\lambda} - n = 0 \tag{20}$$

$$n_0\lambda(1+\theta\lambda) - \frac{n_1}{(1-\theta)} = 0 \tag{21}$$

Eliminating θ from (20) and (21) gives

$$\lambda^2 - \left(\bar{x} - 1 + \frac{n_0}{n}\right)\lambda - \left(\bar{x} - \frac{n_1}{n}\right) = 0 \tag{22}$$

where $\qquad \bar{x} = \sum_{i=1}^{n} x_i \Big/ n$

Simplifying (21), we have

$$\hat{\theta} = \frac{(n_0 - n_1/\lambda)}{n_0 + n_1} \tag{23}$$

Solving the quadratic (22) for λ gives

$$\hat{\lambda} = \frac{1}{2}\left(\bar{x} - 1 + \frac{n_0}{n}\right) + \frac{1}{2}\sqrt{\left(\bar{x} - 1 + \frac{n_0}{n}\right)^2 + 4\left(\bar{x} - \frac{n_1}{n}\right)} \tag{24}$$

The estimate $\hat{\theta}$ may be obtained by substituting (24) into (23).

Cohen [17] determines the asymptotic variances and covariances

$$V\{\hat{\lambda}\} \simeq \frac{\lambda(1+\lambda)}{n(1+\lambda-e^{-\lambda})}$$

$$V\{\hat{\theta}\} \simeq \frac{(1-\theta)(1+\theta\lambda-\theta e^{-\lambda})}{n\lambda e^{-\lambda}(1+\lambda-e^{-\lambda})}$$

and

$$\text{Covariance}(\hat{\lambda},\hat{\theta}) \simeq \frac{(1-\theta)}{n(1+\lambda-e^{-\lambda})}$$

The distribution (19) is fitted by Cohen [17] to data from Bortkiewicz [11] on the number of deaths from the kick of a horse in ten Prussian Army Corps over the twenty years 1875 - 1894, after the data is suitably altered.

Chapter Seven

The Pascal Distribution

7.1 Introduction

In section 3.3 the Pascal, or negative - binomial distribution was introduced and several of its properties developed. The Pascal distribution is an extension of the Poisson series in which the parameter of the Poisson distribution is not constant but varies continuously with a distribution proportional to that of chi - square. In more formal language, the Pascal random variable is a compound Poisson random variable with respect to a chi - square or a gamma compounder. Under certain conditions it has a limiting Poisson distribution. In this chapter, point estimates of the two parameters, k and p, are obtained for both the complete Pascal distribution and the truncated Pascal distribution. Once these estimates have been obtained, the expected frequencies may be computed using the recurrence relation in section 3.3 . Thus, the two major problems in fitting the Pascal distribution to observed data are solved.

7.2 Two Simple Methods of Estimation for the Complete Pascal Distribution

From section 3.3, we notice that a Pascal distribution with parameters k and p, and probability generating function (p. g. f.)

$$(q - pz)^{-k} \tag{1}$$

where $k > 0$, $p > 0$ and $q - p = 1$, has central moments

$$\mu_1' = kp$$

and

$$\mu_2' = kp\left[(k+1)p + 1\right]$$

The method of moments estimates can be obtained as follows. Take a random sample of size n and let n_x be the number of x values observed. Let the first and second sample moments, $\hat{\mu}_1'$ and $\hat{\mu}_2'$, be

$$\hat{\mu}_1' = \sum_{x=0}^{\infty} x n_x \Big/ n$$

and

$$\hat{\mu}_2' = \sum_{x=0}^{\infty} x^2 n_x \Big/ n$$

Then the moment estimates, $\hat{k}$ and $\hat{p}$, are

$$\hat{k} = \frac{\hat{\mu}_1'^{2}}{\hat{\mu}_2' - \hat{\mu}_1'^{2} - \hat{\mu}_1'} \tag{2}$$

and

$$\hat{p} = \hat{\mu}_1' \big/ \hat{k} \tag{3}$$

Equation (3) is a fully efficient equation of estimation. Anscombe [1] has shown that the efficiency of (2) is at least 0.90 for 1) small values of kp when $p < 1/6$, 2) large values of kp when $k > 13$, and 3) moderate values of kp when $\frac{(1 + p)(k + 2)}{p} \geq 15$.

He also shows that the large sample variance of $\hat{k}$ is

$$V\{\hat{k}\} \simeq \frac{2k(k+1)(\hat{k}+\hat{\mu}_1')^2}{\hat{\mu}_1'^2}$$

A second simple method of estimation is the method of the zero proportion. Let n_o be the number of zeroes in the sample. From section 3.3 we have

$$P_o = q^{-k} \tag{4}$$

and

$$\mu_1' = kp \tag{5}$$

where P_o is the probability of obtaining a zero. Equations (4) and (5) suggest that estimates, say $\check{k}$ and $\check{p}$, may be obtained from

$$\check{k}\ \ln\left(1+\frac{\bar{x}}{\check{k}}\right) = \ln\left(\frac{n}{n_o}\right) \tag{6}$$

and

$$\check{p} = \bar{x}/\check{k} \tag{7}$$

Equation (6) may be solved by iteration. Select a trial value for k such that the left side is greater than the constant right side. Then, select a trial value for k such that the left side is less than the right side. Interpolation between these two trial values gives a first approximation of $\check{k}$. The process may be repeated to obtain any desired accuracy. For efficiency at least 0.90, n_o must be greater than or equal to $n/3$. Once $\check{k}$ is obtained, $\check{p}$ is determined from (7). Anscombe [1] has shown that the large sample variance of $\check{k}$ is

$$V\{\check{k}\} \simeq \frac{(1-R)^{-k} - kR - 1}{n\left[-\ln(1-R) - R\right]^2}$$

where

$$R = \frac{\bar{x}}{\check{k}+\bar{x}}$$

7.3 Maximum Likelihood Estimation for the Complete Pascal Distribution

From section 3.3 we note that a random variable x having the Pascal distribution has p. d. f.

$$P_x = \binom{k+x-1}{x} p^x q^{-k-x} \qquad x = 0,1,2,\cdots \quad (8)$$

where $p > 0$, $k > 0$ is an integer, and $q - p = 1$. From (8) we have

$$\frac{\partial \ln P_x}{\partial p} = \frac{x}{p} - \frac{(k+x)}{1+p} \quad (9)$$

and

$$\frac{\partial \ln P_x}{\partial k} = \frac{d \ln (k+x-1)!}{dk} - \frac{d \ln (k-1)!}{dk} - \ln(1+p)$$

$$= F(k+x-1) - F(k-1) - \ln(1+p)$$

$$= \frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{k+x-1} - \ln(1+p) \quad (10)$$

where

$$F(z) = \frac{d \ln z!}{dz} \quad (11)$$

Take a random sample of size n from (8) and let n_x be the number of x values observed. Then, the maximum likelihood estimates, $\hat{p}$ and $\hat{k}$, are the solutions for p and k, respectively, of the following maximum likelihood equations

$$\sum_{x=0}^{\infty} n_x \frac{\partial \ln P_x}{\partial p} = \frac{\sum_{x=0}^{\infty} x n_x}{p(1+p)} - \frac{k}{(1+p)} \sum_{x=0}^{\infty} n_x$$

$$= n \frac{(\bar{x} - kp)}{p(1+p)} = 0 \quad (12)$$

and

$$\sum_{x=0}^{\infty} n_x \frac{\partial \ln P_x}{\partial k} = \sum_{x=0} n_x \left[\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{k+x-1} \right] - n \ln(1+p)$$

$$= \sum_{x=0}^{\infty} \frac{A_x}{(k+x)} - n \ln(1+p) = 0 \quad (13)$$

where
$$A_x = \sum_{j=1}^{\infty} n_{x+j}$$

and
$$\bar{x} = \sum_{x=0}^{\infty} x n_x \Big/ n$$

Simplify (12) and (13) to

$$\bar{x} = \hat{k}\,\hat{p} \tag{14}$$

and
$$\sum_{x=0}^{\infty} \frac{A_x}{\hat{k} + x} = n \ln \left(1 + \frac{\bar{x}}{\hat{k}} \right) \tag{15}$$

Fisher [27] terms the expression

$$z_i = \sum_{x=0}^{\infty} \frac{A_x}{k_i + x} - n \ln \left(1 + \frac{\bar{x}}{k_i} \right) \tag{16}$$

the "score" for the trial value k_i. By trial and error, combined with linear interpolation, a value k, making the score "vanish" is obtained, and this value of k is the maximum likelihood estimate, $\hat{k}$. Then $\hat{p}$ is simply found from (14). Bliss [7] fits the Pascal distribution to data from Garman [30] on the counts of red mites on apple leaves with the parameters estimated using the maximum likelihood technique and a very good fit results.

The sample mean, $\bar{x}$, is the maximum likelihood estimate of the mean kp of the Pascal distribution. Using (1) it is a simple matter to show that the variance of $\bar{x}$, $V(\bar{x})$, is given by

$$V(\bar{x}) = \frac{kp}{n}(p+1)$$

Now, the variance of the maximum likelihood estimate $\hat{k}$

is the reciprocal of the amount of information about k, where, according to Fisher [27], the amount of information about k is the rate at which the score (16) is decreasing as it passes the zero point. A fast way to compute an approximation to the variance is as follows. Suppose that during the trial and error procedure of obtaining $\hat{k}$, z_i and z_{i+1}, are two values of the "score" (16), the first, z_i, being just below zero, and the second, z_{i+1}, just above zero. Let k_i and k_{i+1} be the corresponding trial values of k. Then an approximation to the variance of $\hat{k}$ is

$$V\{\hat{k}\} \simeq \frac{k_i - k_{i+1}}{z_{i+1} - z_i}$$

Bliss [7] maintains that for the cases of "over - dispension" (that is, cases where the sample variance, s^2, exceeds the sample mean, $\bar{x}$) the Pascal distribution is generally more useful than distributions such as the Neyman Type A distribution (Neyman [54]), the Thomas double Poisson distribution (Thomas [79]), and the Polya distribution (Anscombe [1]).

7.4 The U. M. V. U. Estimate of θ^τ

The theory developed by Roy and Mitra [70] for power series distributions and discussed in section 4.4 applies to the Pascal distribution when k is a known positive integer, as in problems of inverse binomial sampling. Now, (8) has the form of (3) in section 4.4

if we let $p = \dfrac{\theta}{(1-\theta)}$ and

$$a(x) = \binom{k+x-1}{x}$$

and $$f(\theta) = (1-\theta)^{-k} .$$

Let $x_1, \ldots, x_n$ be a random sample of size n from (8) with p replaced by $\theta/(1-\theta)$. Let $T = \sum_{i=1}^{n} x_i$, then it can be easily shown that T has p. d. f.

$$P\{T=t\} = \binom{kn+t-1}{t} \theta^t (1-\theta)^{kn} \qquad t=0,1,2,\cdots \tag{17}$$

Then from section 4.4, where $C(t, n)$ and $u_r(t)$ are originally defined, we have

$$C(t,n) = \binom{kn+t-1}{t}$$

and

$$u_r(t) = \begin{cases} 0 & t < r \\ \dfrac{C(t-r,n)}{C(t,n)} & t \geq r \end{cases}$$

$$= \begin{cases} 0 & t < r \\ \dfrac{t^{[r]}}{(kn+t-1)^{[r]}} & t \geq r \end{cases} \tag{18}$$

where $t^{[r]} = t(t-1)\ldots(t-r+1)$. Then $u_r(t)$, given by (18), is the U. M. V. U. estimate of θ^r, where r is a given positive integer. In particular, the U. M. V. U. estimate of θ is

$$u_1(t) = \begin{cases} 0 & t < 1 \\ \dfrac{t}{kn+t-1} & t \geq 1 \end{cases}$$

and the U. M. V. U. estimate of the variance of $u_1(t)$

is, using (9) in section 4.4,

$$\frac{km-t}{(km+t-1)^2(km+t-2)}$$

7.5 Two Moment Methods of Estimation for the Truncated Pascal Distribution

In this section, the simplest and most important case of truncation, truncation away from the zero class, will be examined. If (8) represents the complete Pascal distribution, then the truncated Pascal distribution is represented by

$$P_x' = \binom{k+x-1}{x} \frac{p^x q^{-k-x}}{(1-q^{-k})} \qquad x = 1, 2, \cdots \tag{19}$$

If the central moments of the complete Pascal distribution are denoted by μ_r', the central moments of the truncated Pascal distribution, say μ_{0r}', are given by

$$\mu_{0r}' = \mu_r' / (1-q^{-k}) \tag{20}$$

In particular,

$$\mu_{01}' = kp / (1-q^{-k}) \tag{21}$$

$$\mu_{02}' = kp\left[(k+1)p+1\right] / (1-q^{-k}) \tag{22}$$

and

$$\mu_{03}' = kp\left[(k+1)(k+2)p^2 + 3(k+1)p + 1\right] / (1-q^{-k}) \tag{23}$$

Take a random sample of size n from (19) and let n_x be the number of x values observed, where, of course, $x \geq 1$. Then let

$$\hat{\mu}'_{01} = \frac{\sum_{x=1}^{\infty} x\, n_x}{n} \qquad \hat{\mu}'_{02} = \frac{\sum_{x=1}^{\infty} x^2 n_x}{n}$$

and

$$\hat{\mu}'_{03} = \frac{\sum_{x=1}^{\infty} x^3 n_x}{n}$$

Note that $\hat{\mu}'_{01}$, $\hat{\mu}'_{02}$ and $\hat{\mu}'_{03}$ are consistrnt estimates of μ'_{01}, μ'_{02} and μ'_{03}, respectively. Eliminating k and $(1 - q^{-k})$ from (21), (22) and (23), we obtain the moment estimate, $\hat{p}$, for p,

$$\hat{p} = \frac{\hat{\mu}'_{03} - \hat{\mu}'_{02}}{\hat{\mu}'_{02} - \hat{\mu}'_{01}} - \frac{\hat{\mu}'_{02}}{\hat{\mu}'_{01}} - 1 \tag{24}$$

Now, from (21) and (22) we obtain

$$\frac{\mu'_{02}}{\mu'_{01}} = (\hat{k}+1)\hat{p} + 1 \tag{25}$$

Once $\hat{p}$ has been evaluated using (24), we may evaluate the moment estimate, $\hat{k}$, from (25). This moment method of estimation is simple, yet very inefficient, due to the introduction of the third moment about the origin.

A possible iterative method of moments, using only the first two sample moments, $\hat{\mu}'_{01}$ and $\hat{\mu}'_{02}$, follows. First, we shall rewrite equation (21) in the form

$$q^{-k} = 1 - \frac{kp}{\mu'_{01}} \tag{26}$$

Take the natural logarithm of both sides in (26) so that

$$-k \ln(1+p) = \ln\left(1 - \frac{kp}{\mu'_{01}}\right) \tag{27}$$

Substitute (25) into (27) eliminating k, that is,

$$\frac{(m_1+p)}{p} \ln p = \ln \left(m_2 + \frac{p}{\mu'_{01}} \right) \tag{28}$$

where

$$m_1 = 1 - \frac{\mu'_{02}}{\mu'_{01}}$$

and

$$m_2 = 1 - \frac{m_1}{\mu'_{01}}$$

Rearranging (28) we obtain

$$m_1 = p \, \frac{\ln \left(m_2 + p/\mu'_{01} \right)}{\ln p} - p \tag{29}$$

Now, the estimate for p, say $\check{p}$, may be obtained from (29) by trial and error and linear interpolation, in a similar manner to that used in section 7.2 to solve equation (6). Once $\check{p}$ has been obtained, $\check{k}$, the estimate of k, is obtained from (25). This method yields estimates having higher efficiency than the first method of moments; however, it appears to be a rather tiresome chore to solve (29).

7.6 Maximum Likelihood Estimation for the Truncated Pascal Distribution

As in section 7.5 we shall deal with the Pascal distribution truncated at zero. David and Johnson [21] obtain the maximum likelihood equations for estimating k and p. Denote the estimates by $\hat{k}_0$ and $\hat{p}_0$. Take a random sample of size n from the truncated Pascal distribution represented by (19).

Let n_x be the number of x values observed. Then, the likelihood function is

$$L = \prod_{x=1}^{\infty} P_x'^{\,n_x}$$

so

$$\ln L = \sum_{x=1}^{\infty} n_x \ln P_x'$$

giving

$$\frac{\partial \ln L}{\partial k} = \sum_{x=1}^{\infty} n_x \left[\frac{d \ln (k+x-1)!}{dk} - \frac{d \ln (k-1)!}{dk} - \frac{\ln q}{(1-q^{-k})} \right]$$

$$= \sum_{x=1}^{\infty} n_x \left[\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{k+x-1} \right] - \frac{n \ln q}{(1-q^{-k})} \qquad (30)$$

and

$$\frac{\partial \ln L}{\partial p} = \sum_{x=1}^{\infty} \frac{x n_x}{pq} - \frac{nk}{q(1-q^{-k})} \qquad (31)$$

Note that (30) is similar to equation (10) belonging to the complete case. From (30) and (31) we conclude that the maximum likelihood estimates, $\hat{k}_o$ and $\hat{p}_o$, are the solutions for k and p, respectively, of the following two equations

$$\frac{\sum_{x=1}^{\infty} n_x h(k,x)}{n} = \frac{\ln (1+p)}{(1-q^{-k})} \qquad (32)$$

and

$$\bar{x} = kp \big/ (1-q^{-k}) \qquad (33)$$

where

$$\bar{x} = \sum_{x=1}^{\infty} x n_x \big/ n$$

and

$$h(k,x) = \frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{k+x-1}$$

Then by a very difficult iteration process the estimates, $\hat{k}_o$ and $\hat{p}_o$ may be obtained from (32) and (33).

Hartley [40] has developed a more convenient iterative procedure for obtaining the maximum likelihood estimates. Suppose that n_0 is the unknown number of zeroes missing, so that $N = n_0 + n$ is the "complete" sample size. Using equations (12) and (13), related to the complete case, we define

$$P(k,p) = \frac{(n+n_0)(\bar{x}-kp)}{p(1+p)} \tag{34}$$

and

$$K(k,p) = \sum_{x=0}^{\infty} \frac{A_x}{(k+x)} - (n+n_0)\ln(1+p) \tag{35}$$

where $\bar{x}$ is the mean of a complete sample of size $n + n_0$. Now, the parameters k and p, and also n_0, can be estimated using (34) and (35), and also

$$n_0 = n \; p(0;k,p) \Big/ \big(1 - p(0;k,p)\big) \tag{36}$$

where $p(0;k,p)$ is the probability that a complete Pascal random variable, having parameters k and p, will take on the value, zero. Select, arbitrarily, three "pivotal quantities", which are simple to work with, say $k = 1$, $1/2$ and $1/3$. For each of these values of k, repeated use is made of (34) and (36), as is illustrated for $k = 1/2$ in the following steps.

1) Let $k = 1/2$. Choose a rough estimate of n_0, say ${}_0n_0$, and compute a first estimate of p, say ${}_1p$, from (34), that is

$${}_1p = \frac{\bar{x}}{k} = 2\,\frac{\sum_{x=0}^{\infty} x n_x}{n + {}_0n_0}$$

2) Compute $p(0;\, 1/2,\, {}_1p)$ from (8), that is,

$$p(0;\, 1/2,\, {}_1p) = \frac{1}{(1+{}_1p)^{1/2}}$$

3) Now, compute an "improved" estimate of n_o, say ${}_1n_o$, from

$${}_1n_o = n\,\frac{p(0;\, 1/2,\, {}_1p)}{1-p(0;\, 1/2,\, {}_1p)}$$

4) After 2 1/2 cycles of this procedure have been carried out, we obtain "improving" estimates of n_o, ${}_0n_o$, ${}_1n_o$ and ${}_2n_o$. Let $\delta_1 = {}_1n_o - {}_0n_o$ and $\delta_2 = {}_2n_o - {}_1n_o$. Now, assume that the procedure continues producing further "improving" estimates ${}_3n_o$, ${}_4n_o$, ... with associated differences δ_3, δ_4, ..., defined by $\delta_i = {}_in_o - {}_{i-1}n_o$ for $i = 3, 4, \ldots$. Assume that the differences δ_1, δ_2, δ_3, δ_4, ... form a geometrical progression with constant factor $q = \delta_2/\delta_1$. The sum to infinity of δ_3, δ_4, ... is $\delta_2q/(1-q)$. Thus, the limit of ${}_2n_o$, ${}_3n_o$, ${}_4n_o$, ..., which we shall denote by $n_o(1/2)$, is

$$n_o(1/2) = {}_2n_o + \delta_2 q/(1-q) \tag{37}$$

Compute the final "improved" estimate of n_o, $n_o(1/2)$, from (37).

5) Using $n_o(1/2)$, obtain a final estimate of p, say $\hat{p}(1/2)$, from

$$\hat{p}(1/2) = 2\,\frac{\sum_{x=0}^{\infty} x\,n_x}{n + n_o(1/2)}$$

These 5 steps are repeated for $k = 1$ and $k = 1/3$, and the corresponding estimates of p, $\hat{p}(1)$ and $\hat{p}(1/3)$, are obtained, as well as the estimates of n_o, $n_o(1)$ and $n_o(1/3)$, respectively. For each k and its corresponding $n_o(k)$ and $\hat{p}(k)$, $K(k, p)$ can be evaluated from (35). Thus a table may be formed

k	$\hat{p}(k)$	$K(k, p)$
1	$\hat{p}(1)$	$K(1, p(1))$
1/2	$\hat{p}(1/2)$	$K(1/2, p(1/2))$
1/3	$\hat{p}(1/3)$	$K(1/3, p(1/3))$

By inverse interpolation between k and $K(k, p)$ in the table, a value for k, say $\hat{k}$, can be found making $K(k, p) = 0$. By direct interpolation between k and $\hat{p}(k)$ in the table, a value for $\hat{p}(k)$, corresponding to $\hat{k}$, say $\hat{p}$, can be found. The maximum likelihood estimates are then $\hat{k}$ and $\hat{p}$. A final remark should be made about the selection of the pivotal values for k (here $k = 1$, $1/2$, and $1/3$). Since the pivotal values of k should cover the maximum likelihood estimate, $\hat{k}$, a sound procedure is to compute $\hat{p}(k)$ for $k = 1$ and then depending upon the sign of $K(k, \hat{p}(k))$, choose pivotal values of $k > 1$ or $k < 1$. Estimates of the variances and covariances of $\hat{k}$ and $\hat{p}$ can be obtained by noting that

$$\hat{V}\{\hat{k}\} = -\frac{\partial k}{\partial K(k, p)} \tag{38}$$

$$\hat{V}\left\{\hat{p}\right\} = -\frac{\partial p}{\partial P(k, p)} \tag{39}$$

and

$$\hat{Cov}\left\{\hat{k}, \hat{p}\right\} = -\frac{\partial k}{\partial P(k, p)} = -\frac{\partial p}{\partial K(k, p)} \tag{40}$$

where $\hat{V}(\hat{k})$, $\hat{V}(\hat{p})$ and $\hat{Cov}.(\hat{k}, \hat{p})$ are estimates of $V(\hat{k})$, $V(\hat{p})$ and $Cov.(\hat{k}, \hat{p})$, respectively, where "V" refers to variance and "Cov" refers to covariance. For details, refer to Hartley [40].

Chapter Eight

The Neyman Type A Distribution

8.1 Introduction

The Neyman Type A distribution was introduced in section 3.4 as a generalized Poisson distribution, and a few of its properties were determined. Under certain conditions it has a limiting Poisson distribution. In this chapter, we examine the moment method, and the maximum likelihood method of estimation for the parameters of both the complete Neyman Type A distribution, and the truncated Neyman Type A distribution with the zero class missing.

8.2 The Complete Neyman Type A Distribution

Let λ and λ_1 be the parameters of the complete Neyman Type A distribution, then the method of moments provides simple estimates, $\hat{\lambda}$ and $\hat{\lambda}_1$, of λ and λ_1, respectively. From section 3.4 the first two central moments, μ_1' and μ_2', are

$$\mu_1' = \lambda\lambda_1 \tag{1}$$

and

$$\mu_2' = \lambda\lambda_1(\lambda\lambda_1 + \lambda_1 + 1) \tag{2}$$

Take a random sample of size n from the complete Neyman Type A distribution, and let n_x be the number of x values observed. Let the consistent estimates of μ_1' and μ_2' be

$$\hat{\mu}_1' = \frac{\sum_{x=0}^{\infty} x n_x}{n}$$

and
$$\hat{\mu}_2' = \frac{\sum_{x=0}^{\infty} x^2 n_x}{n}$$

Solving (1) and (2) simultaneously for λ and λ_1, indicates that we take the following, as estimates of λ and λ_1,

$$\tilde{\lambda} = \hat{\mu}_1' / \tilde{\lambda}_1$$

and
$$\tilde{\lambda}_1 = \frac{\hat{\mu}_2'}{\hat{\mu}_1'} - \hat{\mu}_1' - 1$$

Shenton [71] has developed a procedure for obtaining the maximum likelihood estimates, $\hat{\lambda}$ and $\hat{\lambda}_1$ of λ and λ_1, respectively. Let P_x denote the probability that the complete Neyman Type A random variable X takes on the value x. Using the recurrence relation (17) from section 3.4 we have

$$\frac{\partial P_x}{\partial \lambda} = e^{-\lambda_1} \sum_{r=0}^{x} \frac{\lambda_1}{r!} P_{x-r} - P_x$$

$$= \frac{(x+1)}{\lambda\lambda_1} P_{x+1} - P_x \qquad (3)$$

and
$$\frac{\partial P_x}{\partial \lambda_1} = \frac{x P_x}{\lambda_1} - \frac{(x+1)}{\lambda_1} P_{x+1} \qquad (4)$$

From (3) and (4) we obtain the maximum likelihood equations

$$\sum_{x=0}^{\infty} \frac{n_x}{P_x} \frac{\partial P_x}{\partial \lambda} = \sum_{x=0}^{\infty} n_x \frac{(x+1)}{\lambda\lambda_1} \frac{P_{x+1}}{P_x} - \sum_{x=0}^{\infty} n_x = 0 \qquad (5)$$

and
$$\sum_{x=0}^{\infty} \frac{n_x}{P_x} \frac{\partial P_x}{\partial \lambda_1} = \sum_{x=0}^{\infty} \frac{x n_x}{\lambda_1} - \sum_{x=0}^{\infty} n_x \frac{(x+1)}{\lambda_1} \frac{P_{x+1}}{P_x} = 0 \qquad (6)$$

where n_x is the number of x values observed in a sample of size n taken from the complete Neyman Type A distribution. Let $\bar{x}$ be the sample mean and put

$$\pi_x = (x+1)\frac{P_{x+1}}{P_x} \tag{7}$$

then equations (5) and (6) reduce to

$$\sum_{x=0}^{\infty} n_x \pi_x = \lambda\lambda_1 n \tag{8}$$

and

$$\sum_{x=0}^{\infty} n_x \pi_x = n\bar{x} \tag{9}$$

Equations (8) and (9) can be combined more advantageously, so that

$$\bar{x} = \lambda\lambda_1 \tag{10}$$

and

$$\sum_{x=0}^{\infty} n_x \pi_x = n\bar{x} \tag{11}$$

Equations (10) and (11) must be solved for the estimates, $\hat{\lambda}$ and $\hat{\lambda}_1$. Shenton [71] defines

$$F(\lambda_1) = \sum_{x=0}^{\infty} n_x\pi_x - n\bar{x} \tag{12}$$

where λ is considered to be eliminated by (10). Then differentiating $F(\lambda_1)$ with respect to λ_1 gives

$$F^{(1)}(\lambda_1) = \sum_{x=0}^{\infty} \frac{n_x}{P_x^2}(x+1)\left[P_x\frac{dP_{x+1}}{d\lambda_1} - P_{x+1}\frac{dP_x}{d\lambda_1}\right] \tag{13}$$

Now,

$$\begin{aligned}\frac{dP_x}{d\lambda_1} &= \frac{\partial P_x}{\partial\lambda}\frac{d\lambda}{d\lambda_1} + \frac{\partial P_x}{\partial\lambda_1}\\ &= -\frac{\bar{x}}{\lambda_1^2}\left[(x+1)\frac{P_{x+1}}{\lambda\lambda_1} - P_x\right] + \frac{xP_x}{\lambda_1} - \frac{(x+1)P_{x+1}}{\lambda_1}\end{aligned} \tag{14}$$

Substituting (14) into (13) gives

$$F^{(1)}(\lambda_1) = \sum_{x=0}^{\infty}\frac{n_x\pi_x}{\lambda_1} - \frac{(\lambda_1+1)}{\lambda_1^2}\sum_{x=0}^{\infty} n_x\phi_x \tag{15}$$

where

$$\phi_x = \pi_x(\pi_{x+1} - \pi_x) \tag{16}$$

The maximum likelihood estimate $\hat{\lambda}_1$ may be obtained as follows. Let $\hat{\lambda}_{1,i}$ be an i^{th} estimate of $\hat{\lambda}_1$, then a closer approximation is $\hat{\lambda}_{1,i+1}$, given by

$$\hat{\lambda}_{1,i+1} = \hat{\lambda}_{1,i} - F(\hat{\lambda}_{1,i}) \Big/ F^{(1)}(\hat{\lambda}_{1,i}) \tag{17}$$

The initial estimate, $\hat{\lambda}_{1,1}$, may be determined using the method of moments. Repeated use of (17) is made until there is little change in $\hat{\lambda}_{1,i}$. The final value for $\hat{\lambda}_{1,i}$ will be a good approximation to the maximum likelihood estimate, $\hat{\lambda}_1$. Once $\hat{\lambda}_1$ is obtained, the maximum likelihood estimate, $\hat{\lambda}$, is obtained from (10).

Equation (17) is a very tedious equation to work with. Douglas [22] has constructed tables which considerably shorten the work. Let

$$\alpha = \lambda e^{-\lambda_1}$$

then P_o and P_x from (16) in section 4.3 are

$$P_o = e^{\alpha - \lambda}$$

and

$$P_x = e^{-\lambda} \frac{\lambda_1^x}{x!} \sum_{r=0}^{\infty} \frac{r^x}{r!} \alpha^r \qquad x > 0$$

$$= P_o \frac{\lambda_1^x}{x!} \mu'_x \tag{19}$$

where

$$\mu'_x = e^{-\alpha} \sum_{r=0}^{\infty} r^x \frac{\alpha^r}{r!}$$

If we let $p_x = \dfrac{\mu'_{x+1}}{\mu'_x}$, then from (19) we may obtain the relation

$$P_{x+1} = \frac{\lambda_1}{(x+1)} P_x \, p_x \tag{20}$$

Using (20) the maximum likelihood equations (10) and (11)

become
$$\hat{\lambda}\hat{\lambda}_1 = \bar{x}$$

and
$$\hat{\lambda} = \sum_{x=0}^{\infty} n_x p_x \Big/ n$$

Now, from (12) and (13) we have
$$F(\lambda_1) = \lambda_1 \sum_{x=0}^{\infty} n_x p_x - n\bar{x} \tag{21}$$

and
$$F^{(1)}(\lambda_1) = \sum_{x=0}^{\infty} n_x p_x - (1+\lambda_1) \sum_{x=0}^{\infty} n_x q_x \tag{22}$$

where $q_x = p_x(p_{x+1} - p_x)$. Douglas [22] has tabulated p_x and q_x for $\alpha = 0.000\ (0.001)\ 0.03\ (0.01)\ 0.3\ (0.1)\ 3.0$ and $x = 0\ (1)\ 19$. Using equations (17) and (18) in (17) instead of (12) and (15), speeds us the calculation of $\hat{\lambda}_1$ considerably. Douglas [22] fits the complete Neyman Type A distribution to data on the European Corn-Borer, given in Neyman [54], and a good fit results.

8.3 The Truncated Neyman Type A Distribution

We consider the special case of truncation where the zero class is missing. If P_x is the complete Neyman Type A probability, then the truncated Neyman Type A probability is
$$P_x' = \frac{P_x}{1-P_0} \qquad x = 1, 2, \cdots \tag{23}$$

The method of moments estimates, $\check{\lambda}_0$ and $\check{\lambda}_{01}$, of the parameters λ and λ_1, are obtained in a manner similar to that of the complete case. That is,
$$\check{\lambda}_{01} = \frac{\hat{\mu}_2'}{\hat{\mu}_1'} - \hat{\mu}_1' - 1 \tag{24}$$

and

$$\hat{\lambda}_0 = \hat{\mu}_1' / \hat{\lambda}_{01} \tag{25}$$

where
$$\hat{\mu}_1' = \sum_{x=1}^{\infty} x n_x \Big/ n$$

and
$$\hat{\mu}_2' = \sum_{x=1}^{\infty} x^2 n_x \Big/ n$$

Sometimes equations (24) and (25) lead to negative estimates, so that, $\hat{\lambda}_{01}$ and $\hat{\lambda}_0$ are only obtained when rough ideas as to the values of the parameters, λ_1 and λ, are desired.

Along lines similar to those used in the complete case we obtain the maximum likelihood equations

$$\frac{\hat{\lambda}_0 \hat{\lambda}_{01}}{1-\hat{P}_0} = \bar{x}' \tag{26}$$

and
$$1 - \hat{P}_0 e^{-\hat{\lambda}_{01}} = \sum_{x=1}^{\infty} n_x \hat{\pi}_x \Big/ n\bar{x}' \tag{27}$$

where $\hat{\lambda}_0$ and $\hat{\lambda}_{01}$ are the maximum likelihood estimates of λ and λ_1, respectively, and $\hat{P}_0$ and $\hat{\pi}_x$ are the expressions, P_0 and $_x$, with λ and λ_1 replaced by $\hat{\lambda}_0$ and $\hat{\lambda}_{01}$, respectively, and finally, where

$$\bar{x}' = \sum_{x=1}^{\infty} x n_x \Big/ n$$

Write
$$F(\lambda_1) = 1 - P_0 e^{-\lambda_1} - \frac{\sum_{x=1}^{\infty} n_x \pi_x}{n\bar{x}'} \tag{28}$$

where we have considered λ to be eliminated by use of (26). Then, differentiating with respect to λ_1, gives

$$F^{(1)}(\lambda_1) = e^{-\lambda_1} P_0 \left[1 + \lambda e^{-\lambda_1} - (1-e^{-\lambda_1}) \frac{\lambda}{\lambda_1} \left\{ \frac{1-\lambda\lambda_1 e^{-\lambda_1} P_0'}{1-\lambda(1-e^{-\lambda_1}) P_0'} \right\} \right]$$
$$+ \frac{1}{n\lambda_1 \bar{x}'} \left[1 + \frac{1}{\lambda_1} \left\{ \frac{1-\lambda\lambda_1 e^{-\lambda_1} P_0'}{1-\lambda(1-e^{-\lambda_1}) P_0'} \right\} \right] \sum_{x=1}^{\infty} n_x \phi_x$$
$$- \frac{1}{n\lambda_1 \bar{x}'} \sum_{x=1}^{\infty} n_x \pi_x \tag{29}$$

where $P_o' = \frac{P_o}{1 - P_o}$ and both π_x and ϕ_x are as in the complete case. Thus, if $\hat{\lambda}_{01,i}$ is an i^{th} estimate of $\hat{\lambda}_{01}$, a closer approximation is, $\hat{\lambda}_{01,i+1}$, given by

$$\hat{\lambda}_{01,i+1} = \hat{\lambda}_{01,i} - F(\hat{\lambda}_{01,i}) \Big/ F^{(1)}(\hat{\lambda}_{01,i}) \tag{30}$$

If $\hat{\lambda}_{0,i}$ is an i^{th} estimate of $\hat{\lambda}_0$, a closer approximation is, $\hat{\lambda}_{0,i+1}$, given by

$$\hat{\lambda}_{0,i+1} = \frac{\bar{x}'}{\hat{\lambda}_{01,i+1}} \left[1 - e^{-\hat{\lambda}_{0,i}(1 - e^{-\hat{\lambda}_{01,i+1}})} \right] \tag{31}$$

The work involved in using (30) and (31) is much greater than that involved in using (12) and (13) in the complete case. To shorten the amount of labour Douglas [ ] has introduced a procedure similar to that used in the complete case. Omitting the details, the maximum likelihood equations are expressed as

$$\hat{\lambda}_0 \hat{\lambda}_{01} \big/ (1 - \hat{P}_0) = \bar{x}'$$

and

$$\frac{1 - \hat{P}_0 e^{-\hat{\lambda}_{01}}}{\hat{\lambda}_{01}} = \sum_{x=1}^{\infty} \frac{n_x \hat{p}_x}{n \bar{x}'}$$

where $\hat{p}_x$ is p_x with λ and λ_1 replaced by $\hat{\lambda}_0$ and $\hat{\lambda}_{01}$, respectively. Also (28) and (29) are expressed as

$$F(\lambda_1) = 1 - P_0 e^{-\lambda_1} - \frac{\lambda_1}{n\bar{x}'} \sum_{x=1}^{\infty} n_x p_x$$

and

$$F^{(1)}(\lambda_1) = e^{-\lambda} P_0 \left[1 + \lambda e^{-\lambda_1} - (1 - e^{-\lambda_1}) \frac{\lambda}{\lambda_1} \left\{ \frac{1 - \lambda\lambda_1 e^{-\lambda_1} P_0'}{1 - \lambda(1 - e^{-\lambda_1}) P_0'} \right\} \right]$$

$$+ \frac{1}{n\bar{x}'} \left[\lambda_1 + \left\{ \frac{1 - \lambda\lambda_1 e^{-\lambda_1} P_0'}{1 - \lambda(1 - e^{-\lambda_1}) P_0'} \right\} \right] \sum_{x=1}^{\infty} n_x q_x$$

$$- \frac{1}{n\bar{x}'} \sum_{x=1}^{\infty} n_x p_x$$

Again the tables of p_x and q_x provide a more efficient means of obtaining the maximum likelihood estimates. To illustrate the procedure Douglas [22] fits the truncated Neyman Type A distribution to data of leaf counts supplied by Goodall [32], and obtains a good fit ($\chi_4^2 = 6.8$).

A procedure similar to that given in section 7.6 (Hartley [40]) may be used to determine the maximum likelihood estimates of λ and λ_1.

Chapter Nine

The Poisson v Binomial Distribution

9.1 Introduction

In section 3.5, the Poisson v Binomial distribution was introduced and a few of its properties determined. Under certain conditions it has a limiting Poisson distribution. There are two major problems in fitting a fairly complicated distribution : 1) point estimation of the parameters, and 2) determination of the expected frequencies using the estimates of the parameters. Both of these problems are treated in this chapter.

9.2 Simple Methods of Estimation

Let X be a random variable having the Poisson Binomial distribution with parameters λ and p and probabilities, P_x, given by

$$P_x = \sum_{r=0}^{\infty} \binom{nr}{x} p^x q^{nr-x} e^{-\lambda} \frac{\lambda^r}{r!} \tag{1}$$

where $\lambda > 0$, $p > 0$, $q > 0$ and $p + q = 1$, and n is a positive integer. From section 3.5, the first two central moments are

$$\mu_1' = \lambda p n \tag{2}$$

and

$$\mu_2' = \lambda p n \left[\lambda p n + p(n-1) + 1 \right] \tag{3}$$

Let a random sample of size N be taken from (1) and let n_x be the number of x values observed. Let

$$\hat{\mu}_1' = \sum_{x=0}^{\infty} x n_x \Big/ N$$

and $$\hat{\mu}_2' = \sum_{x=0}^{\infty} x^2 n_x \Big/ N$$

Assume that n is known. Solving (2) and (3) for and p, we obtain the method of moments estimates, $\check{\lambda}$ and $\check{p}$, for λ and p, respectively,

$$\check{\lambda} = \frac{(n-1)\ \hat{\mu}_1'^2}{\left(\hat{\mu}_2' - \hat{\mu}_1'^2 - \hat{\mu}_1'\right) n}$$

and
$$\check{p} = \frac{\hat{\mu}_2' - \hat{\mu}_1'^2 - \hat{\mu}_1'}{(n-1)\ \hat{\mu}_1'}$$

A second simple method of estimation is the method of sample zero frequency. Let n_o be the number of zeroes in the sample. If $\check{\lambda}$ and $\check{p}$ are the estimates of λ and p, respectively, then they are obtained from

$$\check{\lambda}\check{p} = \hat{\mu}_1' / n$$

and
$$n_0 = N e^{-\check{\lambda}\left[1 - (1-\check{p})^n\right]}$$

McGuire, Brindley and Bancroft [49] remark that useful values of n are n = 2, 3 or 4, since the Poisson v Binomial distribution approaches the Neyman Type A distribution rapidly as n increases and p decreases. Sprott [73] investigates the efficiency of both these simple methods for n = 2. Only for very small p is the efficiency of the method of moments high. The efficiency of the method of sample zero frequency is around 0.90 for $p \leq 0.3$ but much lower for other values of p. For both methods the efficiency approaches zero as p approaches one. These simple, relatively inefficient, methods are useful for providing initial estimates in long iterative procedures.

9.3 Maximum Likelihood Estimation

Sprott [73] has developed the following maximum likelihood procedure for obtaining the estimates, $\hat{\lambda}$ and $\hat{p}$, of λ and p, respectively. We assume that the value of n is known. From section 3.5 the probabilities, P_x, satisfy the recurrence relation

$$P_{x+1} = \frac{n\lambda p}{(x+1)} \sum_{\tau=0}^{x} \binom{n-1}{\tau} p^{\tau} q^{n-\tau-1} P_{x-\tau} \tag{4}$$

Put
$$S_1(x) = \sum_{\tau=0}^{\infty} \frac{\lambda^{\tau}}{\tau!} \binom{n\tau}{x} p^{x} q^{n\tau-x} \tag{5}$$

Then
$$S_2(x) = \frac{\partial S_1(x)}{\partial \lambda} = \sum_{\tau=0}^{\infty} \frac{\lambda^{\tau-1}}{(\tau-1)!} \binom{n\tau}{x} p^{x} q^{n\tau-x}$$

$$= \frac{(x+1)}{n\lambda p} q\, S_1(x+1) + \frac{x}{n\lambda} S_1(x) \tag{6}$$

Similarly, it can be shown that

$$\frac{\partial S_1(x)}{\partial p} = \frac{x}{p} S_1(x) - \frac{(x+1)}{p} S_1(x+1) \tag{7}$$

Now $S_1(x) = e^{\lambda} P_x$, so that $\ln P_x = \ln S_1(x) - \lambda$. Take a random sample of size N from (1) and let n_x be the number of x values observed. The maximum likelihood equations are

$$\sum_{x=0}^{\infty} n_x \frac{\partial \ln P_x}{\partial \lambda} = -N + \sum_{x=0}^{\infty} n_x \frac{S_2(x)}{S_1(x)} = 0 \tag{8}$$

and $$\sum_{x=0}^{\infty} n_x \frac{\partial \ln P_x}{\partial p} = \sum_{x=0}^{\infty} \frac{n_x}{S_1(x)} \frac{\partial S_1(x)}{\partial p} = 0 \tag{9}$$

Equations (8) and (9) reduce to

$$m \hat{\lambda} \hat{p} = \bar{x} \tag{10}$$

and $$L(\hat{p}) = \sum_{x=0}^{\infty} n_x F(x) - N = 0 \tag{11}$$

where $$F(x) = \frac{(x+1)}{m \hat{\lambda} \hat{p}} \frac{P_{x+1}}{P_x} \tag{12}$$

and $$\bar{x} = \sum_{x=0}^{\infty} x n_x \Big/ N$$

To evaluate the maximum likelihood estimates, $\hat{\lambda}$ and $\hat{p}$, is a rather long procedure. First, note that from (10) and (11) that

$$\hat{q} L(\hat{p}) = \sum_{x=0}^{\infty} n_x \frac{S_2(x)}{S_1(x)} - N \tag{13}$$

Thus if we let $L^{(1)}(p) = \frac{d L(p)}{dp}$, we have

$$\hat{q} L^{(1)}(\hat{p}) = \sum_{x=0}^{\infty} n_x \left[\frac{S_1(x) \frac{dS_2(x)}{d\hat{p}} - S_2(x) \frac{dS_1(x)}{d\hat{p}}}{[S_1(x)]^2} \right] + L(\hat{p}) \tag{14}$$

Using the method of deriving (6) and noticing from (10) that

$$\frac{d\hat{\lambda}}{d\hat{p}} = -\frac{\hat{\lambda}}{\hat{p}}$$

we have the following,

$$\frac{dS_1(x)}{dp} = -\frac{(x+1)}{p}\left(1 + \frac{q}{mp}\right) e^{\lambda} P_{x+1} + \frac{x}{p}\left(1 - \frac{1}{m}\right) e^{\lambda} P_x \tag{15}$$

and $$\frac{dS_2(x)}{dp} = -\left(1 + \frac{q}{mp}\right) q \frac{(x+1)(x+2)}{m\lambda p^2} e^{\lambda} P_{x+2}$$
$$+ \left[\frac{(x+1)(q-p)}{p} - (2x+1)\frac{q}{mp} \right] \frac{(x+1)e^{\lambda} P_{x+1}}{m\lambda p}$$
$$+ \left[1 + x - \frac{x}{m} \right] \frac{x}{m\lambda p} e^{\lambda} P_x \tag{16}$$

Now, substitute (15) and (16) into (14), making use of (12), so that

$$L^{(1)}(\hat{p}) = \sum_{x=0}^{\infty} n_x F(x)\left[\frac{1}{\hat{p}} - \frac{1}{n\hat{p}} - \left(1 + \frac{\hat{q}}{n\hat{p}}\right) n\hat{\lambda}\,\Delta F(x)\right] \quad (17)$$

Now, the maximum likelihood estimate $\hat{p}$ may be evaluated. Let $\hat{p}_i$ be the i^{th} estimate of $\hat{p}$, then a closer approximation to $\hat{p}$ is $\hat{p}_{i+1}$, given by Newton's formula,

$$\hat{p}_{i+1} = \hat{p}_i - L(\hat{p}_i) \Big/ L^{(1)}(\hat{p}_i) \quad (18)$$

By repeated application of (18) we obtain as good an approximation for $\hat{p}$ as desired. The rough initial estimate, $\hat{p}_1$, may be obtained by either the method of moments or the method of sample zero frequency. If the number of zeroes in the sample is large the latter method is favoured. Once $\hat{p}$ has been evaluated, $\hat{\lambda}$ may be determined from (10). Finally, the probabilities P_x can be determined recursively using (4). Thus expected frequencies may be computed and the Poisson $\vee$ Binomial distribution fitted to the data. Sprott [73] fits the Poisson $\vee$ Binomial distribution to data from McGuire et al [49] with the parameters estimated by the method of moments ($\chi^2_{(5)} = 18.90$), the method of sample zero frequency ($\chi^2_{(5)} = (.40)$ and the maximum likelihood method ($\chi^2_{(5)} = 9.88$). The implication of the paper is that for small p, say less than 0.3, the method of sample zero frequency may be used, and the maximum likelihood method for larger values of p.

Sprott [73] determines expressions for the variance and covariance of the maximum likelihood estimates, $\hat{\lambda}$ and $\hat{p}$. Details are omitted here. If "V" refers to variance and "Cov", to covariance,

$$V\{\hat{p}\} = \frac{I_{\lambda\lambda}}{D_{\lambda p} N^2} \qquad V\{\hat{\lambda}\} = \frac{I_{pp}}{D_{\lambda p} N^2}$$

and

$$\mathrm{Cov}(\hat{\lambda}, \hat{p}) = \frac{-I_{\lambda p}}{D_{\lambda p} N^2}$$

where

$$\frac{I_{\lambda\lambda}}{N} = q^2 A + \frac{mp + (m-1)pq}{m\lambda}$$

and

$$\frac{I_{pp}}{N} = m^2 \lambda^2 A + m\lambda \left(1 - m + \frac{1}{p}\right)$$

$$\frac{I_{\lambda p}}{N} = -m\lambda q A + mq + p \ , \qquad A = \sum_{x=0}^{\infty} F(x)^2 P_x - 1$$

and

$$D_{\lambda p} = m\lambda \left(m + \frac{q}{p}\right) A - (m-1)^2$$

Gurland and Shumway [38] have developed a simpler procedure for obtaining the maximum likelihood estimates and computing the Poisson v Binomial probabilities than the procedure suggested by Sprott [73]. He suggests using the recurrence relation (4) to determine the probabilities. The argument against this idea is twofold : 1) the computation of successive probabilities using (4) is tedious because the formula is long and each probability depends on all the preceding ones, and 2) any errors made along the way are carried by all the succeeding probabilities.

First, a simpler recurrence relation than (4) can be developed as follows. Rewrite (1) in the form

$$P_x = e^{-\lambda} \frac{p^x}{q^x} \frac{1}{x!} \sum_{x=0}^{\infty} (nr)^{[x]} \frac{\alpha^r}{r!}$$

$$= e^{\alpha-\lambda} \frac{p^x}{q^x} \frac{\mu_{[x]}}{x!} \tag{19}$$

where $\alpha = \lambda q^n$ and $(nr)^{[x]} = nr(nr-1)\cdots(nr-x+1)$

and

$$\mu_{[x]} = e^{-\alpha} \sum_{r=0}^{\infty} (nr)^{[x]} \frac{\alpha^r}{r!} \tag{20}$$

From (19) and (20) we obtain

$$P_{x+1} = \frac{p}{q} \frac{P_x}{(x+1)} p_{[x]} \tag{21}$$

where

$$p_{[x]} = \mu_{[x+1]} \big/ \mu_{[x]} \tag{22}$$

From (20) we notice that it is possible to consider $\mu_{[x]}$ as the x^{th} factorial moment of the random variable nr where r is a Poisson random variable with parameter α.

Thus,

$$\mu_{[x]} = E\left\{(nr)^{[x]}\right\} = E\left\{\sum_{i=1}^{x} s(x,i)\, n^i r^i\right\}$$

$$= \sum_{i=1}^{x} s(x,i)\, n^i\, E\left\{r^i\right\}$$

$$= \sum_{i=1}^{x} s(x,i)\, n^i\, E\left\{\sum_{j=1}^{i} S(i,j)\, r^{[j]}\right\}$$

$$= \sum_{i=1}^{x} \sum_{j=1}^{i} s(x,i)\, n^i\, S(i,j)\, \alpha^j \tag{23}$$

where the $s(x,i)$ are Stirling numbers of the first kind and the $S(i,j)$ are Stirling numbers of the second kind. Now, the numbers $s(x,i)$ and $S(i,j)$ can easily be obtained from the recurrence relations

$$s(x+1,i) = s(x,i-1) - x \quad s(x,i)$$

and
$$S(i+1,j) = S(i,j-1) + j\,S(i,j)$$

Stirling numbers are defined and discussed in Riordan [69] and Richardson [66]. For given values of n, α and x, $\mu_{[x]}$ may be evaluated from (23). Then $p_{[x]}$ may be determined from (22). The $p_{[x]}$ have been tabulated by Gurland and Shumway [38] for $n = 2$, $x = 0\ (1)\ 9$ and $\alpha = 0.10\ (0.02)\ 1.10$. Then (21) is the simpler recurrence relation promised. It is possible to calculate each probability from the one immediately preceding it, and also, to do so with little labour.

The computation of probabilities is, of course, only possible when we have numerical values for the parameters, λ and p. A shorter procedure for obtaining the maximum likelihood estimates, $\hat{\lambda}$ and $\hat{p}$, follows. Let

$$q_{[x]} = p_{[x]}\left(p_{[x+1]} - p_{[x]}\right)$$

then (17) may be rewritten in the form

$$L^{(1)}(\hat{p}) = \frac{1}{n^2\hat{\lambda}\hat{p}\hat{q}}\left[(n-1)\sum_{x=0}^{\infty} n_x\, p_{[x]} - \left\{\frac{(n-1)\hat{p}+1}{\hat{q}}\right\}\sum_{x=0}^{\infty} n_x\, q_{[x]}\right] \qquad (24)$$

Gurland and Shumway [38] have also tabulated $q_{[x]}$ for the same values of n, x and α as $p_{[x]}$. Thus equation (24) is easier to manipulate than Sprott's equation (17). As before the estimates $\hat{\lambda}$ and $\hat{p}$ are obtained from (18) and (10).

To compute the variances and covariance of $\hat{\lambda}$ and $\hat{p}$, Sprott [73] has shown that the quantity

$$A = \sum_{x=0}^{\infty} \left[F(x)\right]^2 P_x - 1 \tag{25}$$

must be computed. Rewrite (25) in the form

$$A = \frac{1}{m^2 \lambda^2 q^2} \sum_{x=0}^{\infty} P_{[x]}^2 P_x - 1$$

Thus A may be determined using the tables and the estimates of the parameters.

The probabilities P_x may also be computed by a procedure involving matrices developed by Gurland and Shumway [37]. Equation (23) may be written in matrix notation as

$$\mu'_{[\ ]} = S N \underset{\sim}{S} \Lambda$$

where

$$\mu'_{[\ ]} = \begin{pmatrix} \mu'_{[1]} \\ \mu'_{[2]} \\ \vdots \\ \mu'_{[m]} \end{pmatrix} \qquad N = \begin{pmatrix} m & 0 & \cdots & 0 \\ 0 & m^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & m^m \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} \alpha \\ \alpha^2 \\ \vdots \\ \alpha^m \end{pmatrix} \qquad S = \begin{pmatrix} s(1,1) & 0 & \cdots & 0 \\ s(2,1) & s(2,2) & & \vdots \\ \vdots & & \ddots & \\ s(m,1) & \cdots & & s(m,m) \end{pmatrix} \tag{26}$$

and

$$\underset{\sim}{S} = \begin{pmatrix} S(1,1) & 0 & \cdots & 0 \\ S(2,1) & S(2,2) & & \vdots \\ \vdots & & \ddots & \\ S(m,1) & \cdots & & S(m,m) \end{pmatrix}$$

Write $A = SN\underset{\sim}{S}$ and $P = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{pmatrix}$ then $P = BA\Lambda$

where $B_r = P_0 \dfrac{p^r}{q^r} \dfrac{1}{r!}$ and

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & B_m \end{pmatrix} \tag{27}$$

and $P_o = e^{\alpha - \lambda}$. The matrix A has the triangular form

$$A = \begin{pmatrix} A_1^1 & 0_2 & \cdots & 0 \\ A_1^2 & A_2^2 & \ddots & \vdots \\ \vdots & & & \\ A_1^m & \cdots & & A_m^m \end{pmatrix}$$

Gurland and Shumway [37] have tabulated the matrix A for n = 2, 3, 4 and m = 10. To compute the probabilities P_x, first obtain the matrix B from (28) and Λ from (26). Then matrix P may be obtained from (27). This method also offers a great improvement in speed and accuracy over the direct use of the recurrence relation (4).

9.4 The Minimum Chi-Square Method

Gurland and Katti [35] consider the minimum chi-square method of estimating the two parameters, λ and p, of the Poisson $\vee$ Binomial distribution. We again assume that n is known. The most useful values of n are small, say 2, 3 and 4, since the Poisson $\vee$ Binomial distribution rapidly approaches the Neyman Type A distribution with increasing n. Suppose that τ_1, τ_2 and τ_3 are different functions of λ and p, having consistent estimates, t_1, t_2 and t_3, respectively. Let τ and t be the matrices $(\tau_1 \ \ \tau_2 \ \ \tau_3)$ and $(t_1 \ \ t_2 \ \ t_3)$, respectively. Barankin and Gurland [2] show that the estimates t which minimize

$$Q = (t-\tau)\hat{\Omega}^{-1}(t-\tau)' \tag{29}$$

are asymptotically the best, where $\hat{\Omega}$ is a consistent estimate of the covariance matrix, Ω, of t, and $(t-\tau)'$ is the transpose matrix of $(t-\tau)$. After differentiating, the estimation equations are

$$\frac{\partial \tau}{\partial \lambda} \hat{\Omega}^{-1} (t-\tau)' = 0 \tag{30}$$

and

$$\frac{\partial \tau}{\partial p} \hat{\Omega}^{-1} (t-\tau)' = 0 \tag{31}$$

The solutions of (30) and (31) for λ and p, are the minimum chi-square estimates, denoted by $\hat{\lambda}$ and $\hat{p}$, respectively. We shall now sketch the method of obtaining the asymptotic generalized variance of the estimates, $\hat{\lambda}$ and $\hat{p}$. From (30) and (31) we have

$$\left[\frac{\partial \tau}{\partial \lambda}\right]_{\hat{\lambda},\hat{p}} \hat{\Omega}^{-1} \left[(t-\tau)'\right]_{\hat{\lambda},\hat{p}} = 0 \tag{32}$$

and

$$\left[\frac{\partial \tau}{\partial p}\right]_{\hat{\lambda},\hat{p}} \hat{\Omega}^{-1} \left[(t-\tau)'\right]_{\hat{\lambda},\hat{p}} = 0 \tag{33}$$

Expanding $(t-\tau)'$ in powers of $(\hat{\lambda}-\lambda)$ and $(\hat{p}-p)$, and neglecting all second order and higher terms, we obtain asymptotically,

$$\frac{\partial \tau}{\partial \lambda} \Omega^{-1} (t-\tau)' = \frac{\partial \tau}{\partial \lambda} \Omega^{-1} \frac{\partial \tau'}{\partial(\lambda,p)} \begin{pmatrix} \hat{\lambda}-\lambda \\ \hat{p}-p \end{pmatrix} \tag{34}$$

and

$$\frac{\partial \tau}{\partial p} \Omega^{-1} (t-\tau)' = \frac{\partial \tau}{\partial p} \Omega^{-1} \frac{\partial \tau'}{\partial(\lambda,p)} \begin{pmatrix} \hat{\lambda}-\lambda \\ \hat{p}-p \end{pmatrix} \tag{35}$$

where the matrix $\frac{\partial \tau'}{\partial(\lambda,p)}$ is more explicitly written as

$$\frac{\partial \tau'}{\partial(\lambda,p)} = \begin{pmatrix} \frac{\partial \tau_1}{\partial \lambda} & \frac{\partial \tau_1}{\partial p} \\ \frac{\partial \tau_2}{\partial \lambda} & \frac{\partial \tau_2}{\partial p} \\ \frac{\partial \tau_3}{\partial \lambda} & \frac{\partial \tau_3}{\partial p} \end{pmatrix}$$

The generalized variance G is then

$$G = \left| E\left\{ \begin{pmatrix} \hat{\lambda}-\lambda \\ \hat{p}-p \end{pmatrix} \begin{pmatrix} \hat{\lambda}-\lambda & \hat{p}-p \end{pmatrix} \right\} \right| = \left| \frac{\partial \tau}{\partial(\lambda,p)} \Omega^{-1} \frac{\partial \tau'}{\partial(\lambda,p)} \right|^{-1}$$

Suppose that we select the first two factorial cumulants, $K_{[1]}$ and $K_{[2]}$, and the natural logarithm of the zero proportion, $\ln P_0$, as particular cases of Υ_1, Υ_2 and Υ_3 respectively. Let $\hat{K}_{[1]}$, $\hat{K}_{[2]}$ and the $\ln \hat{P}_0$ be consistent estimates of $K_{[1]}$, $K_{[2]}$ and $\ln P_0$, respectively, so that the matrix $t = (\hat{K}_{[1]} \; \hat{K}_{[2]} \; \ln \hat{P}_0)$ is a consistent estimate of the matrix $\Upsilon = (K_{[1]} \; K_{[2]} \; \ln P_0)$. From section 3,5 we have that for the Poisson $\vee$ Binomial distribution

$$K_{[1]} = \lambda p n \qquad\qquad K_{[2]} = \lambda p^2 n(n-1) \qquad (36),(37)$$

and

$$\ln P_0 = \lambda (q^n - 1) \qquad (38)$$

Using (36), (37) and (38) the estimation equations (30) and (31) become

$$\lambda = \frac{(np \quad n(n-1)p^2 \quad q^n - 1)\,\hat{\Omega}^{-1}\,(\hat{K}_{[1]} \quad \hat{K}_{[2]} \quad \ln \hat{P}_0)}{(np \quad n(n-1)p^2 \quad q^n - 1)\,\hat{\Omega}^{-1}\,(np \quad n(n-1)p^2 \quad q^n - 1)} \qquad (39)$$

and

$$\lambda = \frac{(1 \quad 2(n-1)p \quad -q^{n-1})\,\hat{\Omega}^{-1}\,(\hat{K}_{[1]} \quad \hat{K}_{[2]} \quad \ln \hat{P}_0)}{(1 \quad 2(n-1)p \quad -q^{n-1})\,\hat{\Omega}^{-1}\,(np \quad n(n-1)p^2 \quad q^n - 1)} \qquad (40)$$

The covariance matrix, Ω, of t is the symmetric matrix

$$\Omega = \frac{1}{N}\begin{pmatrix} \mu_2' - \mu_1'^2 & \mu_3' - \mu_1'\mu_2' & -\mu_1' \\ \mu_3' - \mu_1'\mu_2' & \mu_4' - \mu_2'^2 & -\mu_2' \\ -\mu_1' & -\mu_2' & \frac{1}{P_0} - 1 \end{pmatrix}$$

Now Ω^{-1} can be written in the form $K\Lambda$, where $K = N|N\Omega|$ and Λ is the symmetric matrix

$$\Lambda = \begin{pmatrix} \dfrac{\mu_4' - \mu_2'^2}{P_0} - \mu_4' & \dfrac{\mu_1'\mu_2' - \mu_3'}{P_0} + \mu_3' & \mu_1'\mu_4' - \mu_2'\mu_3' \\ \dfrac{\mu_1'\mu_2' - \mu_3'}{P_0} + \mu_3' & \dfrac{\mu_2' - \mu_1'^2}{P_0} - \mu_2' & \mu_2'^2 - \mu_1'\mu_3' \\ \mu_1'\mu_4' - \mu_2'\mu_3' & \mu_2'^2 - \mu_1'\mu_3' & \begin{array}{c} \mu_4'(\mu_2' - \mu_1'^2) - \mu_2'^3 \\ + 2\mu_3'^2\mu_2' - \mu_3'^2 \end{array} \end{pmatrix}$$

Since (39) and (40) are homogeneous of degree zero in Ω, Ω^{-1} may be replaced by Λ with the constant K omitted. Take a random sample of size N from the population. Compute the sample moments $\hat{\mu}'_1, \hat{\mu}'_2, \hat{\mu}'_3$ and $\hat{\mu}'_4$ and $\hat{P}_0$ from the data using

$$\hat{\mu}'_i = \sum_{x=0}^{\infty} x^i n_x \Big/ N \qquad i = 1, 2, 3, 4$$

and

$$\hat{P}_0 = n_0 / N$$

where n_x is the number of x values observed. Then the matrix Λ and $\ln \hat{P}_0$ may be determined. The sample factorial moments may be obtained from the equations

$$\hat{\kappa}_{[1]} = \hat{\mu}'_1$$

and

$$\hat{\kappa}_{[2]} = \hat{\mu}'_2 - \hat{\mu}'^2_1 - \hat{\mu}'_1$$

Equations (39) and (40) enable one to tabulate λ as a function of p. However, Gurland and Katti [35] have not formed such a table. As they stand, the two equations may be solved by an iterative trial and error process with an initial value of p taken to be a simple number near the method of moments estimate of p.

Gurland and Katti [35] illustrate the procedure with an example from McGuire et al [49]. They fit the Poisson v Binomial distribution to the data, estimating the parameters,

λ and p, by the 1) method of moments $(\chi^2_4 = 17.47)$, 2) the method of the first moment and the zero frequency $(\chi^2_4 = 5.57)$, 3) the maximum likelihood method $(\chi^2_4 = 5.71)$ and 4) the minimum chi-square method $(\chi^2_4 = 6.71)$.

The efficiency of the minimum chi-square method is tabulated for all combinations of $n = 2, 3, 5$, $\lambda = 0.1$, 0.3, 0.5, 1.0, and 2.0 and $p = 0.1$, 0.3 and 0.5. A general high efficiency prevails throughout the region $2 \leq n \leq 15$, $0 < \lambda \leq 2$, and $0 < p \leq 0.5$. The efficiency of the method of the sample zero frequency is tabulated for the same λ, n and p, and generally high efficiencies prevail, each value being only slightly below that of the minimum chi-square method. The efficiency of the method of moments is tabulated and generally, the values are very low. From the table one would conclude that only for $p << 0.1$ is the efficiency satisfactory, that is, around 0.90.

Gurland and Katti [35] conclude, mainly on the basis of the high efficiency results, that the minimum chi-square method, using the first two factorial cumulants and the logarithm of the zero frequency, may be used instead of the asymptotically efficient maximum likelihood method, when the parameters are in the region for which the tables are constructed. Beyond this region the Poisson $\vee$ Binomial may be replaced by other simpler distributions, such as the Poisson distribution in one limiting case.

Chapter Ten

The Poisson ∨ Pascal Distribution

10.1 Introduction

The Poisson ∨ Pascal Distribution was introduced in section 3.6 as a generalized Poisson distribution and a few of its properties were developed. Under certain conditions it has a limiting Poisson distribution. In this chapter we are concerned with fitting the distribution to observed data. The two major problems in fitting distributions to observed data are 1) computation of the expected frequencies and 2) estimation of the unknown parameters. Three simple "ad hoc" methods of estimation, as well as the maximum likelihood method, are investigated.

10.2 Three Simple Methods of Estimation

The three "ad hoc" methods considered in this section permit quick evaluation of the estimates, however, these estimates are all less efficient than the asymptotically efficient maximum likelihood estimates.

Before we consider the methods of estimation we shall establish some needed relations between various moments and cumulants. Let $u(t)$ and $v(t)$ be the factorial moment generating function and the factorial cumulant generating function, respectively, of any random variable. More explicitly, let $\mu_{[r]}$ and $k_{[r]}$ be the r^{th} factorial moment and the r^{th} factorial cumulant, respectively, so that

$$u(t) = \sum_{r=0}^{\infty} \mu_{[r]} t^r / r!$$

and

$$v(t) = \sum_{r=0}^{\infty} \kappa_{[r]} t^r / r!$$

Now by definition $u(t)$ and $v(t)$ are related as follows

$$v(t) = \ln u(t) \tag{1}$$

By differentiating both sides of (1) we can establish the following

$$\kappa_{[1]} = \mu_{[1]} \qquad \kappa_{[2]} = \mu_{[2]} - \mu_{[1]}^2$$
$$\kappa_{[3]} = \mu_{[3]} - 3\,\mu_{[1]}\,\mu_{[2]} + 2\,\mu_{[1]}^3 \tag{2}$$

The first "ad hoc" method to be considered is the method of the first three factorial cumulants. From section 3.6 we may deduce that the factorial moment generating function, $u(t)$, of the Poisson $\vee$ Pascal distribution is

$$u(t) = e^{\lambda \left[\{q - p(1+t)\}^{-k} - 1 \right]} \tag{3}$$

where $\lambda > 0$, $p > 0$, $k > 0$ and $q - p = 1$. Then, the factorial cumulant generating function, $v(t)$, is

$$v(t) = \lambda \left[\{q - p(1+t)\}^{-k} - 1 \right] \tag{4}$$

From (4) we may obtain the first three factorial cumulants

$$\kappa_{[1]} = \lambda k p \qquad \kappa_{[2]} = \lambda k (k+1)\, p^2$$
$$\kappa_{[3]} = \lambda k (k+1)(k+2)\, p^3 \tag{5}$$

and

Now let $\hat{\mu}_{[1]}$, $\hat{\mu}_{[2]}$ and $\hat{\mu}_{[3]}$ be consistent estimates of the first three factorial moments $\mu_{[1]}$, $\mu_{[2]}$ and $\mu_{[3]}$, respectively, and let $\hat{\kappa}_{[1]}$, $\hat{\kappa}_{[2]}$ and $\hat{\kappa}_{[3]}$ be consistent estimates of $\kappa_{[1]}$, $\kappa_{[2]}$ and $\kappa_{[3]}$, respectively. Solving the equations (5) simultaneously for λ, k and p, suggests that the estimates of λ, k and p, say $\check{\lambda}$, $\check{k}$ and $\check{p}$, respectively, be taken as

$$\check{k} = \frac{2\hat{\kappa}_{[2]}^2 - \hat{\kappa}_{[1]}\hat{\kappa}_{[3]}}{\hat{\kappa}_{[1]}\hat{\kappa}_{[3]} - \hat{\kappa}_{[2]}^2} \qquad \check{p} = \frac{\hat{\kappa}_{[2]}}{(\check{k}+1)\hat{\kappa}_{[1]}} \tag{6}$$

and
$$\check{\lambda} = \frac{\hat{\kappa}_{[1]}}{\check{k}\check{p}}$$

To obtain $\check{k}$, $\check{p}$ and $\check{\lambda}$ numerically, compute $\hat{\mu}_{[1]}$, $\hat{\mu}_{[2]}$ and $\hat{\mu}_{[3]}$ from the data, then compute $\hat{\kappa}_{[1]}$, $\hat{\kappa}_{[2]}$ and $\hat{\kappa}_{[3]}$ from equations (2), and substitute into (6). Generally, this method is favoured among the "ad hoc" methods when the sample mean and variance are large.

The second "ad hoc" method is the method of the first two factorial moments and the propertion of zeroes. Let P_0 be the propertion of zeroes, or the probability of obtaining the value zero. We shall obtain three simultaneous equations involving $\ln P_0$, $\kappa_{[1]}$ and $\kappa_{[2]}$. By replacing $\ln P_0$, $\kappa_{[1]}$ and $\kappa_{[2]}$ with their consistent estimates $\ln \hat{P}_0$, $\hat{\kappa}_{[1]}$ and $\hat{\kappa}_{[2]}$, respectively, and solving for λ, k and p, we obtain the estimates, $\check{\lambda}$, $\check{k}$ and $\check{p}$, respectively. That is

$$\kappa_{[1]} = \lambda k p \qquad \kappa_{[2]} = \lambda k p^2 (k+1) \tag{7}$$

and
$$\ln P_0 = \lambda(q^{-k} - 1)$$

The last of these equations is obtained from (22) in section 3.6, and the first two, from (5). The estimates $\check{p}$, $\check{k}$ and are then obtained from

$$\check{p}\,\ln\left[1+\left(\frac{\hat{\kappa}_{[2]}}{\hat{\kappa}_{[1]}}-\check{p}\right)\frac{\ln\hat{P}_0}{\hat{\kappa}_{[1]}}\right]+\left(\frac{\hat{\kappa}_{[2]}}{\hat{\kappa}_{[1]}}-\check{p}\right)\ln(1+\check{p})=0 \qquad (8)$$

$$\check{k}=\frac{\hat{\kappa}_{[2]}}{\check{p}\;\hat{\kappa}_{[1]}}-1 \qquad \check{\lambda}=\frac{\hat{\kappa}_{[1]}}{\check{k}\;\check{p}} \qquad (9),(10)$$

From the data we obtain $\hat{\mu}_{[1]}$, $\hat{\mu}_{[2]}$ and $\hat{P}_0$, and then, $\hat{\kappa}_{[1]}$, $\hat{\kappa}_{[2]}$ are obtained from (2). The estimates $\check{p}$, $\check{k}$ and $\check{\lambda}$ can then be obtained from (8), (9) and (10). Equation (8) can be solved by an iterative trial and error process. The method of the first two factorial moments and the proportion of zeroes is generally favoured among the "ad hoc" methods when the sample mean and variance are moderate and the proportion of zeroes observed is large.

The third "ad hoc" method is the method of the first two factorial moments and the ratio of the first two frequencies. From (22) in section 3.6 we have that

$$P_0 = e^{\lambda(q^{-k}-1)}$$

and

$$P_1 = \frac{\lambda k p}{q^{k+1}}\,e^{\lambda(q^{-k}-1)}$$

so that

$$\frac{P_0}{P_1} = \frac{q^{k+1}}{\lambda k p} \qquad (11)$$

Solving (11) and the first two equations of (5), simultaneously for λ, k and p, and replacing $\kappa_{[1]}$, $\kappa_{[2]}$ and P_0/P_1 by their consistent estimates $\hat{\kappa}_{[1]}$, $\hat{\kappa}_{[2]}$ and $\hat{P}_0/\hat{P}_1$, respectively, we obtain the estimates $\check{p}$, $\check{k}$ and $\check{\lambda}$, of p, k and λ, respectively, as follows

$$\frac{\ln(1+\hat{p})}{\hat{p}} = \frac{\hat{\kappa}_{[1]}}{\hat{\kappa}_{[2]}} \ln\left[\hat{\kappa}_{[1]} \frac{\hat{P}_0}{\hat{P}_1}\right] \qquad (12)$$

and $$\hat{k} = \frac{\hat{\kappa}_{[2]}}{\hat{\kappa}_{[1]}\hat{p}} - 1 \quad , \quad \hat{\lambda} = \frac{\hat{\kappa}_{[1]}}{\hat{k}\hat{p}} \qquad (13),(14)$$

$\hat{P}_0/\hat{P}_1$, $\hat{\mu}_{[1]}$ and $\hat{\mu}_{[2]}$ are computed from the data, then, $\hat{\kappa}_{[1]}$ and $\hat{\kappa}_{[2]}$ are computed from (2). Equation (12) can be solved for $\hat{p}$ by an iterative trial and error procedure, and then $\hat{k}$ and $\hat{\lambda}$ may be obtained from (13) and (14), respectively. This "ad hoc" method is favoured when the first two frequencies are relatively large.

Katti and Gurland [36] have computed tables of the efficiencies of the three "ad hoc" methods. The region of tabulation in all cases is $0.1 \leq \lambda \leq 5.0$, $.1 \leq p \leq 1.0$ and $0.1 \leq k \leq 2.0$. The method of the first three factorial moments has generally poor efficiency. However, the other two methods have generally, high efficiency values, and one of them may be used (which one depends upon circumstances) without much loss of information. Since the Poisson $\vee$ Pascal distribution rapidly approaches the Neyman Type A distribution as $k \rightarrow \infty$ and $p \rightarrow 0$, only small values of k need be considered. For $\lambda \rightarrow \infty$ the Poisson $\vee$ Pascal distribution approaches either the Pascal distribution or the Poisson distribution, so that only small values of λ need be considered.

Katti and Gurland [36] have fitted the Poisson

Pascal distribution to two sets of data from Beall and Rescia [3] with the parameters estimated for the first set by the method of the first three factorial moments ($\chi^2_8 = 9.58$), and for the second set, by the method of the first two factorial moments and the zero frequency ($\chi^2_8 = 6.88$). For comparison purposes, the Neyman Type A distribution was fitted to both the first set of data ($\chi^2_9 = 42.97$) and the second set of data ($\chi^2_9 = 13.75$). The Poisson $\vee$ Pascal gives a relatively good fit.

10.3 Maximum Likelihood Estimation

The efficiencies of the second two "ad hoc" methods are high, still it is of value to be able to obtain the asymptotically efficient maximum likelihood estimates, especially if a simple enough procedure involving tables and such can be developed. We shall now give the rather long derivation of the maximum likelihood estimates, $\hat{\lambda}$, $\hat{k}$ and $\hat{p}$, of λ, k and p, respectively, of the Poisson $\vee$ Pascal distribution as found in Gurland and Shumway [37]. From section 3.6 we have that the Poisson $\vee$ Pascal probabilities, P_x, are given by

$$P_x = \sum_{r=0}^{\infty} \binom{kr+x-1}{x} \frac{p^x}{q^x} q^{-kr} e^{-\lambda} \frac{\lambda^r}{r!} \tag{15}$$

where $\lambda > 0$, $k > 0$, $p > 0$ and $q - p = 1$. By differentiating (15) with respect to λ and p, respectively, we have

$$\frac{\partial P_x}{\partial \lambda} = \frac{(x+1)q}{\lambda k p} P_{x+1} - \frac{x}{\lambda k} P_x - P_x$$

and
$$\frac{\partial P_x}{\partial p} = \frac{x P_x}{p} - \frac{(x+1)}{p} P_{x+1}$$

These results lead to the maximum likelihood equations

$$L(p) = \sum_{x=0}^{\infty} \frac{n_x}{P_x} \frac{\partial P_x}{\partial p} = \sum_{x=0}^{\infty} n_x M_x - N = 0 \qquad (16)$$

and
$$\sum_{x=0}^{\infty} \frac{n_x}{P_x} \frac{\partial P_x}{\partial \lambda} = \frac{q}{\lambda k p} N \hat{\mu}_1' - \frac{N \hat{\mu}_1'}{\lambda k} - N = 0 \qquad (17)$$

where n_x is the observed number of x values in a sample of size N,

$$M_x = \frac{(x+1)}{\lambda k p} \frac{P_{x+1}}{P_x}$$

and
$$\hat{\mu}_1' = \sum_{x=0}^{\infty} x\, n_x \Big/ N$$

Equations (16) and (17) reduce to

$$L(p) = \sum_{x=0}^{\infty} n_x M_x - N = 0 \qquad (18)$$

and
$$\hat{\mu}_1' = \hat{\lambda} \hat{k} \hat{p} \qquad (19)$$

Now, we must obtain $\frac{\partial P_x}{\partial k}$, so as to obtain the third maximum likelihood equation. Let $g(z)$ and $h(z)$ be the probability generating functions (p.g.f.'s) of the Poisson $\vee$ Pascal and Pascal distributions, respectively. Then

$$g(z) = e^{\lambda[(q-pz)^{-k} - 1]} = e^{\lambda[h(z)-1]} \qquad (20)$$

and
$$\frac{\partial g(z)}{\partial k} = -\lambda\, g(z)\, h(z) \ln(q-pz)$$

Using Leibnitz's formula we have

$$\frac{d^x}{dz^x} \frac{\partial g(z)}{\partial k} = \lambda \left[\sum_{\tau=0}^{x-1} \binom{x}{\tau} \sum_{j=0}^{\tau} g^{(j)}(z) h^{(\tau-j)}(z)\, (x-\tau-1)! \left(\frac{p}{q-pz}\right)^{x-\tau} - \sum_{\tau=0}^{x} \binom{x}{\tau} g^{(\tau)}(z) h^{(x-\tau)}(z) \ln(q-pz) \right] \qquad (21)$$

Now,
$$P_x = \frac{1}{x!}\, g^{(x)}(z)\Big|_{z=0} \tag{22}$$

so that
$$\frac{\partial P_x}{\partial k} = \frac{\partial}{\partial k}\frac{1}{x!}\, g^{(x)}(z)\Big|_{z=0}$$
$$= \frac{1}{x!}\,\frac{d^x}{dz^x}\,\frac{\partial}{\partial k}\, g(z)\Big|_{z=0} \tag{23}$$

Also, if π_r is the r^{th} Pascal probability,
$$\pi_r = \frac{1}{r!}\,\frac{d^r}{dz^r}\, h(z)\Big|_{z=0} \tag{24}$$

Thus, substituting (21), (22) and (24) into (23), we obtain
$$\frac{\partial P_x}{\partial k} = \lambda\left[\sum_{r=0}^{x-1}\frac{1}{(x-r)}\sum_{j=0}^{r} P_j\,\pi_{r-j}\,\frac{p}{q}^{x-r} - \sum_{r=0}^{x} P_r\,\pi_{x-r}\ln q\right] \tag{25}$$

Using equations (4) and (23) from sections 3.2 and 3.6 respectively we have
$$\frac{\partial P_x}{\partial k} = \frac{1}{k}\sum_{r=0}^{x-1}\frac{1}{(x-r)}\,\frac{p^{x-r+1}}{q^{x-r+1}}\left[(r+1)P_{r+1} - \frac{pr}{q}P_r\right] - \frac{q}{kp}\left[(x+1)P_{x+1} - \frac{px}{q}P_x\right]\ln q$$
$$= -\frac{1}{k}\sum_{r=0}^{x-1}\frac{p^{x-r}}{q^{x-r}}\,\frac{r\,P_r}{(x-r)(x-r+1)} + \frac{x}{k}P_x - \frac{q}{kp}\left[(x+1)P_{x+1} - \frac{x}{k}P_x\right]\ln q$$

The third maximum likelihood equation is then
$$\sum_{x=0}^{\infty}\frac{n_x}{P_x}\,\frac{\partial P_x}{\partial k} = N\lambda k p\left(1 - \frac{\ln q}{p}\right) - \sum_{x=0}^{\infty}\frac{n_x}{P_x}\sum_{r=1}^{x-1} B_r^x P_r = 0 \tag{26}$$

where
$$B_r^x = \frac{p^{x-r}}{q^{x-r}}\,\frac{r}{(x-r)(x-r+1)} \tag{27}$$

The B_r^x are tabulated by Gurland and Shumway [37] for $r = 1\,(1)\,8$ and $x = 2\,(1)\,9$. Equation (26) may be written in matrix notation

$$\hat{k} = \frac{1}{N\hat{\lambda}(\hat{p} - \ln \hat{q})} \begin{pmatrix} \frac{n_2}{P_2} & \frac{n_3}{P_3} & \cdots \end{pmatrix} \begin{pmatrix} B_1^1 & 0 & \cdots & 0 \\ B_1^2 & B_2^2 & & \vdots \\ \vdots & & \ddots & \\ B_1^{m-1} & \cdots & & B_m^{m-1} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{m-1} \end{pmatrix} \tag{28}$$

Now, let k be fixed, and differentiate $L(p)$, given by (18), with respect to p, treating λ as a function of p (this is valid because of (19)). Then

$$P_x' = \frac{dP_x}{dp} = -(x+1)P_{x+1}\left(\frac{q+kp}{kp^2}\right) + xP_x\left(\frac{k+1}{kp}\right) + \frac{\lambda P_x}{p}$$

and
$$L^{(1)}(\hat{p}) = \left.\frac{dL(p)}{dp}\right|_{\hat{p}} = \sum_{x=0}^{\infty} n_x M_x \left[\frac{\hat{k}+1}{\hat{k}\hat{p}} - \Delta M_x \frac{(\hat{q}+\hat{k}\hat{p})}{\hat{p}}\right] \tag{29}$$

We are now equipped to obtain the maximum likelihood estimates, $\hat{\lambda}$, $\hat{p}$ and $\hat{k}$. First, obtain initial estimates $\hat{\lambda}_1$, $\hat{p}_1$ and $\hat{k}_1$, from the method of the first three factorial moments. Then calculate an improved estimate of $\hat{p}$, say $\hat{p}_2$, from Newton's equation

$$\hat{p}_2 = \hat{p}_1 - L(\hat{p}_1) / L^{(1)}(\hat{p}_1)$$

Calculate an improved estimate, $\hat{k}_2$, of $\hat{k}$, from the matrix equation (28). Then, obtain an improved estimate, $\hat{\lambda}_2$, of $\hat{\lambda}$, from (19), which is

$$\hat{\lambda}_2 = \hat{\mu}_1' / \hat{k}_2 \hat{p}_2$$

This procedure is repeated until no substantial change occurs in the estimates.

Once the parameters λ, k and p have been estimated, the expected frequencies may be computed by computing the probabilities P_x using the recurrence relation (24) in section 3.6. However, this has two drawbacks, already mentioned in connection with the Poisson Binomial distribution. First, the formula is long and each probability depends upon all the preceding probabilities, and secondly, errors made in computing any one probability are carried along by the succeeding probabilities. Gurland and Shumway [37] have developed a matrix procedure which shortens the labour. The probabilities P_x, given by (15), may be rewritten as

$$P_x = e^{\alpha-\lambda} \frac{p^x}{q^x} \frac{1}{x!} \mu'_{[x]} \tag{30}$$

where $\alpha = \lambda q^{-k}$

and
$$\mu'_{[x]} = \sum_{r=0}^{\infty} (kr)^{(x)} e^{-\alpha} \frac{\alpha^r}{r!} \tag{31}$$

where $(kr)^{(x)} = kr(kr+1) \ldots (kr+x-1)$. Now, (31) suggests that we may treat $\mu'_{[x]}$ as the x^{th} factorial moment of kr, where r is a Poisson random variable with parameter α.

Consider equation (31),

$$\begin{aligned}
\mu'_{[x]} &= \sum_{r=0}^{\infty} (kr)^{(x)} e^{-\alpha} \alpha^r / r! \\
&= \sum_{r=0}^{\infty} \sum_{j=1}^{x} S^*(x,j)\,(kr)^j e^{-\alpha} \alpha^r / r! \\
&= \sum_{r=0}^{\infty} \sum_{j=1}^{x} \sum_{i=1}^{j} S^*(x,j)\, k^j S(j,i)\,(t)^{(i)} e^{-\alpha} \alpha^r / r! \\
&= \sum_{j=1}^{x} \sum_{i=1}^{j} S^*(x,j)\, S(j,i)\, k^j \lambda^i
\end{aligned} \tag{32}$$

where $S(j,i)$ are Stirling numbers of the second kind and are discussed in Riordon [69], and the $S^*(x,j)$ are defined by the expansion

$$X^{(x)} = X(X+1)\cdots(X+x-1)$$
$$= \sum_{j=1}^{x} S^*(x,j) X^j$$

The $S^*(x,j)$ can be shown, easily enough, to satisfy the recurrence relation

$$S^*(x+1,j) = S^*(x,j-1) + x\, S^*(x,j)$$

and so, a table may be constructed as in the case of the Stirling numbers of the first and second kinds. Now equations (30) and (32) may be written in matrix notation as

$$P = B S^* K \underline{S} \Lambda = B A^* \Lambda \tag{33}$$

where

$$\Lambda = \begin{pmatrix} \alpha \\ \alpha^2 \\ \vdots \\ \alpha^m \end{pmatrix} \qquad K = \begin{pmatrix} k & 0 & \cdots & 0 \\ 0 & k^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & k^m \end{pmatrix}$$

$$P = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{pmatrix} \qquad S^* = \begin{pmatrix} S^*(1,1) & 0 & \cdots & 0 \\ S^*(2,1) & S^*(2,2) & & \vdots \\ \vdots & & \ddots & \\ S^*(m,1) & \cdots & & S^*(m,m) \end{pmatrix}$$

$$\underline{S} = \begin{pmatrix} S(1,1) & 0 & \cdots & 0 \\ S(2,1) & S(2,2) & & \vdots \\ \vdots & & \ddots & \\ S(m,1) & \cdots & & S(m,m) \end{pmatrix}$$

and

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & B_m \end{pmatrix}$$

where $B_x = P_0 \dfrac{p^x}{q^x} \dfrac{1}{x!}$ and $P_0 = e^{\alpha - \lambda}$

Now $A^* = S^* K \underline{S}$ has the triangular form

$$A^{*} = \begin{pmatrix} \overset{*}{A}{}_1^{1} & 0 & \cdots & 0 \\ \overset{*}{A}{}_1^{2} & \overset{*}{A}{}_2^{2} & & \vdots \\ \vdots & & \ddots & \\ \overset{*}{A}{}_1^{m} & & \cdots & \overset{*}{A}{}_m^{m} \end{pmatrix}$$

and is tabulated by Gurland and Shumway [37]. Thus the probabilities P_x may be obtained by determining the matrix P form (33). The matrix procedure offers a great improvement in speed and accuracy over the use of the recurrence relation (24) in section 3.6.

Chapter Eleven

Confidence Intervals

11.1 Introduction

We now consdier the problem of obtaining confidence intervals for the parameter λ of the complete Poisson distribution. Since only a single parameter is involved, Fisher's concept of fiducial limits and Neyman's concept of a confidence interval (although they differ basicly) may each be considered in terms of the other. Then, the terminology will be a convenient blending of that of both "schools". There are basicly two types of confidence intervals, non-randomized and randomized. For the former, it is not possible to make "exact" statements as to the probability that the interval contains the parameter. For example, we may only say that the probability is at least 0.95 that the confidence region contains the parameter. Randomized confidence intervals involve the performing of an auxiliary experiment using tables of random numbers and permit one to make "exact" probability statements, such as, the probability is 0.95 that the confidence region contains the parameter. In this chapter we discuss both types of confidence intervals. A brief section deals with the Poisson process and the special approach that may be used. We conclude the chapter with a short discussion of confidence intervals based on approximations.

11.2 Non-randomized Confidence Intervals

Suppose that x is a single observation of a discrete random variable having parameter π, and that, on this basis, we desire a "confidence interval", denoted by, $\delta(x)$, such that, given $0 < \epsilon < 1$,

$$P\left\{ \pi \in \delta(x) \middle| x \right\} \geq 1-\epsilon \tag{1}$$

We call $1 - \epsilon$ the "confidence coefficient". Such a region, $\delta(x)$, may be constructed, according to Neyman [55], as follows. For each π, determine an "acceptance region", denoted by $A(\pi)$, as a subset of the sample space such that

$$P\left\{ x \in A(\pi) \middle| \pi \right\} \geq 1-\epsilon \tag{2}$$

Then, let $\delta(x)$ consist of those values and only those values π whose corresponding acceptance region $A(\pi)$ contains x. Then the so-constructed $\delta(x)$ satisfies (1). For, let π_0 be the actual value of the parameter and x_0, the observed value of the random variable. Then

$$P\left\{ x_0 \in A(\pi_0) \middle| \pi_0 \right\} \geq 1-\epsilon$$

But, by the definition of $\delta(x_0)$, $\pi_0 \in \delta(x_0)$ if and only if $x_0 \in A(\pi_0)$, so that

$$P\left\{ \pi_0 \in \delta(x_0) \middle| x_0 \right\} = P\left\{ x_0 \in A(\pi_0) \middle| \pi_0 \right\} \geq 1-\epsilon$$

Thus $\delta(x_0)$ satisfies (1).

For the Poisson distribution, the parameter is $\pi = \lambda$ and the sample space consists of the non-negative integers. Given $1 - \epsilon$, for each λ we consider acceptance regions $A(\lambda)$ as consisting of consecutive, non-negative integers, $x_1, x_1+1, \ldots, x_2$, such that

$$P\{x \in A(\lambda)\} = \sum_{x=x_1}^{x_2} e^{-\lambda}\frac{\lambda^x}{x!} \geq 1-\epsilon \qquad (3)$$

Now, x_1 and x_2 in (3) are not unique, so that restrictions upon them shall follow, and for each restriction, a different confidence interval shall result.

The first restriction defines the familiar "central" confidence intervals considered by Garwood [31], Ricker [67] and Pearson and Hartley [59]. They are often referred to as the "usual" or "standard" confidence intervals. Choose x_1 and x_2 such that x_1 is the largest integer so that

$$\sum_{x=0}^{x_1-1} e^{-\lambda}\frac{\lambda^x}{x!} \leq \frac{\epsilon}{2}$$

and x_2 is the smallest integer so that

$$\sum_{x=x_2+1}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} \leq \frac{\epsilon}{2}$$

Garwood [31] constructs tables giving $\delta_1(x) = (\lambda_1(x), \lambda_2(x))$, where $\lambda_1(x)$ and $\lambda_2(x)$ are the lower and upper limits, respectively, to the confidence interval for λ, for $x =$ 9 (1) 50, using the fact that Poisson sums can be expressed as integrals over gamma and chi-square distributions. Thus, for the lower tail, we see from section 2.3 that,

$$\sum_{x=0}^{x_1-1} e^{-\lambda}\frac{\lambda^x}{x!} = \int \frac{e^{-t}t^{x_1-1}}{(x_1-1)!}dt = 1 - P\{T \leq \lambda\}$$

$$= 1 - P\{\chi^2_{2x_1} \leq 2\lambda\}$$

where T is a gamma random variable with parameter x_1 and $\chi^2_{2x_1} = 2T$ is a chi-square random variable with $2x_1$ degrees of freedom. For each given x_1, a value for λ can be found from chi-square tables. A similar treatment is used for the upper tail. Once the acceptance regions are determined, the confidence regions are easily obtained. Ricker [67] gives tables of confidence limits for $1 - \epsilon = 0.95$ and 0.99 and $x = 0\ (1)\ 50$. The Poisson sums are obtained from tables given in Soper [72] and Whitaker [84]. For $x > 50$, Ricker [67] suggests using the normal approximation to the Poisson distribution. Crow and Gardner [19] have determined an accurate approximation to $\delta_1(x)$ by averaging the two large sample approximations, the normal approximation with continuity correction and the square root normal approximation with continuity correction. Both these approximations were established in section 2.8. If λ_u and λ_L are the upper and lower confidence limits, respectively, then

$$\left.\begin{matrix}\lambda_u \\ \lambda_L\end{matrix}\right\} \simeq x + \frac{3}{8}\beta_\epsilon^2 \pm \frac{1}{2} \pm \beta_\epsilon \sqrt{(x \pm 1/2) + \frac{\beta_\epsilon^2}{8}}$$

where β_ϵ is the upper $100\epsilon/2\ \%$ point of the standard normal distribution. Central confidence intervals are, of course, non-randomized, and are larger than necessary to have the parameter covered with probability at least $1 - \epsilon$. The two tails each have probability less than $\epsilon/2$ of containing the parameter.

A second restriction upon x_1 and x_2 is due to Sterne [74]. Let the acceptance region, denoted by $A_2(\lambda)$, consist of the values x having the largest probabilities. Thus, for each λ, the most probable value of x is a member of $A_2(\lambda)$, the second most probable value is also a member, and so on, until (3) is satisfied. The acceptance region $A_2(\lambda)$ is determined according to Crow and Gardner [19], by proceeding continuously from small values to larger values of λ, starting at $\lambda = 0$, and altering x_1 and x_2 so as to maintain (3). Suppose that for $\lambda = \lambda_o$, the acceptance region $A_2(\lambda_o)$, is $x_1, x_1+1, \ldots, x_1+r = x_2$. Then

$$\sum_{x=x_1}^{x_1+r} e^{-\lambda_o} \frac{\lambda_o^x}{x!} \geq 1-\epsilon$$

Now, because of the definition of Sterne's acceptance region, no probability sum of less than r+1 terms is as large as $1 - \epsilon$ for $\lambda = \lambda_o$. Because of the inequality (3) in section 2.3, no probability sum of less than r+1 terms is as large as $1 - \epsilon$ for $\lambda > \lambda_o$. Thus, combining these two statements, we have that the length of the acceptance cannot decrease with increasing λ. It is then desirable to keep the acceptance region at the same length, if possible, as λ increases. Let λ_{x_1+1, x_2+1} be the value of λ that maximizes

$$\sum_{x=x_1+1}^{x_2+1} e^{-\lambda} \frac{\lambda^x}{x!}$$

Then, the same length is maintained, if possible, by substituting x_2+1 for x_1 at $\lambda = \lambda_{x_1+1, x_2+1}$. If $\sum_{x=x_1}^{x_2} e^{-\lambda} \frac{\lambda^x}{x!}$ drops to $1 - \epsilon$ before such a substitution is possible, the acceptance region must be enlarged to $x_1, x_1+1, \ldots, x_2+1$. Once the acceptance regions $A_2(\lambda)$ have been determined, the confidence regions, denoted by $\delta_2(x)$, are easily obtained.

A third type of confidence interval is due to a slight modification of Sterne's acceptance region by Crow and Gardner [19]. If in the process of determining Sterne's acceptance region, we replace x_1 by x_2+1 at the value of λ where

$$\sum_{x=x_1+1}^{x_2+1} e^{-\lambda} \frac{\lambda^x}{x!}$$

first equals $1-\epsilon$, instead of at $\lambda = \lambda_{x_1+1, x_2+1}$ we obtain a different accpetance region, which we denote by $A_3(\lambda)$. If $\sum_{x=x_1}^{x_2} e^{-\lambda} \frac{\lambda^x}{x!}$ drops to $1-\epsilon$ before the substitution is possible, the acceptance region is enlarged to $x_1, x_1+1, \ldots, x_2+1$. The confidence regions obtained by this modification are denoted by $\delta_3(x)$ and have been tabulated by Crow and Gardner [19] for $1-\epsilon = 0.80, 0.90, 0.95, 0.99$ and 0.99 and $x = 0\ (1)\ 300$.

Crow and Gardner [19] compare the confidence intervals $\delta_1(x)$, $\delta_2(x)$ and $\delta_3(x)$ by considering their lengths d_1, d_2 and d_3, respectively, and computing the relative percentage deviations $100(d_1 - d_3)/d_1$ and $100(d_1 - d_2)/d_1$. The improvement of $\delta_3(x)$ over $\delta_1(x)$ is appreciable for small values of x and decreases as x increases. The relative improvement of $\delta_3(x)$ over $\delta_2(x)$ is also appreciable for small x. Of all non-randomized confidence intervals for λ, $\delta_2(x)$ and $\delta_3(x)$ have the shortest total lengths for a given confidence coefficient. In addition, $\delta_3(x)$ has the advantage over $\delta_2(x)$, of having the smallest upper confidence limits. However, all three of $\delta_1(x)$, $\delta_2(x)$ and $\delta_3(x)$ are at a disadvantage when one-sided confidence intervals are desired.

11.3 Randomized Confidence Intervals

A shortcoming of the confidence intervals, $\delta_2(x)$ and $\delta_3(x)$, is that there is no statement possible about the probabilities, first that the parameter lies below the lower confidence limit, and second, that the parameter lies above the upper confidence limit. When the statistician is interested in only one limit (such as the opper limit of the average number of defects on a manufactured item), this additional information is desired. Stevens [75] introduces randomized confidence intervals as an interesting solution to the problem.

Let X be a Poisson random variable having p.d.f. given by

$$p(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x = 0,1,2,\cdots \tag{4}$$

where $\lambda > 0$. Define $G(x,\lambda)$ as

$$G(x;\lambda) = \sum_{\tau=x}^{\infty} p(\tau;\lambda) = \sum_{\tau=x}^{\infty} e^{-\lambda}\frac{\lambda^\tau}{\tau!} \tag{5}$$

Define the lower P_0 limit of λ, denoted by $\lambda_0(x)$, as the solution for λ of

$$G(x,\lambda) = P_0 \tag{6}$$

where P_0 is a given real number such that $0 \leq P_0 \leq 1$ and

$\lambda_0(0)$ is defined to be identically zero. Similarly, define the upper P_1 limit of λ, denoted by $\lambda_1(x)$, as the solution for λ of

$$G(x+1, \lambda) = 1 - P_1 \tag{7}$$

From (5) we have that

$$h_0(\lambda) = \frac{\partial G(x,\lambda)}{\partial \lambda} = \sum_{r=x}^{\infty} \left[e^{-\lambda} \frac{\lambda^{r-1}}{(r-1)!} - e^{-\lambda} \frac{\lambda^{r}}{r!} \right] = e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \tag{8}$$

Similarly, from (5) we have

$$h_1(\lambda) = \frac{\partial G(x+1, \lambda)}{\partial \lambda} = e^{-\lambda} \frac{\lambda^{x}}{x!} \tag{9}$$

From (6), (7), (8) and (9) we have

$$\int_0^{\lambda_0(x)} h_0(\lambda)\, d\lambda = P_0$$

and

$$\int_0^{\lambda_1(x)} h_1(\lambda)\, d\lambda = 1 - P_1$$

Then, $h_0(\lambda)$ and $h_1(\lambda)$ supply limits to the parameter λ, corresponding to any pair of significance levels, P_0 and P_1. Now, suppose that x is the observed value of the Poisson random variable, and let

$$y = x + z$$

where z is any number belonging to the rectangular distribution and chosen from a table of random numbers. We note that both x and z are uniquely determined if y is given. Let $y_0 = x_0 + z_0$ where x_0 and z_0, and so y_0, are fixed. Then from simple probability concepts we have

$$\begin{aligned} P\{y \geq y_0\} &= P\{x > x_0\} + P\{x = x_0\} P\{z \geq z_0\} \\ &= G(x_0+1, \lambda) + p(x_0; \lambda)(1 - z_0) \\ &= z_0 G(x_0+1, \lambda) + (1 - z_0) G(x_0, \lambda) \end{aligned} \tag{10}$$

Differentiate both sides with respect to λ, and drop the zero suffixes on x_o and z_o, so that

$$h_z(\lambda) = z\,h_1(\lambda) + (1-z)\,h_0(z)$$
$$= z\,e^{-\lambda}\frac{\lambda^x}{x!} + (1-z)\,e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} \qquad (11)$$

For any given x, there exists an $h_z(\lambda)$ for each possible value of z. Thus to make $h_z(\lambda)$ unique the following rule is employed. Perform the main experiment and obtain a value x of the Poisson random variable. Then, one, and only one selection is made, from a table of random numbers, of the value z. The unique distribution, $h_z(\lambda)$, thus determined, is called "the fiducial distribution of λ". The fiducial limits, obtained by integrating $h_z(\lambda)$, always lie wholly within the limits obtained for λ by the Ricker [67] or Garwood [31] approach. Two useful properties are displayed by the resulting confidence limits : 1) Because of the continuous nature of the random variable y, probability statements may be made in terms of equalities rather than inequalities. 2) The probabilities, first, of the parameter lying below the lower fiducial limit, and second, of the parameter lying above the upper fiducial limit, are easily obtained.

Suppose that λ_o and λ_o' are the lower limits corresponding to the observed value x, and $x+1$, respectively, obtained from Garwood [31], Ricker [67] or Pearson and Hartley [59]. We shall calculate the lower fiducial limit of Stevens' method by interpolating between the two consecutive values of y, that is, $y = x$ and $y = x+1$. But first, we must obtain the functional relationship between y and λ. For a lower limit, corresponding to any given, fixed probability P, we have

$$P = z\, G(x+1, \lambda) + (1-z)\, G(x, \lambda) \tag{12}$$

so that
$$\frac{dP}{dx} = \frac{\partial P}{\partial z} + \frac{\partial P}{\partial \lambda}\frac{d\lambda}{dz}$$

$$= -p(x;\lambda) + h_z(\lambda)\frac{d\lambda}{dz} = 0 \tag{13}$$

Rearranging (13) we obtain

$$\frac{d\lambda}{dz} = \frac{p(x;\lambda)}{h_z(\lambda)} = \frac{\lambda}{\lambda z + x(1-z)} \tag{14}$$

Since the equation (12) approaches $G(x+1, \lambda) = P$, the equation appropriate to the value $y = x+1$, as z approaches one, λ is a continuous function of z, although the derivative $\frac{d\lambda}{dy}$ is discontinuous at integral values of y.

Now
$$\frac{d^2\lambda}{dz^2} = \frac{\partial}{\partial z}\left(\frac{d\lambda}{dz}\right) + \frac{\partial}{\partial \lambda}\left(\frac{d\lambda}{dz}\right)\frac{d\lambda}{dz} \tag{15}$$

From (14) and (15) we obtain the following results,

$$\left.\frac{d\lambda}{dz}\right|_{z=0} = \frac{\lambda_0}{x}, \qquad \left.\frac{d^2\lambda}{dz^2}\right|_{z=0} = \frac{\lambda_0^2}{x^2}\left(\frac{x+1}{\lambda_0} - 1\right) \tag{16),(17}$$

and
$$\left.\frac{d\lambda}{dz}\right|_{z=1} = 1, \qquad \left.\frac{d^2\lambda}{dz^2}\right|_{z=1} = \frac{x}{\lambda_0} - 1$$

Let λ_{05} be Stevens' lower fiducial limit. Using Taylor's series and either expanding about the point $z = 0$, or expanding about $z = 1$, we obtain

$$\lambda_{05} \simeq \lambda_0 + \left.\frac{d\lambda}{dz}\right|_{z=0} z + \frac{1}{2}\left.\frac{d^2\lambda}{dz^2}\right|_{z=0} z^2 \tag{18}$$

or $$\lambda_{os} \simeq \lambda_o' + \left.\frac{d\lambda}{dz}\right|_{z=1}(z-1) + \frac{1}{2}\left.\frac{d^2\lambda}{dz^2}\right|_{z=1}(z-1)^2 \qquad (19)$$

In summary, then, to calculate Stevens' fiducial limits, proceed as follows. For the lower fiducial limit, λ_{os}, obtain from suitable tables (Pearson and Hartley [59]) the lower limits corresponding to the observed value of x, and also, $x+1$. Denote these by λ_o and λ_o', respectively. Obtain a value for z by referring to a table of random numbers. If $z \leq 1/2$, compute $\left.\frac{d\lambda}{dz}\right|_{z=0}$ and $\left.\frac{d^2\lambda}{dz^2}\right|_{z=0}$ from (16) and (17). Then compute λ_{os} from (19). If $z > 1/2$, compute λ_{os} from (20). The same basic procedure is used to obtain Stevens' upper fiducial limits. Thus, Stevens' useful, randomized fiducial limits may be obtained from existing tables of confidence intervals by a simple interpolation procedure.

To conclude our treatment of randomized confidence intervals for the Poisson parameter we present the concept of the Neyman - shortest unbiased confidence interval as given in Blyth and Hutchinson [9]. Generally, a random subset A of the possible values of a parameter λ is an "unbiased" confidence interval for λ, with confidence coefficient $1-\epsilon$, if the following two conditions hold.

$$P\{\lambda \in A \mid \lambda\} \geq 1-\epsilon \quad \text{for all } \lambda \qquad (20)$$

$$P\{\lambda' \in A \mid \lambda\} \leq P\{\lambda \in A \mid \lambda\} \quad \text{for all } \lambda' \qquad (21)$$

A "uniformly" Neyman - shortest confidence interval, A, is one which, in addition to satisfying (20) and (21), also minimizes $P(\lambda' \in A \mid \lambda)$ for all pairs λ' and λ. A process of randomization is needed for the construction of such confidence intervals when the random variable involved is discrete.

Let X be the Poisson random variable with parameter λ and let Y be a random variable having the rectangular distribution, such that X and Y are independent. Eudey [23] has shown that a uniformly most powerful unbiased test of

$$H_0: \lambda = \lambda_*$$
$$H_1: \lambda \neq \lambda_*$$

having significance level $1 - \alpha$ and based on a single observation X, is given by the acceptance region

$$A(\lambda_*) = \left\{ (x,y) \,\middle|\, x_0 + \gamma_0 \leq x+y \leq x_1 + \gamma_1 \right\} \tag{22}$$

where x_0 and x_1 are integers and $0 \leq \gamma_0 \leq 1$, $0 \leq \gamma_1 \leq 1$, and $A(\lambda_0)$ satisfies the two conditions

$$P\left\{ x_0 + \gamma_0 \leq x+y \leq x_1+\gamma_1 \,\middle|\, \lambda_* \right\} = \alpha \;, \qquad \frac{d\,P}{d\lambda}\left\{ x_0+\gamma_0 \leq x+y \leq x_1+\gamma_1 \,\middle|\, \lambda \right\} \Bigg|_{\lambda=\lambda^*} \tag{23),(24}$$

More specifically the conditions (23) and (24) are

$$\gamma_0 = \frac{(x_1 - \lambda_*)\left[P\left\{ x_0 \leq x \leq x_1 - 1 \,\middle|\, \lambda_* \right\} - \alpha \right] - x_0 P\left\{ x = x_0 \,\middle|\, \lambda_* \right\} + x_1 P\left\{ x = x_1 \,\middle|\, \lambda_* \right\}}{(x_1 - x_0)\; P\left\{ x = x_0 \,\middle|\, \lambda_* \right\}} \tag{25}$$

$$\gamma_1 = \frac{(x_0 - \lambda_*)\left[P\left\{ x_0 \leq x \leq x_1 - 1 \,\middle|\, \lambda_* \right\} - \alpha \right] - x_0 P\left\{ x = x_0 \,\middle|\, \lambda_* \right\} + x_1 P\left\{ x = x_1 \,\middle|\, \lambda_* \right\}}{(x_1 - x_0)\; P\left\{ x = x_1 \,\middle|\, \lambda_* \right\}} \tag{26}$$

Eudey [23]'s test is unique except for the way in which randomization is carried out. Blyth and Hutchinson [9] have tabulated Neyman - shortest unbiased confidence intervals for the Poisson parameter, λ, for $\alpha = 0.95$ and 0.99 and

$$x+y = 0.01\,(0.01)\,0.10\,(0.02)\,0.20\,(0.05)\,1.00\,(0.1)\,10.0$$
$$10.0\,(0.2)\,40.0\,(0.5)\,55.0\,(1)\,250$$

using an ILLIAC digital computer. For given values of λ_* and α, trial values of x_0 and x_1 were substituted into (25) and (26) until both γ_0 and γ_1 were in the interval $[0, 1]$.

Blyth and Hutchinson [9], in comparing Stevens' fiducial limits with those based on Eudey [23]'s results, show that Stevens' are shorter for $1.6 \leq x + y \leq 9.4$ and longer elsewhere. However, it should be mentioned that shortness is not the only criterion of desirability for confidence intervals.

11.4 The Poisson Process

Often it is possible to interpret observations as resulting from the continuing Poisson process characterized by the parameter λ. In such a case, observations stating the number of events during a certain amount of observation, or, the amount of observation, T, required for a certain number of events, n, to occur, are possible. From experimental designs based on this type of observation, confidence intervals and test of hypotheses, involving the parameter λ, may be determined.

In section 2.6 we have shown that $2\lambda T$ has a chi-square distribution with $2n$ degrees of freedom. Thus, given confidence coefficient $1 - \epsilon$, numbers C and D may be found from chi-square tables so that

$$P\left\{ C \leq 2\lambda T \leq D \right\} = 1 - \epsilon$$

Thus $\left(\frac{C}{2T}, \frac{D}{2T} \right)$ is a confidence interval for λ having confidence coefficient $1 - \epsilon$.

If two Poisson processes, characterized by λ_1 and λ_2, are to be compared by arriving at conclusions concerning the ratio $\gamma = \lambda_2/\lambda_1$, a confidence interval for γ can be easily constructed. Let T_1 and T_2 be the amounts of observation required for n_1 and n_2 events to occur in the λ_1 - Poisson process and the λ_2 - Poisson process, respectively. Then $2\lambda_1 T_1$ and $2\lambda_2 T_2$ have both chi-square distributions with $2n_1$ and $2n_2$ degrees of freedom, respectively. Then the statistic

$$F = \frac{2\lambda_1 T_1}{2n_1} \Bigg/ \frac{2\lambda_2 T_2}{2n_2} = \frac{T_1 n_2}{\gamma T_2 n_1}$$

has the F distribution with $(2n_1, 2n_2)$ degrees of freedom. Given confidence coefficeint, $1 - \epsilon$, two numbers C and D may be found from tables of the F distribution such that

$$P\left\{ C \leq \frac{T_1 n_2}{\gamma T_2 n_1} \leq D \right\} = 1 - \epsilon$$

Thus $\left(\frac{T_2 n_1}{T_1 n_2} D , \frac{T_2 n_1}{T_1 n_2} C \right)$ is a confidence interval for having confidence coefficient $1 - \epsilon$.

Birnbaum [4] shows that it two Poisson processes characterized by λ_1 and λ_2 may be observed simultaneously, then a confidence interval for $\gamma = \lambda_2/\lambda_1$ can be obtained. Let x_1 and x_2 be the number of events occuring in the λ_1 - and λ_2 - Poisson processes, respectively, during a fixed amount of observation. Let p be the probability that, starting at a certain "time", the first event to occur will come from the λ_1 - Poisson process. Clearly, then, $p = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$. Then $1 - p$ is the probability that the first event will come from the λ_2 - Poisson process. We may treat the observations x_1 and x_2 as the number of "successes" and "failures", respectively, of a binomial sample. Using any of the procedures developed for obtaining confidence intervals for the "proportion" parameter of a binomial distribution, we may obtain a confidence interval for p. If (C, D) is the confidence interval for p having confidence coefficient $1 - \epsilon$, then (D^{-1}, C^{-1}) is a confidence interval for γ having confidence coefficient $1 - \epsilon$.

11.5 Approximate Confidence Intervals

In section 2.8 we showed that the statistic $\dfrac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$ where the X_i, $i = 1, 2, \ldots, n$ are independent Poisson distributions each having parameter λ, has an asymptotic normal distribution with mean λ and variance λ/n. This fact may be used to form an approximate confidence interval for λ when n is large. Let the desired confidence coefficient be $1 - \epsilon$. Then, from standard normal tables, a real number β_ϵ can be found such that

$$P\left\{-\beta_\epsilon \leq \frac{\bar{X}-\lambda}{\sqrt{\lambda/n}} \leq +\beta_\epsilon\right\} \simeq 1-\epsilon$$

The resulting confidence interval for λ is the asymptotically shortest $100\,(1 - \epsilon)\,\%$ confidence interval. Let the upper and lower confidence limits for λ be λ_u and λ_L, respectively. Then

$$\lambda^2 - \lambda\left(\frac{\beta_\epsilon^2}{n} + 2\bar{x}\right) + \bar{x}^2 = 0$$

and solving for λ gives

$$\left.\begin{matrix}\lambda_u\\ \lambda_L\end{matrix}\right\} = \bar{x} + \frac{\beta_\epsilon^2}{2n} \pm \frac{\beta_\epsilon}{2}\sqrt{\frac{\beta_\epsilon^2}{n^2} + \frac{4\bar{x}}{n}} \qquad (27)$$

We have also shown in section 2.8 that the statistic $2\sqrt{\bar{X}}$ has an asymptotic normal distribution with mean $2\sqrt{\lambda}$ and variance $1/n$. Thus

$$P\left\{-\beta_\epsilon \leq 2\sqrt{n}\left(\sqrt{\bar{X}}-\sqrt{\lambda}\right) \leq \beta_\epsilon\right\} \simeq 1-\epsilon$$

Solving for the upper and lower confidence limits, λ_u and λ_L, respectively, we have

$$\left.\begin{matrix}\lambda_u\\ \lambda_L\end{matrix}\right\} = \bar{x} + \frac{\beta_\epsilon^2}{4n} \pm \beta_\epsilon\sqrt{\frac{\bar{x}}{n}} \qquad (28)$$

As we have already mentioned in section 11.2, Crow and Gardner [19] have averaged (27) and (28) to obtain a remarkably accurate approximation to the Garwood [31] and Ricker [67] confidence intervals, $\delta_1(x)$. When terms of order $1/n$ are neglected, both (27) and (28) simplify to

$$\left.\begin{matrix}\lambda_u\\ \lambda_L\end{matrix}\right\} \simeq \bar{x} \pm \beta_\epsilon\sqrt{\frac{\bar{x}}{n}}$$

a result commonly seen.

11.6 Concluding Comments

The different methods of obtaining confidence intervals discussed in this chapter are in the main based on a single observation of the Poisson random variable. However, because of the additivity property of the Poisson distribution, established in section 2.4, the sum of n independent Poisson random variables each having parameter λ, is again a Poisson random variable, and has parameter $n\lambda$. A random sample of size n can be taken and a confidence interval obtained for $n\lambda$. Thus a confidence interval for λ is obtained based on n observations of the Poisson random variable.

Most of the distinctions made between the different confidence intervals discussed in this chapter are appreciable only for small samples. In the majority of cases, the normal approximate confidence intervals and the simple Garwood [31] and Ricker [67] confidence intervals are satisfactory.

Chapter Twelve

Hypothesis Testing

12.1 Introduction

In this chapter we deal with the problem of hypothesis testing for the Poisson distribution. Goodness of fit and homogeneity tests are investigated. The Poisson distribution truncated at zero is briefly discussed. In continuation of our treatment of the Poisson process, a chapter is offered, and the simplicity of the methods involved should be noted. The problems of hypothesis testing and combing tests when observations are from a discrete distribution are considered in the final two chapters.

12.2 Goodness of Fit Tests

Suppose that the sample space of a Poisson random variable x is divided into a finite number r of mutually exclusive and exhaustive cells or classes $S_1, S_2, \ldots, S_r$. Let $p_1, \ldots, p_r$ where $\sum_{i=1}^{r} p_i = 1$ be the probabilities that an observation chosen at random falls in $S_1, S_2, \ldots, S_r$, respectively. Take a random sample of size n from the population and let $o_1, o_2, \ldots, o_r$ be the number of observations falling in $S_1, S_2, \ldots, S_r$, respectively, so that $\sum_{i=1}^{r} o_i = n$. We wish to test the null hypothesis H_o that the population has a Poisson distribution with known parameter λ. Then consider the quantity called "the chi-square of the grouped sample"

$$\chi^2 = \sum_{i=1}^{r} \frac{(o_i - n p_i)^2}{n p_i} \tag{1}$$

If the null hypothesis H_o is true, np_i will be the expected number of observations falling in S_i, where $i = 1, 2, \ldots, r$ and the value of χ^2 will be small, so that χ^2 is apparently a convenient measure of the deviation of the hypothetical Poisson distribution from the actual distribution of the population. Now, under the null hypothesis H_o, the random variables $o_1, o_2, \ldots, o_r$ have a multinomial distribution with parameters n and $p_1, p_2, \ldots, p_{n-1}$. Cramér [18] (P. 417 - 419) proves that χ^2 has an asymptotic chi-square distribution with $r-1$ degrees of freedom. As $n \to \infty$, the multinomial distribution approaches the multivariate normal, so that χ^2, which can then be written as the sum of $r-1$ standard normal random variables, has a chi-square distribution. The most significant aspect of this result is that the limiting distribution of χ^2 is independent of the original Poisson distribution, and only depends upon the number of cells r. Large values of χ^2 indicate a poor fit so that the critical region of the test is selected to be the "right-hand" tail of the chi-square distribution with $r-1$ degrees of freedom. When n is fairly large the chief source of discrepancy is that the exact distribution of χ^2 is discontinuous whereas the approximating distribution is continuous. The standard correction for continuity may be applied. Since the random variable χ^2 has a finite range and its continuous approximation, infinite range, there exists a tendency to overestimate the true probabilities near the end of the "right-hand" tail. For finite n, the approximation is satisfactory for most purposes if $np_i \geq 5$ where $i = 1, 2, \ldots, r$. It is desirable, if accuracy is not forsaken, to have the cells each as small as possible so that as little information as possible from the sample is lost. Cramér [18] shows that under the null hypothesis H_0, the mean and variance of χ^2 are

$$E\{\chi^2\} = r-1 \tag{2}$$

and

$$V\{\chi^2\} = 2(r-1) + \frac{1}{n}\left[\sum_{i=1}^{r} \frac{1}{p_i} - r^2 - 2r + 2\right] \tag{3}$$

Thus, if the number of observations n is so small as to rule out the application of the chi-square test some information may be obtained from (2) and (3). The well - known results of this paragraph apply, of course, to any discrete hypothetical distribution whether it is Poisson or not.

Suppose that we now wish to test the null hypothesis H_0 that a population has a Poisson distribution with unknown parameter. The chi-square test may be used to test the "goodness of fit" of the Poisson distribution to a random sample taken from the population. The unknown parameter may be estimated by the manimum likelihood estimate, the sample mean $\bar{x}$. Cramér [18] (P. 424 - 434) shows that

$$\chi^2 = \sum_{i=1}^{r} \frac{(o_i - n\hat{p}_i)^2}{n\hat{p}_i} \tag{4}$$

where $\hat{p}_i$ is p_i with the parameter estimated by $\bar{x}$, has a chi-square distribution with $r-2$ degrees of freedom. Then a right - tail test based on χ^2 may be used to test the goodness of fit.

As Fisher [28] points out, χ^2, as a measure of discrepancy, may be inaccurate with $\bar{x}$ is small and the Poisson series short (that is, small r), since an indication of discrepancy may then come chiefly from frequencies with small expectation. Fisher [28] proposes a "generalized measure of deviation"

$$L = \sum_{i=1}^{r} o_i \ln\left(\frac{o_i}{n \hat{p}_i}\right) \tag{5}$$

which he describes as "the logarithmic difference in likelihood between the most likely Poisson series and the most likely theroetical series without restriction". The statistic $-2L$ has an asymptotic chi-square distribution when the Poisson parameter is large. Cochran [12] also investigates the statistic L and suggests that the likelihood approach is more appropriate than the χ^2 approach as a test criterion.

We shall now present a test for deviation in the zero frequency. Section 12.4, dealing with the Poisson distribution truncated at zero, may be read along with this paragraph. Sometimes it happens when using a goodness of fit test, that the quantity χ^2 is significant because the zero observation is over - represented. Thus, a method of examining whether or not the zero frequency is responsible for the significant χ^2 is of value. Let n_x, where $x = 0, 1, 2, \ldots,$ be the number of x values observed in a random sample of size n taken from a population that is Poisson under the null hypothesis H_o. Let $n' = n - n_o$. Let χ^2_p be the total chi-square quantity, that is

$$\chi^2_p = \sum_{x=0}^{\infty} \frac{(n_x - np(x))^2}{np(x)}$$

where $p(x)$ are the complete Poisson probabilities. Let χ^2_T be the chi-square quantity for the Poisson distribution truncated at zero, that is

$$\chi^2_T = \sum_{x=1}^{\infty} \frac{(n_x - n'p'(x))^2}{n'p'(x)}$$

where $p'(x)$ are the truncated Poisson probabilities. Define χ^2_Z by

$$\chi^2_Z = \chi^2_p - \chi^2_T$$

Then χ^2_Z has a chi-square distribution with one degree of freedom and may be used to detect significant departures in the zero frequency.

Similarly, a maximum likelihood ratio goodness of fit test may yield a significant value of L. Let L_p be the total L, that is

$$L_p = \sum_{x=0}^{\infty} n_x \ln\left(\frac{n_x}{np(x)}\right)$$

and let L_T be the "L value" for the Poisson distribution truncated at zero, that is

$$L_T = \sum_{x=1}^{\infty} n_x \ln\left(\frac{n_x}{n'p(x)}\right)$$

Define L_Z by

$$L_Z = L_p - L_T$$

Then $-2L_Z$ has an approximate chi-square distribution with one degree of freedom, and may be used to test the zero frequency. Both the approaches mentioned are only valid for large samples.

An exact expression may be easily obtained for the probability that a certain number, say n_0, of the observations in the sample are "zero" observations. The problem is equivalent to a classical occupancy problem where the zero observation corresponds to a cell being empty when $T = \sum_{x=0}^{\infty} n_x$ objects are placed in n cells. The probability that n_0 objects or observations are zero observations may be determined from Feller [25] (P. 92) to be

$$P_{n_0} = \binom{n}{n_0} \sum_{j=0}^{n-n_0-1} (-1)^j \frac{(n-n_0-j)^T}{n^T} \binom{n-n_0}{j}$$

$$= \frac{1}{n^T} \binom{n}{n_0} \Delta^{n-n_0} 0^T$$

Thus, to detect both positive and negative departures of n_0 from the expected we set up a critical region

$$C = \left\{ n_0 \,\middle|\, n_0 = 0, 1, 2, \cdots, r_1 - 1 \text{ and } n_0 = r_2 + 1, \cdots, n-1, n \right\}$$

where r_1 is the largest positive integer and r_2, the smallest positive integer such that

$$\sum_{n_0=0}^{r_1-1} P_{n_0} \leq \alpha/2$$

and

$$\sum_{n_0=r_2+1}^{n} P_{n_0} \leq \alpha/2$$

where α is the upper limit to the significance level. Values of $\frac{\Delta^r 0^r}{r!}$ have been tabulated in Stevens [76] and in Fisher and Yates [29] so that the values of P_{n_o} may be obtained with only a little computation. If n is large, the normal approximation may be used (see Weiss [83]). If $T >> n$, P_{n_o} may be approximated by the Poisson distribution (see Feller [25] (P. 94)),

$$P_{m_o} \simeq e^{-\lambda} \frac{\lambda^{m_o}}{m_o!}$$

where $\lambda = n\, e^{-T/m}$. Thus P_{n_o} may be obtained from Poisson Tables.

12.3 Homogeneity Tests

a) The Index of Dispersion Test

Suppose that a random sample consists of n observations $x_1, x_2, \ldots, x_n$ from a Poisson population, and that $T = \sum_{i=1}^{m} x_i$. We have seen in section 4.4 that T is a sufficient statistic for the Poisson parameter. If T and n are fixed, the conditional random variables $x_1, x_2, \ldots, x_n$ may be treated as multinomial random variables each having probability $1/n$, so that the conditional probability of the observations is

$$P\left\{ x_1, x_2, \ldots, x_m \,\middle|\, T, m \right\} = \frac{T!}{x_1!\, x_2! \ldots x_m!} \left(\frac{1}{m} \right)^T \tag{16}$$

We note that $E\left\{ x_i \,\middle|\, T ; n \right\} = T/n = \bar{x}$. We would like to test the null hypothesis H_o that each of the observations $x_1, x_2, \ldots, x_n$ arise from the same Poisson population and not from a compound Poisson population. The chi-square test applied to the conditional observations yields the well-known quantity I called "the Poisson index of dispersion"

$$I = \sum_{i=1}^{m} \frac{\left(x_i - E\{x_i|T\}\right)^2}{E\{x_i|T\}} = \sum_{i=1}^{m} \frac{(x_i-\bar{x})^2}{\bar{x}} \qquad (7)$$

Cramér [18] (P. 445 - 449) discusses the chi-square test as a test for the homogeneity of a number of samples, and in so doing, shows that I is approximately a chi-square random variable with n-1 degrees of freedom. Hoel [41] considers approximations for the first four moments of I and from his work it appears that the chi-square approximation is highly satisfactory for $\bar{x} \geq 5$. For slightly smaller $\bar{x}$ the approximation is still fairly accurate for the Poisson distribution (but not for some other distributions such as the binomial distribution). Rao and Chakravarti [65] have also investigated the chi-square approximation and conclude that it yields "good" results for $\bar{x} > 3$ but that it "may be misleading" if $\bar{x} < 1$. It should be noted that the numerator of I equals n times the sample variance and that under the null hypothesis H_0, the denominator is also an estimate of the variance of a Poisson distribution. Thus, in cases where the sample is non-homogeneous and "over-dispersion" occurs, the quantity I will be large, so that a right tail test based on the chi-square distribution is possible. This test is sometimes called "the variance test of homogeneity". If it so happens that exact probabilities are desired, the expression (6) may be used.

b) The Likelihood Ratio Test

Suppose that we wish to test the null hypothesis H_o that each of the n observations comes from the same Poisson population having parameter λ against the alternative hypothesis H_1 that each of the n observations comes from different Poisson populations and having parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ as in part a) of this section. Thus we wish to test

$$H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$$

$$H_1 : \lambda_i \neq \lambda_j \qquad i \neq j \quad i,j = 1,2,\ldots,n$$

The likelihood functions of the sample $x_1, x_2, \ldots, x_n$ under H_o and H_1 are $L(H_o)$ and $L(H_1)$, respectively, where

$$L(H_0) = \frac{e^{-n\lambda}\lambda^{T}}{\prod_{i=1}^{n} x_i!}$$

and

$$L(H_1) = \prod_{i=1}^{n} e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!}$$

After the maximizing process we obtain the maximum likelihood functions $\hat{L}(H_o)$ and $\hat{L}(H_1)$ of $L(H_o)$ and $L(H_1)$, respectively, to be

$$\hat{L}(H_0) = \frac{e^{-T}(T/n)^{T}}{\prod_{i=1}^{n} x_i!}$$

and

$$\hat{L}(H_1) = e^{-T} \prod_{i=1}^{n} \frac{x_i^{x_i}}{x_i!}$$

Thus the maximum likelihood ratio is

$$\Lambda = \frac{\hat{L}(H_0)}{\hat{L}(H_1)} = \frac{(T/n)^{T}}{\prod_{i=1}^{n} x_i^{x_i}}$$

so that $$\ln \Lambda = T \ln T - T \ln n - \sum_{i=1}^{n} x_i \ln x_i$$

If T/n is large, $-2 \ln \Lambda$ has an approximate chi-square distribution with $n-1$ degrees of freedom. Then, the maximum likelihood ratio test based on the statistic $-2 \ln \Lambda$ may be used to test H_0 against H_1. The statistics $-2 \ln \Lambda$ and the index of dispersion I are asymptotically equivalent so that neither is preferable to the other when $\bar{x}$ is large. Under H_0 the expectation of a particular value x of the Poisson series $x = 0, 1, 2, \ldots$ is given by

$$m_x = n e^{-\lambda} \frac{\lambda^x}{x!}$$

If the class expectations m_x are small, Cochran [12] suggested that the statistic $-2 \ln \Lambda$ is preferable to I. Also, when class expectations are small Fisher [28] gives examples of samples illustrating the superiority of the statistic $-2 \ln \Lambda$. Rao and Chakravarti [65], in a follow up of Fisher [28]'s work, come to the conclusion that for small samples and/or small $\bar{x}$, then $-2 \ln \Lambda$ has advantages over the index of dispersion I. They remark that for small samples I "tends to be heavily grouped" so that the value of I having "a cumulative probability less than or equal to 5% may actually correspond to a much lower level of significance because of the gaps in I". Using $-2 \ln \Lambda$ "much closer percentages are obtained and consequently it has better chance of rejecting the null hypothesis". A table gives values of $\sum_{i=1}^{n} x_i^2$ such that all values of

$\sum_{i=1}^{m} x_i^2$ greater than or equal to the tabulated values are significant at less than or equal to 5%, for $T = 3\ (1)\ 10$ and $n = 3\ (1)\ 10\ (10)\ 100$. The actual cumulative probabilities accompany each tabulated $\sum_{i=1}^{m} x_i^2$ value and are obtained using the exact probability expression (6). A similar table deals with the statistic $\sum_{i=1}^{m} x_i \ln x_i$ involved in the likelihood test.

c) Conditional Tests

The basic approach of this sectioh was originated by Przyborowski and Wilenski [62] . We first consider Hoel[42]'s follow up of their paper. Let x_1 and x_2 be observations from Poisson populations having means λ_1 and λ_2, respectively. We wish to test the hypothesis

$$\begin{aligned} H_0:\ & \frac{\lambda_2}{\lambda_1} = r \\ H_1:\ & \frac{\lambda_2}{\lambda_1} < r \end{aligned} \tag{8}$$

where r is a specified number. The probability of obtaining x_1 and x_2 is given by

$$p(x_1, x_2) = e^{-\lambda_1} \frac{\lambda_1^{x_1}}{x_1!} e^{-\lambda_2} \frac{\lambda_2^{x_2}}{x_2!}$$

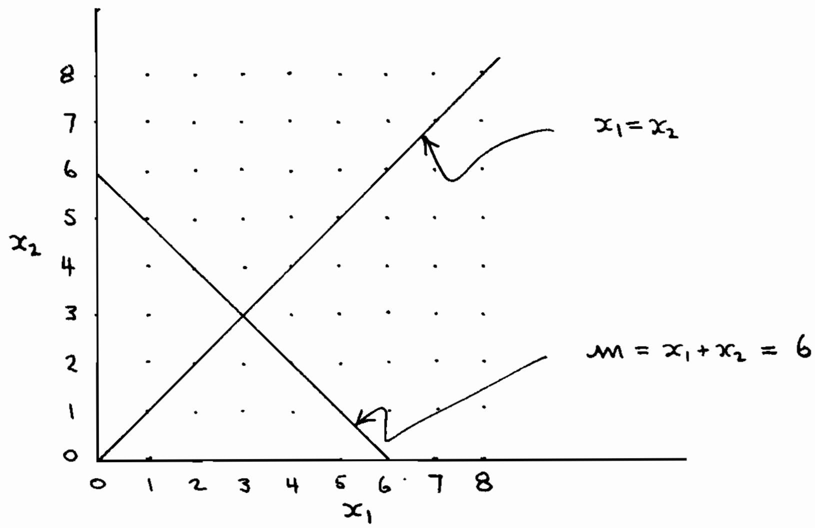
Put $\mu = \lambda_1 + \lambda_2$, $p = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ and $m = x_1 + x_2$. Then we may rewrite $p(x_1, x_2)$ as

$$p(x_1, x_2) = e^{-\mu} \frac{\mu^m}{m!} \frac{m!}{x_1!\,(m-x_1)!} p^{x_1} (1-p)^{m-x_1} \tag{9}$$

and the hypotheses (8) may be rewritten as

$$\begin{aligned} H_0:\ & p = \frac{1}{1+r} \\ H_1:\ & p > \frac{1}{1+r} \end{aligned}$$

A diagram of the sample space follows. It should be noted that for each diagonal line, $m = x_1 + x_2$ equals a positive integer.



The probability $p(x_1, x_2)$ given by (9) can be looked on as the product of two chances, namely,

1) the chance $$p(m \mid \mu) = e^{-\mu} \frac{\mu^m}{m!}$$

the probability that the point (x_1, x_2) falls on the line $m = x_1 + x_2$

2) and the chance $$p(x_1 \mid m, p) = \frac{m!}{x_1!\,(m-x_1)!}\, p^{x_1} (1-p)^{m-x_1} \qquad (10)$$

the probability that for a given m, the observed partition into x_1 and $m - x_1$, occurs.

To test the hypotheses we seek a critical region ω such that

$$P\left\{ (x_1, x_2) \in \omega \;\middle|\; p = \frac{1}{r+1}, \mu \right\} = \alpha$$

where α is the significance level. However, because of the discreteness of x_1 and x_2, no such region exists. If m and p are large, the binomial function (10) can be approximated by a continuous normal function, and an α-region on each line $m = x_1 + x_2$ can be determined so that

$$P\{\text{that a point } (x_1, x_2) \text{ on } m = x_1 + x_2 \text{ will fall in region}\} = \alpha$$

The totality of such α-regions, determined for $m = 0, 1, 2, \ldots,$ constitutes a critical region C of size α, which is independent of μ because the probability that a point (x_1, x_2) lies in the critical region is

$$\sum_{m=0}^{\infty} e^{-\mu} \frac{\mu^m}{m!} \sum_{C} \frac{m!}{x_1!\,(m-x_1)!} p^{x_1} (1-p)^{m-x_1} = \sum_{m=0}^{\infty} e^{-\mu} \frac{\mu^m}{m!} \alpha = \alpha$$

Among all the similar critical regions C that may be determined as outlined there exists a 'best' critical region in the Neyman - Pearson sense, for testing

$$H_0: \; p = p_0$$

$$H_1: \; p = p_1$$

if there exists a best critical region on each diagonal line $m = x_1 + x_2$. The best critical region on such a line, if it exists, is a region satisfying

$$\frac{f(x_1;\, p_0)}{f(x_1;\, p_1)} \leq k \qquad (11)$$

where f is the continuous normal function and k is a constant determined so that the probability under H_0 of a point (x_1, x_2) falling in the critical region is α. Writing condition (11) more explicitly and letting $q_1 = 1 - p_1$ and $q_0 = 1 - p_0$ we have

$$\sqrt{\frac{p_1 q_1}{p_0 q_0}}\; e^{\frac{1}{2}\left[\frac{(x_1 - m p_1)^2}{m p_1 q_1} - \frac{(x_1 - m p_0)^2}{m p_0 q_0}\right]} \leq k$$

which may be altered to

$$e^{\frac{1}{m}\left[\frac{1}{p_1 q_1} - \frac{1}{p_0 q_0}\right]\left[x_1 - \frac{m\left(\frac{1}{q_1} - \frac{1}{q_0}\right)}{\left(\frac{1}{p_1 q_1} - \frac{1}{p_0 q_0}\right)}\right]^2} \leq c \qquad (12)$$

where c is a constant independent of x_1. Let x_0 be the value of x_1 such that

$$P\left\{ x_1 > x_0 \mid p = p_0 \right\} = \alpha \qquad (13)$$

then (12) holds for $x_1 > x_0$ provided $p_1 > p_0$. Thus the region defined by (13) is a 'best' critical region for alternative hypotheses of the form $H_1 : p_1 > p_0$ on the line $m = x_1 + x_2$. The totality of all such regions for $m = 0, 1, 2, \ldots$ constitutes a 'best' critical region A similar treatment applies to testing with alternative hypotheses of the form $H_1 : p_1 < p_0$. Przyborowski and Wilenski [62] deal with the special case where the null hypothesis is $H_0 : r = 1$ and so $H_0 : p = p_0 = 1/2$. They chose to work with the binomial function rather than approximate it by some continuous function. In testing

$$H_0 : p = 1/2$$
$$H_1 : p \neq 1/2$$

they determined a critical region C^* as follows. For each m, $m = 0, 1, 2, \ldots$, let $w(m,\alpha)$ be all the sample points (x_1, x_2) falling on the diagonal line $m = x_1 + x_2$ and satisfying $x_1 \leq k(m,\alpha)$ and $x_1 \geq m - k(m,\alpha)$ where $k(m,\alpha)$ is the largest positive integer such that

$$\sum_{x_1=0}^{k(m,\alpha)} \frac{m!}{x_1!\,(m-x_1)!} \leq \frac{\alpha}{2}$$

The totality of regions $w(m,\alpha)$ for $m = 0, 1, 2, \ldots$ constitutes a critical region of size less than or equal to α.

They state that the test "seems likely to be as efficient as any other alternative test in detecting departures in p from 1/2". A table gives the boundary values $k(m,\alpha)$ for $m = 1\ (1)\ 80$ and $\alpha = 0.01, 0.05, 0.10$ and 0.20. The power function of the test is given by

$$P\left\{(x_1,x_2) \in C^* \,\middle|\, p, \mu\right\} = \sum_{m=0}^{\infty} e^{-\mu} \frac{\mu^m}{m!} \sum_{\omega(m,\alpha)} \frac{m!}{x_1!\,(m-x_1)!}\, p^{x_1}(1-p)^{m-x_1}$$

and has been computed by Przyborowski and Wilenski [62] for $\mu = 2\ (1)\ 15\ (5)\ 50$ and $p = 0.0\ (0.1)\ 0.5$ with $\alpha = 0.10, 0.05$. Hoel [42] points out that the approach of Przyborowski and Wilenski [62] has the disadvantage that special tables or charts are needed. Hoel [42], besides using a continuous approximation for the binomial function, suggests a further modification of Przyborowski and Wilenski [62]. He chooses x_o to be the integer which most nearly satisfies (13), rather than the smallest integer for which the left side does not exceed α. Hoel [42] remarks that there are only two values of m, namely $m = 3$ and 9, for $m \leq 30$ for which the chi-square test and his modification of Przyborowski and Wilenski [62]'s paper might yield different decisions.

In conclusion, it seems that the conditional tests offer nothing that is of any great advantage, and so either the chi-square approach or the likelihood approach it to be preferred.

12.4 The Truncated Poisson Distribution

The special case of truncation at zero will only be considered. From section 5.4 we have for $c = 0$ that the truncated Poisson probabilities are given by

$$P_*(x;\lambda,0) = \frac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})\,x!} \qquad x = 1,2,\cdots \tag{14}$$

First, we derive an exact expression for the probability of the observed frequencies. Let n_x be the number of x values observed from the truncated Poisson population. Let

$$n = \sum_{x=1}^{\infty} n_x$$

and

$$T = \sum_{x=1}^{\infty} x n_x$$

The conditional probability of $n_1, n_2, \ldots$ given n is

$$P\{n_1, n_2, \cdots \mid n\} = \frac{n!}{\prod_{x=1}^{\infty} n_x!} \prod_{x=1}^{\infty} \left(\frac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})\,x!} \right)^{n_x}$$

$$= \frac{n! \; e^{-n\lambda} \; \lambda^T}{\prod_{x=1}^{\infty} n_x!\,(x!)^{n_x} \; (1-e^{-\lambda})^n} \tag{15}$$

The conditional probability of T given n is given by

$$P\{T \mid n\} = \frac{e^{-n\lambda}\,\lambda^T}{(1-e^{-\lambda})^n\;T!} \sum_{j=0}^{n} (-1)^j \binom{n}{j} (n-j)^T \tag{16}$$

which may be obtained by noting that the probability generating function of a truncated Poisson random variable is

$$\sum_{x=1}^{\infty} P_*(x;\lambda,0)\, z^x = \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})\,x!} z^x = \sum_{x=1}^{\infty} \frac{e^{-\lambda}(\lambda z)^x}{(1-e^{-\lambda})\;x!} = \frac{e^{-\lambda}(e^{\lambda z}-1)}{(1-e^{-\lambda})}$$

so that the probability generating function of the random variable T is

$$E\{z^T\} = e^{-n\lambda}\left(e^{\lambda z}-1\right)^n \Big/ \left(1-e^{-\lambda}\right)^n$$

$$= \frac{e^{-n\lambda}}{(1-e^{-\lambda})^n} \sum_{T=0}^{\infty} \lambda^T \sum_{j=0}^{n} (-1)^j \binom{n}{j} \frac{(n-j)^T}{T!} z^T$$

From expressions (15) and (16) we obtain the conditional probability of $n_1, n_2, \ldots$ given T and n to be

$$P\{n_1, n_2, \cdots \mid T, n\} = P\{n_1, n_2, \cdots \mid n\} \Big/ P\{T \mid n\}$$

$$= \frac{n! \quad T!}{\phi(T,n) \prod_{x=1}^{\infty} n_x! \, (x!)^{n_x}} \tag{17}$$

where
$$\phi(T,n) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} (n-j)^T$$
$$= \Delta^n 0^T$$

We note that due to the sufficiency of the statistic T (see Tukey [81]), the expression for $P(n_1, n_2 ,,, \mid T, n)$ is independent of the parameter λ. The expression (17) may be used to compute cumulative probabilities.

If we wish to perform a goodness of fit test on the data we may use either the chi-square statistic

$$\chi^2_T = \sum_{x=1}^{\infty} \frac{(n_x - m_x)^2}{m_x}$$

or the likelihood ratio statistic

$$L_T = \sum_{x=1}^{\infty} n_x \, \ln\left(\frac{n_x}{m_x}\right)$$

where the expected frequencies m_x, $x = 1, 2, \ldots,$ are given by

$$m_x = n \frac{e^{-\lambda} \lambda^x}{(1-e^{-\lambda})\, x!}$$

and the parameter λ is estimated, as we have seen in section 5.4, by the formula

$$\frac{\lambda}{1-e^{-\lambda}} = \frac{T}{m}$$

Both χ^2_T and $-2L_T$ have asymptotic chi-square distributions with $m-1$ degrees of freedom. As in section 12.3 there is little to choose between the two statistics χ^2_T and $-2L_T$ if the class frequencies m_x are large, while the likelihood ratio statistic is preferable if class frequencies are small. The "accuracy" of either statistic may be investigated by computing cumulative probabilities using the expression (17). However, as computations are lengthy, this is only practical for small samples.

We would now like to consider the problem of testing the homogeneity of a random sample $x_1, x_2, \ldots, x_n$. The statistic in the truncated case that corresponds to the Poisson index of dispersion is

$$I_T = \sum_{i=1}^{m} \frac{(x_i - \bar{x}^*)^2}{\bar{x}^*(1+\lambda-\bar{x}^*)}$$

where $\bar{x}^* = \frac{T}{m} = \frac{\lambda}{1 - e^{-\lambda}}$. The statistic I_T has an asymptotic chi-square distribution with $m-1$ degrees of freedom. We shall now develop the likelihood ratio statistic. The likelihood function of the random sample $x_1, x_2, \ldots, x_n$ under the null hypothesis H_0 of a homogeneous sample is

$$L(H_0) = \frac{e^{-m\lambda}\,\lambda^T}{(1-e^{-\lambda})^m \prod_{i=1}^{m} x_i!}$$

The value of λ maximizing $L(H_o)$ satisfies the equation

$$\frac{\lambda}{(1-e^{-\lambda})} = \frac{T}{m}$$

The likelihood function under H_1, the hypothesis of a non-homogeneous sample is

$$L(H_1) = \prod_{i=1}^{m} \frac{e^{-\lambda_i} \lambda_i^{x_i}}{(1-e^{-\lambda_i})\, x_i!}$$

The value of λ_i, $i = 1, 2, \ldots, n$, maximizing $L(H_1)$ satisfies

$$\frac{x_i}{\lambda_i} = \frac{1}{1-e^{-\lambda_i}} \qquad i = 1, 2, \cdots, m$$

Thus, the likelihood ratio becomes

$$\Lambda_T = e^{-m\lambda} \left(\frac{T}{\lambda m}\right)^m \lambda^T \Big/ \prod_{i=1}^{m} e^{-\lambda_i \frac{x_i}{\lambda_i}} \lambda_i^{x_i}$$

The statistic $-2 \ln \Lambda_T$ has an approximate chi-square distribution with $m-1$ degrees of freedom. Thus both the statistics I_T and $-2 \ln \Lambda_T$ may be used to test the homogeneity of the sample. Again, the statistic $-2 \ln \Lambda_T$ is to be preferred to I_T when class expectations are small.

12.5 The Poisson Process

Birnbaum [5] offers solutions to certain problems in hypothesis testing when it is natural to conceive of the observations as originating from a λ-Poisson process. Let T be the amount of observation required for a specified number of events n to occur in a λ-Poisson process. Then we have seen that $2\lambda T$ has a chi-square distribution with 2n degrees of freedom. Thus we may test at significance level α, the hypotheses where $\lambda_o > 0$ is a real number. A number β may be found from chi-square tables such that under the null hypothesis H_o

$$P\{2\lambda_o T > \beta\} = \alpha$$

If the observed value of T exceeds $\beta/2\lambda_o$ we reject $H_o : \lambda = \lambda_o$ in favour of $H_1 : \lambda > \lambda_o$. Otherwise we accept H_o. A similar approach may be used to test $H_o : \lambda = \lambda_o$ against $H_1 : \lambda < \lambda_o$ or $H_1 : \lambda \neq \lambda_o$.

Suppose that two Poisson processes characterized by λ_1 and λ_2 may be observed simultaneously. Let the amount of observation t be the same in both cases. From section 2.6 we have that the number of events observed in the respective processes x_1 and x_2 have Poisson distribution with parameters $\mu_1 = \lambda_1 t$ and $\mu_2 = \lambda_2 t$, respectively. Let

$$\gamma = \frac{\mu_2}{\mu_1} = \frac{\lambda_2}{\lambda_1}$$

We may think of x_1 as the number of "successes" observed in a series of n trials where a "success" refers to an event occuring from the λ_1-Poisson process, so that the probability of a "success" is

$$p = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1}{1+\gamma} \qquad (1B)$$

Thus, we may treat x_1 and x_2 as a sample from the binomial distribution having parameters $m = x_1 + x_2$ and p. We may test $H_o : \mu_1 = \mu_2$ or $\lambda_1 = \lambda_2$ against $H_1 : \mu_1 \neq \mu_2$

or $\lambda_1 \neq \lambda_2$ by testing H_0 : $p = 1/2$ against H_1 : $p \neq 1/2$ because of (18). This may be done using any procedure for testing binomial parameters. For the sake of simplicity Birnbaum [5]'s detailed discussion of the test procedure and of the duration of the experiment is omitted.

Suppose that two Poisson processes, characterized by λ_1 and λ_2 are separated in space or time and that T_1 and T_2 are the amounts of observation required to observe n_1 and n_2 events, respectively. Then $2\lambda_1 T_1$ and $2\lambda_2 T_2$ have chi-square distributions with $2n_1$ and $2n_2$ degrees of freedom, respectively. Then

$$F = \frac{2\lambda_1 T_1}{2n_1} \Big/ \frac{2\lambda_2 T_2}{2n_2} = \frac{n_2 \lambda_1 T_1}{n_1 \lambda_2 T_2}$$

has an F - distribution with $(2n_1, 2n_2)$ degrees of freedom. We wish to test at significance level α, the hypotheses

$$H_0: \lambda_1 = \lambda_2$$
$$H_1: \lambda_1 \neq \lambda_2$$

Under the null hypothesis H_0 we have $F = n_2 T_1 / n_1 T_2$ so that a real number β may be found from F - tables such that

$$P\left\{ \frac{n_2 T_1}{n_1 T_2} > \beta \right\} = \alpha$$

If $\dfrac{n_2 T_1}{n_1 T_2}$, computed from the observations, is greater than β we reject $H_0 : \lambda_1 = \lambda_2$ in favour of H_1. Otherwise we accept H_0.

Wald [82] has developed "the sequential probability ratio test" for testing null hypotheses against one-sided alternatives. We will very briefly outline the testing procedure as applied to the λ-Poisson process and given in Birnbaum [5]. We wish to test at significance level α hypotheses of the form

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda = \lambda_0 + \Delta$$

where $\Delta > 0$ is a real number. Let the variable x be the number of events observed in the λ-Poisson process, and the variable t, the corresponding amount of observation performed. Observe the process only as long as the variables x and t satisfy the inequality

$$b + st < x < a + st$$

where
$$a = \ln\left(\frac{1-\beta}{\alpha}\right) \ln\left(\frac{\lambda_0}{\lambda_0+\Delta}\right)$$

$$b = \ln\left(\frac{\beta}{1-\alpha}\right) \ln\left(\frac{\lambda_0}{\lambda_0+\Delta}\right)$$

and
$$s = \Delta \ln\left(\frac{\lambda_0}{\lambda_0+\Delta}\right)$$

and α is the desired maximum probability of rejecting H_0 when $\lambda \leq \lambda_0$, and β is the desired maximum probability of rejecting H_0 when $\lambda \geq \lambda_0 + \Delta$. Then, observation of the process will be stopped if either $x \leq b + st$ or $x \geq a + st$. If stoppage is because of $x \geq a + st$, we reject $H_0 : \lambda = \lambda_0$ in favour of H_1. Otherwise, we accept H_0.

12.6 Tests of Significance in Discrete Distributions

In this section we deal with discrete probability density functions in general. Let x be a discrete random variable having sample space $x = 0, 1, 2, \ldots, k$ where k is a positive integer or ∞ and probability density function $p(x|\lambda)$ where λ is a single unknown parameter. We wish to test

$$H_0: \lambda = \lambda_0$$
$$H_1: \lambda > \lambda_0$$

Let the desired significance level be α and let the critical region of the test be $C = \left\{ x \mid x = 0, 1, 2, \ldots, c \right\}$ where c is the smallest value of x such that

$$\sum_{x=c}^{k} p(x|\lambda_0) \leq \alpha \tag{19}$$

The cumulative probabilities in (19) may be determined from special tables in the case of the Poisson or binomial distributions. If an observation of x is taken and the value belongs to C the null hypothesis H_0 is rejected in favour of H_1. Although this test is simple it has several drawbacks. First, the size of the test is less than the significance level. This means that the probability of rejecting the null hypothesis H_0 when it is true is smaller than the desired value α. Also, 'best' tests in the sense of Neyman and Pearson are not possible as no two regions can be found having exactly the same size.

To overcome these drawbacks Pearson [57] and Tocher [80] propose the performing of an auxiliary experiment. We present Tocher [80]'s contribution here and Pearson [57]'s, in section 12.7. In both cases the essential idea is that of adding "a continuous dimension" to the sample space of x which permits the application of the theory for continuous sample spaces. We wish to find a 'best' test for testing at significance level α, the hypotheses

$$H_0: \lambda = \lambda_0$$
$$H_1: \lambda = \lambda_1$$

First, let the likelihood ratio Λ_x of the value x be

$$\Lambda_x = p(x \mid \lambda_1) \Big/ p(x \mid \lambda_0) \tag{20}$$

and order the sample points such that

$$\Lambda_1 \geq \Lambda_2 \geq \cdots \tag{21}$$

and such that if $\Lambda_x = \Lambda_{x+1}$, then $p(x \mid \lambda_0) \leq p(x+1 \mid \lambda_0)$. If $\Lambda_x = \Lambda_{x+1}$ and $p(x \mid \lambda_0) = p(x+1 \mid \lambda_0)$ then, $p(x \mid \lambda_1) = p(x+1 \mid \lambda_1)$ and the events x and $x+1$ are equivalent in a certain sense and so may be pooled together and considered to be a single event having probability $2p(x \mid \lambda)$.

Any set of numbers $\{w_x\}$ where $0 \leq w_x \leq 1$ and $x = 1, 2, \ldots, k$ defines a test procedure $\Omega(w_x)$ which rejects H_0 in favour of H_1 with probability w_x if the value x is observed. Let $\beta = \sum_{x=1}^{k} w_x \, p(x \mid \lambda_1)$.

We desire a test procedure $\Omega(\hat{w}_x)$ such that the set $\{\hat{w}_x\}$ satisfies the two conditions

$$\sum_{x=1}^{k} \hat{\omega}_x \, p(x \mid \lambda_0) \leq \alpha$$

and

$$\sum_{x=1}^{k} \hat{\omega}_x \, p(x \mid \lambda_1) \quad \text{is a maximum}$$

Define $\{\hat{w}_x\}$ such that

$$\hat{\omega}_x = \begin{cases} 1 & x = 1, 2, \cdots, x_0 \\ \theta = \dfrac{\alpha - \sum_{x=1}^{x_0} p(x \mid \lambda_0)}{p(x_0+1 \mid \lambda_0)} & x = x_0+1 \\ 0 & x \geq x_0+2 \end{cases} \tag{23}$$

where x_0 is the value of x satisfying

$$\sum_{x=1}^{x_0} p(x \mid \lambda_0) \leq \alpha < \sum_{x=1}^{x_0+1} p(x \mid \lambda_0)$$

Let $\hat{\beta} = \sum_{x=1}^{k} \hat{w}_x \, p(x \mid \lambda_1)$. We now show that $\hat{\beta} \geq \beta$.

First,

$$\hat{\beta} - \beta = \sum_{x=1}^{x_0} (1-\omega_x)\, p(x \mid \lambda_1) + \left[\frac{\alpha - \sum_{x=1}^{x_0} p(x \mid \lambda_0)}{p(x_0+1 \mid \lambda_0)} \right] p(x_0+1 \mid \lambda_1) - \sum_{x=x_0+1}^{k} \omega_x \, p(x \mid \lambda_1)$$

so that

$$\hat{\beta} - \beta \geq \sum_{x=1}^{x_0} (1-\omega_x)(\Lambda_x - \Lambda_{x+1})\, p(x \mid \lambda_0) + \sum_{x=x_0+1}^{k} \omega_x (\Lambda_{x+1} - \Lambda_x)\, p(x \mid \lambda_0) \tag{24}$$

From (21) we have that the right-hand side of (24) is a positive quantity, so that $\hat{\beta} \geq \beta$. The equality holds only if

$$\sum_{x=1}^{x_0} \omega_x \, p(x \mid \lambda_0) = \alpha$$

and

$$\begin{cases} 1 - \omega_x = 0 & x = 1, 2, \cdots, x_0 \\ \omega_x = 0 & x = x_0+2, \cdots, k \end{cases}$$

Therefore, equally powerful test procedures only differ from $\Omega(\hat{w}_x)$ in the w values assigned to the value $x = x_0 + 1$. The test procedure is as follows. Take an observation, and call it x^0. If $x^0 \in \left\{ x \mid x = 0, 1, 2, \ldots, x_0 \right\}$

reject H_o. If $x^o = x_o + 1$, perform the auxiliary experiment by selecting a number between 0 and 1 from a table of random numbers, and call it z^o. If

$$z^{\circ} \leq \theta = \frac{\alpha - \sum_{x=1}^{x_0} p(x \mid \lambda_0)}{p(x_0+1 \mid \lambda_0)} \tag{25}$$

reject H_o in favour of H_1. We also note that

$$P\left\{C \mid H_0\right\} = \sum_{x=1}^{k} \hat{\omega}_x \, p(x \mid \lambda_0) = \alpha$$

Thus the test $\Omega(\hat{w}_x)$ has the following two points in its favour :

1) The probability of rejecting the null hypothesis H_o when it is true is known exactly and equals the desired value α.

2) The test $\Omega(\hat{w}_x)$ is a 'best' test in the sense of Neyman and Pearson.

Tocher [80] extends the concepts of Neyman and Pearson in testing to include unbiased test procedures and also, testing composite hypotheses.

To counterbalance the advantages of the auxiliary random experiment there are two disadvantages :

1) From the same experimental data different decisions as to the rejection or acceptance of the null hypothesis are possible.

2) Under certain experimental conditions computations may prove troublesome (see Lancaster [46]). If the probability $p(x_o + 1 \mid \lambda_o)$ is large, the number of observations of $x_o + 1$ may be large, so that the auxiliary random experi-

ment may have to be used repeatedly, thus it is possible that the effect of the auxiliary random experiment may be greater than the effect of the remainder of the experimental data in the decision to reject or accept the null hypothesis. Lancaster [46] offers an approach which avoids these difficulties. Lancaster [46] selects the median probability $P_m(x)$ defined by

$$P_m(x) = \frac{\sum_{j=1}^{x} p(j \mid \lambda_0) + \sum_{j=1}^{x-1} p(j \mid \lambda_0)}{2} \tag{26}$$

as test function. If x^0 is the observed value of x the rule of rejection is to reject H_0 if

$$P_m(x^0) \leq \alpha \tag{27}$$

When this rule is applied to the observation $x_0 + 1$ it is equivalent to a rule of rejection when $\theta \geq 1/2$, where θ is given by (25). The decision based on the rule (27) will always be the same for given H_0 and a given set of experiemntal data. Lancaster [46] considers the agreement between the auxiliary random experiment and the median probability approach and concludes that there is 75 % agreement in marginal cases and 100% agreement in all other cases. Both approaches may involve lengthy computations, however the median probability has a "good approximation" in the crude chi-square distribution as is shown in Lancaster [46] .

12.7 The Combination of Tests

As conditions often differ from experiment to experiment it is desirable to have some method of combining a number of independent experiments which have all been planned to test a common hypothesis. If continuous distributions are involved there is little difficulty; however, the matter is not so simple in the case of discrete distributions. We shall deal with discrete distributions in general.

Lancaster [47] mentions the need for an adequate procedure for combining independent experiments when the number of observations in any experiment is small and discrete distributions are involved.

It will be helpful to have in mind the procedure generally used in the combination of tests involving a continuous distribution. Let x be a continuous random variable with probability density function $p(x)$ defined as a positive quantity for $a \leq x \leq b$ and zero elsewhere. Then a random variable y defined by

$$y = \int p(x)\,dx$$

has the uniform distribution in $[0, 1]$. Clearly, the random variable $1 - y$ is also uniformly distributed in $[0, 1]$. It is then easy to show that the random variables $-2 \ln y$ and $-2 \ln (1-y)$ each have chi-square distributions with 2 degrees of freedom. Also, if M is the median of the distribution $p(x)$ defined such that

$$\int_a^M p(x)\,dx = \int_M^b p(x)\,dx = \frac{1}{2}$$

then the random variable y' defined by

$$y' = \begin{cases} 2\int_a^x p(x)\,dx & x \leq M \\ 2\int_x^b p(x)\,dx & x > M \end{cases}$$

is uniformly distributed in $[0, 1]$, so that $-2 \ln y'$ has a chi-square distribution with 2 degrees of freedom. These results may be used to combine a number of independent tests as follows. Let x_i be the continuous test statistic of the i^{th} experiment where $i = 1, 2, \ldots, l$ and let $p(x_i \mid H_o)$ be its probability density function under a null hypothesis H_o, common to the l experiments. Define y_i and y_i' for $i = 1, 2, \ldots, l$ by

$$y_i = \int_{a_i}^{x_i} p(x_i \mid H_o)\,dx_i$$

and

$$y_i' = \begin{cases} 2\int_{a_i}^{x_i} p(x_i \mid H_o)\,dx_i & x_i \leq M_i \\ 2\int_{x_i}^{b_i} p(x_i \mid H_o)\,dx_i & x_i > M_i \end{cases}$$

where M_i is the median of the distribution given by $p(x_i \mid H_o)$ and a_i and b_i are related to x_i as a and b are related to x. Then the three statistics Q_1, Q_2, Q_3 defined by

$$Q_1 = -2\sum_{i=1}^{l} \ln y_i \quad , \quad Q_2 = -2\sum_{i=1}^{l} \ln(1-y_i)$$

and
$$Q_3 = -2 \sum_{i=1}^{\ell} \ln y_i'$$

each have a chi-square distribution with 2ℓ degrees of freedom. The statistics Q_1 and Q_2 are used in single-tail tests while Q_3 is used in two-tail tests.

The question arises as to whether or not a similar procedure may be used for discontinuous distributions. David and Johnson [20] have partly answered the question. Let x be a discrete random variable taking on values $1, 2, \ldots, k$ with probabilities p_1, $p_2, \ldots, p_k$ where k is finite or ∞ and $\sum_{x=1}^{k} p_x = 1$. Define v_i and u_i by

$$v_i = \sum_{x=1}^{i} p_x \tag{28}$$

and
$$u_i = \sum_{x=1}^{i-1} p_x + \frac{p_i}{2} \qquad i = 1, 2, \cdots, k \tag{29}$$

Then we define the new random variables v and u as taking on values v_i and u_i, respectively, for $i = 1, 2, \ldots, k$. It is easy to verify that
$$P\left\{ v = v_i \right\} = p_i$$

and
$$P\left\{ u = u_i \right\} = p_i \qquad i = 1, 2, \cdots, k$$

and also that
$$E\left\{ v \right\} = \sum_{i=1}^{k} \left(\sum_{x=1}^{i} p_x \right) p_i = \frac{1}{2}\left(1 + \sum_{i=1}^{k} p_i^2 \right)$$

and
$$E\left\{ u \right\} = \sum_{i=1}^{k} \left(\sum_{x=1}^{i-1} p_x + \frac{p_i}{2} \right) p_i = \frac{1}{2}$$

and
$$E\left\{ u^r \right\} = \sum_{i=1}^{k} \left(\sum_{x=1}^{i-1} p_x + \frac{p_i}{2} \right)^r p_i \qquad r = 1, 2, \cdots$$

and in particular $$V\{u\} = \frac{1}{12}\left(1 - \sum_{i=1}^{k} p_i^3\right)$$

It is interesting to examine these results and compare them with the continuous case. First, the random variable v is defined in an analogous manner to the continuous random variable y. However, the expectation of v is greater than $1/2$, the expectation of y. The random variable u is defined as a slight modification of v, so as to have expectation equal to $1/2$. The variance of u is slightly smaller than $1/12$, the variance of y. As $k \to \infty$ the moments of u approach those of a random variable that is uniformly distributed in $[0, 1]$. David and Johnson [20] numerically investigate the departure of u and v from rectangularity for x both a Poisson and a binomial random variable. The conclusion is that the assumption of rectangularity for u and v may be misleading. The minimum departure occurs when all the p_i values are equal. The random variable $-2 \ln v$ is analogous to the continuous random variable $-2 \ln y$ which has a chi-square distribution with 2 degrees of freedom. After some manipulation we obtain the $r = 1, 2, \ldots$

$$E\left\{(-\ln u)^r\right\} = (-1)^r \sum_{i=1}^{k} \left[\ln\left(1 - \sum_{x=i+1}^{k} p_x - \frac{p_i}{2}\right)\right]^r p_i$$

$$= (-1)^r \sum_{\omega=1}^{\infty} \frac{(1/2)^{\omega-1}}{(\omega-1)!} \sum_{i=1}^{k} p_i^{\omega} \frac{d^{\omega-1}}{d\left(\sum_{x=i+1}^{k} p_x\right)^{\omega-1}} \left[\ln\left(1 - \sum_{x=i+1}^{k} p_x\right)\right]^r \qquad (30)$$

where d is an operator defined by $e^d = E$ and E is the shift operator defined on any function of n, $f(n)$, such that $E\, f(n) = f(n+1)$. The cumulants of $-2 \ln u$ may be obtained from (30), although the resulting expressions are rather long. As $k \to \infty$ and $p_i \to 0$ from the positive direction for $i = 1, 2, \ldots, k$ the moments of $-2 \ln u$ tend to those of a random variable having a chi-square distribution with 2 degrees of freedom. That is

$$\lim_{k\to\infty,\, p_i \to 0} E\left\{-2 \ln u\right\}^{\tau} = 2^{\tau}\, \tau!$$

The statistic $-2 \ln u$ is also considered by Lancaster [47] where it is denoted by χ'^2_m. The mean value of $-2 \ln v$ on the assumption that v is uniformly distributed is denoted by χ^2_m by Lancaster [47] and is given by

$$\chi^2_m = \int_{v_{i-1}}^{v_i} (-2 \ln v)\, dv \Big/ (v_i - v_{i+1})$$

$$= 2 - 2\left[\frac{v_i \ln v_i - v_{i-1} \ln v_{i-1}}{v_i - v_{i-1}}\right]$$

where v_i is the observed value of v. According to Lancaster [47] both χ^2_m and χ'^2_m may be approximated by the chi-square distribution with 2 degrees of freedom. The statistic χ^2_m has mean equal to 2 and variance slightly less than 4. The quantities χ^2_m or χ'^2_m may be obtained for each of the l independent experiments and the sum of the χ^2_m's or the χ'^2_m's will have an approximate chi-square distribution with 2ℓ degrees of freedom. The statistic χ'^2_m is to be preferred to χ^2_m because it can be more easily evaluated. Both χ'^2_m and χ^2_m are superior to $-2 \ln v$ looked upon as a chi-square random variable.

A radically different approach to the problem of combining independent experiments follows from the work of Eudey [23], Stevens [75], Tocher [80], and Pearson [57]. The basic idea is to add a continuous uniformly distributed random variable to the discrete random variable, so that the sum will be a continuous random variable, and the procedure already outlined for continuous random variables, applicable. The idea has already been discussed in section 11.3 and 12.6. Let x be a discrete random variable having probability density function $p(x)$ where $x = 1, 2, \ldots, k$ and k is finite or ∞. Let z be a continuous random variable uniformly distributed on $[0, 1]$ and independent of x. Define a random variable y as in section 11.3 by

$$y = x + z$$

We note that y is a continuous random variable. Now

$$P\{y \leq y_0\} = P\{x < x_0\} + z_0\, p(x_0)$$
$$= \sum_{x=1}^{x_0-1} p(x) + z_0\, p(x_0)$$

for any values x_0 and z_0 of x and z, respectively. Drop the zeroes on x_0 and z_0 and define $y(x,z)$ by

$$y(x,z) = \sum_{j=1}^{x-1} p(j) + z\, p(x)$$

Then $y(x,z)$ is a continuous random variable and is uniformly distributed on $[0, 1]$. Then $-2 \ln y(x,z)$ is a chi-square random variable having 2 degrees of freedom, so that a number of independent experiments may be combined. It is interesting to note that the mean of $-2 \ln y(x,z)$ is χ^2_m and that the median of $-2 \ln y(x,z)$ is χ'^2_m. Pearson [57] also provides the following results for fixed x.

$$E\{\text{mean of } -2 \ln y(x,z)\} = 2$$

$$V\{\text{mean of } -2 \ln y(x,z)\} < 4$$

$$E\{\text{median of } -2 \ln y(x,z)\} \text{ is a little less than } 4$$

and

$$V\{\text{median of } -2 \ln y(x,z)\} < 4$$

These are to be compared with 2 and 4, the mean and variance, respectively, of a random variable having the chi-square distribution with 2 degrees of freedom.

We conclude this chapter with a brief comparison of the approaches offered by Lancaster [47] and Pearson [57]. When small samples are involved there is some uncertainty as to just how accurate Lancaster [47] 's two statistics are. Tocher [80] has shown that for a fixed significance level the test based on $y(x,z)$ is more powerful in the sense of Neyman and Pearson than any other test. Also the test procedure is as quick to carry out as that involving χ'^2_m, and much quicker than that of χ^2_m. The main argument against the use of $y(x,z)$ is that it seems difficult to have to make a decision to reject or accept a null hypothesis on the basis of an auxiliary random experiment which is in a sense independent of the observations of the random variable x. Either Lancaster [47]'s χ'^2_m or Pearson [57]'s $-2 \ln y(x,z)$ may be used in combining a number of independent tests, depending upon the specific nature of the tests, the accuracy desired, and the experimenters convictions regarding the use of the auxiliary random experiment.

Chapter Thirteen

Topics for Original Research

The complete Poisson distribution has been thoroughly investigated, and excellent tables have been formed of probabilities, estimates and confidence intervals. There is still research to be done with the truncated and censored Poisson distributions.

The Poisson distribution truncated on the right was discussed in section 5.3 and it was noted that an unbiased estimate for the parameter λ in this case does not exist. Now, for any integer $m \leq d$ where d is the truncation point, we have that

$$\lambda = \frac{\sum_{x=0}^{m} x e^{-\lambda} \frac{\lambda^x}{x!}}{\sum_{x=0}^{m-1} e^{-\lambda} \frac{\lambda^x}{x!}}$$

so that

$$\frac{\sum_{x=0}^{m} x n_x}{\sum_{x=0}^{m-1} n_x}$$

is an estimate for λ, where n_x is the number of x values observed in a sample. An interesting aspect of this estimate for λ is that it is independent of the truncation point d. By putting $m = d$ we obtain Moore [52] 's estimate λ_M. If the truncation point d

is known, the maximum likelihood estimate $\hat{\lambda}$ may be obtained from Cohen [13]'s tables. If d is unknown, the maximum likelihood estimate of d is

$$\hat{d} = \max\{x_1, x_2, \cdots, x_n\}$$

where $x_1, x_2, \ldots, x_n$ is a random sample of size n. An investigation of $\hat{d}$ may be worthwhile. First, we determine an expression for the probabilities of $\hat{d}$

$$P\{\hat{d} = \alpha\} = \sum_{k=1}^{n} \binom{n}{k} P\{x_j < \alpha \;\forall j \neq 1,2,\ldots,k \;;\; x_1 = \cdots = x_k = \alpha\}$$

$$= \sum_{k=1}^{n} \binom{n}{k} \left(\sum_{i=0}^{\alpha-1} \frac{e^{-\lambda}\lambda^{i}}{F(d)\, i!} \right)^{n-k} \left(\frac{e^{-\lambda}\lambda^{\alpha}}{F(d)\, \alpha!} \right)^{k}$$

$$= \sum_{k=1}^{n} \binom{n}{k} \left(\frac{F(\alpha-1)}{F(d)} \right)^{n-k} \left(\frac{e^{-\lambda}\lambda^{\alpha}}{F(d)\, \alpha!} \right)^{k}$$

$$= \left(\frac{F(\alpha-1)}{F(d)} + \frac{e^{-\lambda}\lambda^{\alpha}}{F(d)\, \alpha!} \right)^{n} - \left(\frac{F(\alpha-1)}{F(d)} \right)^{n}$$

$$= \frac{\left(F(\alpha)\right)^{n} - \left(F(\alpha-1)\right)^{n}}{\left(F(d)\right)^{n}}$$

where

$$F(d) = \sum_{i=0}^{d} e^{-\lambda} \frac{\lambda^{i}}{i!}$$

Since $\hat{d}$ is an estimate of d we are interested in the probability that $\hat{d}$ takes on the value d, that is

$$P\{\hat{d} = d\} = 1 - \left(\frac{F(d-1)}{F(d)}\right)^m$$

We may note that

$$\lim_{m \to \infty} P\{\hat{d} = d\} = \lim_{m \to \infty} \left[1 - \left(\frac{F(d-1)}{F(d)}\right)^m\right] = 1$$

thus $\hat{d}$ is a consistent estimate of d. The mean of $\hat{d}$ is obtained as follows

$$E\{\hat{d}\} = \sum_{\alpha=0}^{d} \alpha\, P\{\hat{d} = \alpha\} = \sum_{\alpha=0}^{d} \alpha \left[\frac{(F(\alpha))^m - (F(\alpha-1))^m}{(F(d))^m}\right]$$

$$= d - \sum_{j=0}^{d-1} \left(\frac{F(j)}{F(d)}\right)^m$$

Thus $\hat{d}$ is an "under-estimate" of d having a negative bias. However, we have that

$$\lim_{m \to \infty} E\{\hat{d}\} = d$$

The variance of $\hat{d}$ after a little manipulation is found to be

$$V\{\hat{d}\} = \sum_{j=0}^{d-1} (2d - 2j - 1)\left(\frac{F(j)}{F(d)}\right)^m - \left[\sum_{j=0}^{d-1}\left(\frac{F(j)}{F(d)}\right)^m\right]^2$$

In summary of the case of truncation on the right, if both d and λ are unknown, d may be estimated by $\hat{d}$, and then $\hat{\lambda}$ may be obtained from Cohen [13]'s tables.

The Poisson distribution truncated on the left at c was discussed in section 5.4. If the truncation point c is known two simple estimates for λ are Rider [68]'s λ_R and the unbiased estimate given in Tate and Goen [77], $V_c(x_1, x_2, \ldots, x_n)$. The maximum likelihood estimate of λ, $\hat{\lambda}$ is the solution for λ of equation (28) in section 5.4, that is

$$\bar{x} = \lambda \frac{\left(1 - F(c-1)\right)}{\left(1 - F(c)\right)}$$

However, tables are only available for the case where $c = 0$ (see Cohen [14]). Complete tables giving λ as a function of $\bar{x}$ for different values of c would be valuable. If the truncation point c is unknown $c + 1$ may be estimated by the maximum likelihood estimate

$$(\widehat{c+1}) = \min\left\{x_1, x_2, \cdots, x_n\right\}$$

where $x_1, x_2, \ldots, x_n$ is a random sample of size n. We shall now examine $(\widehat{c + 1})$ briefly. The probabilities of $(\widehat{c + 1})$ are given by

$$P\left\{\widehat{c+1} = \alpha\right\} = P\left\{\min\left\{x_1, x_2, \cdots, x_n\right\} = \alpha\right\}$$

$$= \sum_{k=1}^{n} \binom{n}{k} P\left\{x_j > \alpha \ \ \forall j \neq 1, 2, \cdots, k \ ; \ x_1 = \cdots = x_k = \alpha\right\}$$

$$= \sum_{k=1}^{n} \binom{n}{k} \left(\sum_{j=\alpha+1}^{\infty} \frac{e^{-\lambda}\lambda^{j}}{\left(1-F(c)\right) j!}\right)^{n-k} \left(\frac{e^{-\lambda}\lambda^{\alpha}}{\left(1-F(c)\right)\alpha!}\right)^{k}$$

$$= \sum_{k=1}^{n} \binom{n}{k} \left(\frac{1-F(\alpha)}{1-F(c)}\right)^{n-k} \left(\frac{e^{-\lambda}\lambda^{\alpha}}{\left(1-F(c)\right)\alpha!}\right)^{k}$$

$$= \left(\frac{1-F(\alpha)}{1-F(c)} + \frac{e^{-\lambda}\lambda^{\alpha}}{(1-F(c))\,\alpha!} \right)^{n} - \left(\frac{1-F(\alpha)}{1-F(c)} \right)^{n}$$

$$= \frac{\left(1-F(\alpha-1)\right)^{n} - \left(1-F(\alpha)\right)^{n}}{\left(1-F(c)\right)^{n}}$$

The probability that the estimate $(\widehat{c+1})$ takes on the value $c + 1$ is then given by

$$P\left\{ \widehat{c+1} = c+1 \right\} = 1 - \left(\frac{1-F(c+1)}{1-F(c)} \right)^{n}$$

We may also note that

$$\lim_{n \to \infty} P\left\{ \widehat{c+1} = c+1 \right\} = \lim_{n \to \infty} \left[1 - \left(\frac{1-F(c+1)}{1-F(c)} \right)^{n} \right] = 1$$

We now obtain the mean of the estimate $(\widehat{c+1})$

$$E\left\{ \widehat{c+1} \right\} = \sum_{\alpha = c+1}^{\infty} \alpha \left[\frac{\left(1-F(\alpha-1)\right)^{n} - \left(1-F(\alpha)\right)^{n}}{\left(1-F(c)\right)^{n}} \right]$$

$$= (c+1) + \sum_{j=c+1}^{\infty} \left(\frac{1-F(j)}{1-F(c)} \right)^{n}$$

so that $(\widehat{c+1})$ is an "over-estimate" of $c + 1$ having a positive bias. The variance of $(\widehat{c+1})$ is found to be

$$V\{\widehat{c+1}\} = \sum_{j=c+1}^{\infty} (2j-2c-1)\left(\frac{1-F(j)}{1-F(c)}\right)^{m} - \left[\sum_{j=c+1}^{\infty}\left(\frac{1-F(j)}{1-F(c)}\right)^{m}\right]^2$$

Once an estimate for c has been obtained, an estimate for λ may be obtained by the methods already discussed.

In section 6.2 the censored Poisson distribution was discussed. Let n_x be the number of x values observed in a random sample of size n. Then if n_x is known for all $x \leq c$ and only the total number of observations is known for $x > c$ Murakami et al [53] have constructed nomograms to facilitate determination of the maximum likelihood estimate of λ. However, these graphs are not very accurate, and tables would be more useful. Other cases of censoring, such as the case where n_x is known for $x \geq c$ and only the total number of observations is known for $x < c$, may be treated in a similar manner.

The generalized Poisson distribution obtained by assuming that the parameter λ of a Poisson distribution has the normal distribution truncated away from negative real numbers may be considered. If we impose the restricting condition that the mean and variance be equal, a relatively simple generalized distribution results. No details are given as it is not at all certain at this time that this idea is of practical value or not. If the idea is useful, there is of course, a large number of estimation and fitting problems to be considered.

Bibliography

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

2. Barankin, E.W. and Gurland, J., "On asymptotically normal, efficient estimations", University of California Publications in Statistics (1951) pp. 89 - 129

3. Beall, G. and Rescia, R., "A generalization of Neyman's contagious distributions", Biometrics (1953) pp. 354 - 386

4. Birnbaum, A., "Some procedures for comparing Poisson processes or populations", Biometrika (1953) pp. 447 - 449

5. Birnbaum, A., "Statistical methods for Poisson processes and exponential populations", J.A.S.A. (1954) pp. 254 - 266

6. Blackwell, D., "Conditional expectation and unbiased sequential estimation", A.M.S. (1947) pp. 105 - 110

7. Bliss, C.I., "Fitting the negative binomial distribution to biological data", Biometrics (1953) pp. 176 - 196

8. Bliss, C.I., "Estimation of the mean and its error from incomplete Poisson distributions", Connecticut Agricultural Experiment Station Bulletin 513 (1948)

9. Blyth, C.R. and Hutchinson, D.W., "Table of Neyman - shortest unbiased confidence intervals for the Poisson parameter", Biometrika (1961) pp. 191 - 194

10. Blyth, C.R. and Hutchinson, D.W., "Table of Neyman - shortest unbiased confidence intervals for the binomial parameter", Biometrika (1960) pp. 381 - 391

11. Bortkiewicz, L. von, "Das Gesetz der Kleinen Zahlen", Leipzig : Teubner, 1898

12. Cochran, W.G., "The chi-square distribution for the binomial and Poisson series with small expectations", Annals of Eugenics (1936) pp. 207 - 217

13. Cohen, A.C. Jr., "Estimating the Poisson parameter from samples that are truncated on the right", Technometrics (1961)

14. Cohen, A.C. Jr., "Estimating the parameter in a conditional Poisson distribution", Biometrics (1960) pp. 203 - 211

15. Cohen, A.C. Jr., "An extension of a truncated Poisson distribution", Biometrics (1960) pp. 446 - 450

16. Cohen, A.C. Jr., "Estimation in the truncated Poisson distribution when zeroes and some ones are missing", J.A.S.A. (1960) pp. 342 - 348

17. Cohen, A.C. Jr., "Estimating the parameters of a modified Poisson distribution", J.A.S.A. (1960) pp. 139 - 143

18. Cramér, H., "Mathematical Methods of Statistics", Princeton University Press, 1945

19. Crow, E.L. and Gardner, R.S., "Confidence intervals for the expectation of a Poisson variate", Biometrika (1959) pp. 441 - 453

20. David, F.N. and Johnson, N.L., "The probability integral transformation when the variable is discontinuous", Biometrika (1950) pp. 42 - 49

21. David, F.M. and Johnson, N.L., "The truncated Poisson", Biometrics (1952) pp. 275 - 285

22. Douglas, J.B., "Fitting the Neyman type A (two parameter) contagious distribution", Biometrics (1955) pp. 149 - 173

23. Eudey, M.W., "On the treatment of discontinuous random variables", Technical Report No. 13, Statistical Laboratory, University of California, Berkeley

24. Feller, W., "On a general class of 'contagious' distributions", A.M.S. (1943) pp. 389 - 400

25. Feller, W., "An Introduction to Probability Theory and Its Applications" Vol 1 2nd ed., John Wiley, New York, 1957

26. Finney, G.C. and Varley, D.J., "An example of the truncated Poisson distribution", Biometrics (1955) pp. 387 - 394

27. Fisher, R.A., "Note on the efficient fitting of the negative binomial", Biometrics (1953) pp. 197 - 200

28. Fisher, R.A., "The significance of deviations from expectations in a Poisson series", Biometrics (1950)

29. Fisher, R.A. and Yates, F., "Statistical Tables for Biological, Agricultural and Medical Research", 4th ed., Oliver and Boyd, Edinburgh, 1953

30. Garman, P., "Original data on European red mite on apple leaves", Connecticut, 1951

31. Garwood, F., "Fiducial limits for the Poisson distribution", Biometrika (1936) pp. 437 - 442

32. Goodall, D.W., unpublished data on leaf counts of Leucopogon Virgatus, University of Melbourne

33. Gurland, J., "Some interrelations among compound and generalized distributions", Biometrika (1957) pp. 265 - 266

34. Gurland, J., "A generalized class of contagious distributions", Biometrics (1958) pp. 229 - 249

35. Gurland, J. and Katti, S.K., "Some methods of estimation for the Poisson binomial distribution", Technical Report No. 212, Mathematical Research Center, United States Army, University of Wisconsin (1960)

36. Gurland, J. and Katti, S.K., "The Poisson Pascal distribution", Biometrics (1961) pp. 527 - 538

37. Gurland, J. and Shumway, R., "A fitting procedure for some generalized Poisson distributions", Technical Report No. 205, Mathematics Research Center, United States Army, University of Wisconsin (1960)

38. Gurland, J. and Shumway, R., "Fitting the Poisson binomial distribution", Biometrics (1960) pp. 522 - 533

39. Guttman, I., "A note on a series solution of a problem in estimation", Biometrika (1958) pp. 565 - 567

40. Hartley, H.O., "Maximum likelihood estimation from incomplete data", Biometrics (1958) pp. 174 - 194

41. Hoel, "On indices of dispersion", A.M.S. (1943) pp. 155 - 162

42. Hoel, "Testing the homogeneity of Poisson frequencies", A.M.S. (1945) pp. 362 - 368

43. Irwin, J.O., "Note on the estimation of the mean of a Poisson distribution from a sample with the zero class missing", Biometrics (1959) pp. 324 - 326

44. Jordon, C., "Calculus of Finite Differences", Chelsea, New York, 1950

45. Kitagawa, T., "Tables of the Poisson Distribution", Tokyo : Baifukan, 1952

46. Lancaster, H.O., "Significant tests in discrete distributions", J.A.S.A. (1961) pp. 223 - 234

47. Lancaster, H.O., "The combination of probabilities arising from data in discrete distributions", Biometrika (1949) pp. 370 - 382

48. Lehmann, E.L. and Scheffé, H., "Completeness, similar regions and unbiased estimation", Part I Sankhya (1950) pp. 305 - 340

49. McGuire, J.U., Brindley, T.A. and Bancroft, T.A., "The distribution of European corn-borer larvae Pyrausta Nubilalis (H bn) in field corn", Biometrics (1956) pp. 65 - 78

50. Miksa, F.L., as yet unpublished table of Stirling numbers of the second kind

51. Molina, E.C., "Poisson's Exponential Binomial Limit", D. Van Nostrand, 1942

52. Moore, P.G., "The estimation of the Poisson parameter

from a truncated distribution", Biometrika (1952) pp. 247 - 251

53. Murakami, M., Asai, A., and Kawamura, M., "The estimation of the Poisson parameter from a truncated distribution and a censored sample", Journal of the College of Arts and Sciences, Chiba University (1954)

54. Neyman, J., "On a new class of 'contagious' distributions, applicable in entomology and bacteriology", A.M.S. (1939) pp. 35 - 57

55. Neyman, J., "Outline of a theory of statistical estimation based on the classical theory of probability", Phil. Trans. (1937) pp. 333

56. Noak, A., "A class of random variables with discrete distributions", A.M.S. (1950) pp. 127 - 132

57. Pearson, E.S., "On questions raised by the combination of tests based on discontinuous distributions", Biometrika (1950) pp. 383 - 398

58. Pearson, E.S. and Clopper. C.J., "The use of confidence or fiducial limits illustrated in the case of the binomial", Biometrika (1934) pp. 404 - 413

59. Pearson, E.S. and Hartley, H.O., "Tables of the chi-square integral and the cumulative Poisson distribution", Biometrika (1950) pp. 313 - 325

60. Plackett, R.L., "The truncated Poisson distribution", Biometrics (1953) pp. 485 - 488

61. Poisson, S.D., "Recherches sur la probabilite des jugements en matiere criminelles et en matiere civile", Paris, France, 1837

62. Przyborowski, J. and Wilenski, H., "Homogeneity of results in testing samples from Poisson series", Biometrika (1940) pp. 313 - 323

63. Raff, M.S., "On approximating the Point Binomial", J.A.S.A. (1956) pp. 293 - 303

64. Rao, C.R., "Information and accuracy attainable in the estimation of statistical parameters", Bull. Cal. Math. Soc. (1945) pp. 81 - 91

65. Rao, C.R. amd Chakravarti, I.M., "Some small sample tests of significance for a Poisson distribution", Biometrics (1956) pp. 264 - 282

66. Richardson, C.H., "An Introduction to the Calculus of Finite Differences", D. Van Nostrand Co., Inc. New York, 1954

67. Ricker, W.E., "The concept of confidence or fiducial limits applied to the Poisson frequency distribution", J.A.S.A. (1937) pp. 349 - 356

68. Rider, P.R., "Truncated Poisson distributions", J.A.S.A. (1953) pp. 826 - 830

69. Riordon, J., "An Introduction to Combinatorial Analysis", John Wiley, New York, 1958

70. Roy, J. and Mitra, S.K., "Unbiased minimum variance estimation in a class of discrete distributions", Sankhya (1957) pp. 755 - 765

71. Shenton, L.R., "On the efficiency of the method of moments and Neyman's type A distribution", Biometrics (1949) pp. 450 - 454

72. Soper, H.E., table of the general term of Poisson's exponential expansion, Biometrika (1918) p. 25

73. Sprott, D.A., "The method od maximum likelihood

applied to the Poisson binomial distribution", Biometrics (1958) pp. 97 - 106

74. Sterne, T.E., "Some remarks on confidence or fiducial limits", Biometrika (1954) pp. 275 - 278

75. Stevens, W.L., "Fiducial limits of the parameter of a discontinuous distribution", Biometrika (1950) pp. 117 - 129

76. Stevens, W., "Significance of grouping", Annals of Eugenics (1937) p. 57

77. Tate, R.F. and Goen, R.L., "Minimum variance unbiased estimation for the truncated Poisson distribution", A.M.S. (1958) pp. 755 - 765

78. Teicher, H., "An inequality on Poisson probabilities", A.M.S. (1955) pp. 147 - 149

79. Thomas, M., "A generalization of Poisson's binomial limit for use in ecology", Biometrika (1949) pp. 18 - 25

80. Tocher, K.D., "Extension of the Neyman - Pearson theory of tests to the discontinuous case", Biometrics (4950) pp. 130 - 144

81. Tukey, J.W., "Sufficiency, truncation, selection", A.M.S. (1949) pp. 309 - 311

82. Wald, A., "Sequential Analysis", John Wiley, New York, 1947

83. Weiss, I., "Limiting distributions in some occupancy problems", Technical Report No. 28, Office of Naval Research, Stanford University, 1955

84. Whitaker, L., table of the general term of Poisson's exponential expansion, Biometrika (1918) p. 36

85. Wilks, S., "Mathematical Statistics", John Wiley, New York, 1962

Note : Abbreviations

A.M.S. : Annals of Mathematical Statistics

J.A.S.A. : Journal of the American Statistical Association