THE POISSON DISTRIBUTION

by

Jeremy Collis-Bird, B.Sc., A.S.A.

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science.

Department of Mathematics,
McGill University,
Montreal.

August 1963.

PREFACE

It was intended to present the subject matter of this thesis in such a manner that a final year undergraduate student of mathematical statistics might read the following pages without difficulty. With this consideration in mind, a rather long introduction to chapter three and an introduction to chapter four have been included. Further, some points that have received emphasis or considerable explanation may well be perfectly straightforward to the mature Statistician.

J. C-B.

August, 1963

TABLE OF CONTENTS

TABLE OF CONTENTS, continued

. . .

HISTORICAL NOTE

The reader of modern statistical literature will, in many areas of science, frequently come across a random variable that is said to have a "Poisson distribution"; or the random variable may be described as one that obeys the "Poisson Law".

Before commencing a study of the properties and structure of the Poisson distribution, it will be interesting and instructive to examine some of the early literature on the subject. By the year 1914 many writers had independently discovered the probability distribution. However, the name by which it is now known is, properly, that of the initial discoverer.

The French mathematician Simeon D. Poisson published a large treatise "Recherches sur la probabilité des jugements..." in 1837. The principal objective of this work was to establish a theorem, which he called the "Law of Great Numbers". This is an extension of the famous theorem of Jacques Bernouilli which appeared in "Ars Conjectandi" in 1713. Consequently, another result obtained in "Recherches" was overlooked for more than half a century, for any interest the mathematicians of the day may have had in this work was focused on the "Law of Great Numbers".

Poisson's derivation of a limiting approximation to the binomial under certain conditions is given below. While the main attempt has been to preserve his argument, some modifications seemed desirable. For not only is the material widely scattered in "Recherches", but the development is too long to reproduce conveniently.

Let p be the probability of a success, q the probability of a failure in a series of m + n Bernouilli trials; p + q = 1.

Let P be the probability of occurrence of at least m failures. Then P will be the probability that there will not occur more than n successes, and is given by the sum of the first n + 1 terms in the expansion of $(q + p)^{m+n}$

i.e. $$P = q^{m+n} + (m+n)q^{m+n-1}\cdot p + \cdots\cdots + \frac{(m+n)\cdots(m+1)}{n!}q^m p^n \qquad (1)$$

After a lengthy combinatorial argument, Poisson arrives at an equivalent expression for P.

$$P = q^m\left[1 + mp + \frac{m(m+1)}{2!}p^2 + \cdots\cdots + \frac{m(m+1)\cdots(m+n-1)}{n!}p^n\right] \qquad (2)$$

The two expressions can be shown to be equivalent by factoring out q^m from (1), replacing q by 1 - p in the remainder and collecting coefficients of powers of p.

Now suppose p to be a very small fraction. If n be the number of successes and m the number of failures then in a very great number of trials $\frac{n}{m+n}$ will also be a very small fraction.

Setting $p(m + n) = \lambda$, $p = \frac{\lambda}{m+n}$

so that $q^m = (1 - \frac{\lambda}{m+n})^m = \left(1 - \frac{\lambda}{m+n}\right)^{m+n} \cdot \left(1 - \frac{\lambda}{m+n}\right)^{-n}$

The first factor in the value for q^m can be replaced by the exponential $e^{-\lambda}$ and the second by unity. So that, the expression (2) can be written

$$e^{-\lambda}\left[1 + \frac{m+n-n}{m+n}\cdot\lambda + \frac{(m+n-n)(m+n+1-n)}{2!\,(m+n)^2}\cdot\lambda^2 + \cdots\cdots + \frac{(m+n-n)\cdots(m+n-1)}{n!\,(m+n)^n}\cdot\lambda^n\right]$$

and omitting the fraction $\frac{n}{m+n}$

P reduces to $e^{-\lambda}[1 + \lambda + \frac{\lambda^2}{2!} + \cdots\cdots + \frac{\lambda^n}{n!}]$

or one can write P in the form

$$P = 1 - \frac{e^{-\lambda}\cdot\lambda^{n+1}}{(n+1)!}\left[1 + \frac{\lambda}{n+2} + \frac{\lambda^2}{(n+2)(n+3)} + \cdots\cdots \text{etc.}\right]$$

Consider the case $\lambda = 1$ and suppose n = 10. Then the difference 1 - P will be approximately one hundred millionth. Thus, where the probability of a single success is the very small fraction $\frac{1}{m+n}$, the probability that there will not occur more than ten successes in the m + n trials becomes almost certainty.

The occurrence of comparatively rare events was first studied by Quetelet and Bortkiewicz. Adolphe Quetelet, a professor of mathematics at the age of nineteen, wrote a number of papers on the application of probability to demographic statistics. In particular, he studied the rates of crime and suicide in different countries and social classes. In 1898, Ladislaus von Bortkiewicz published a small treatise on the occurrence of rare events. He took several examples of rare phenomena from the statistical records of Germany, and using the Poisson probability function to calculate the expected frequencies, he exhibited a close fit with the actual recorded data. His results led him to formulate the "Law of Small Numbers" by which the Poisson probability function provided the theoretical explanation for the observed frequency of occurrence of rare events.

One example Bortkiewicz gave deserves particular mention from the historical viewpoint. This is the data on the number of deaths from the kick of a horse in 14 Army Corps for 20 years, which can nowadays be found in most textbooks on elementary probability. It might be mentioned that the original data for 14 Corps involved 280 deaths, and that Bortkiewicz removed 4 Corps or 80 deaths from his data to achieve the best fit of theory with experience. The first tabulation of the function $e^{-\lambda}\frac{\lambda^x}{x!}$ can be found in this 1898 publication. The tables are given to 4 decimal places for suitably chosen x, and λ from 0.1 to 10.0.

Another treatment of rare events, somewhat similar to that of Bortkiewicz, was given by Mortara (1912). He used data on the marriages of uncle and niece in Italian provinces from 1900 - 1909, and other illustrations.

It appears that Poisson's binomial limit escaped the attention of English speaking mathematicians until after the turn of the century. A textbook on Statistics by A.L. Bowley (1901) contains a numerical example of the binomial expansion with p very small; and in a footnote Bowley remarks that the work of Bortkiewicz has just come to his attention, since preparation of the text for publication.

In 1907 "Student" published a paper in which he arrived at the same limiting expression of the binomial as had Poisson. Student had been engaged in bacteriological research. Briefly, some of his experiments consisted of counting the number of yeast cells in a sample drop of liquid, that settled on each of the 400 squares of a haemacytometer. His theoretical consideration of these experiments was as follows.

Assume the liquid to be thoroughly mixed and spread out in a thin layer over N units of area. Let there be on an average λ cells per unit area, that is $N\lambda$ altogether. A given cell will have an equal chance of falling on any unit area, that is, the chance of its falling in a given unit area is $\frac{1}{N}$ and of its not doing so is $1 - \frac{1}{N}$.

Consequently, considering all the $N\lambda$ cells the chances of 0, 1, 2, 3... cells falling on a given area are given by the terms of the binomial expansion $\left[\left(1-\frac{1}{N}\right) + \frac{1}{N}\right]^{N\lambda}$

The $(r + 1)^{th}$ term is

$$\left(1-\frac{1}{N}\right)^{N\lambda - r}\left(\frac{1}{N}\right)^{r}\frac{N\lambda(N\lambda-1)(N\lambda-2)\cdots\cdots(N\lambda-r+1)}{r!}$$

$$= \left(1-\frac{1}{N}\right)^{N\lambda-r}\cdot\frac{\lambda\left(\lambda-\frac{1}{N}\right)\left(\lambda-\frac{2}{N}\right)\cdots\cdots\left(\lambda-\frac{r-1}{N}\right)}{r!}$$

$$= \left[1 - \frac{N\lambda - r}{N} + \frac{(N\lambda-r)(N\lambda-r-1)}{2!\,N^2} - \cdots\cdots + (-1)^s\frac{(N\lambda-r)\cdots(N\lambda-r-s+1)}{s!\,N^s} + \cdots\right]$$

$$\times\ \frac{\lambda\left(\lambda-\frac{1}{N}\right)\left(\lambda-\frac{2}{N}\right)\cdots\cdots\left(\lambda-\frac{r-1}{N}\right)}{r!}$$

But in the limit as N tends to infinity the terms $\frac{1}{N}, \frac{2}{N}, \cdots \frac{r-1}{N}$ and $\frac{r}{N}, \frac{r+1}{N}, \ldots \frac{r+s-1}{N}$ are all negligibly small. Thus, the $(r + 1)^{th}$ term reduces to

$$\left[1 - \lambda + \frac{\lambda^2}{2!} + \cdots\cdots + (-1)^s\frac{\lambda^s}{s!} + \cdots\cdots\right]\frac{\lambda^r}{r!} = \frac{e^{-\lambda}\lambda^r}{r!}$$

and the binomial expansion to $e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots\cdots + \frac{\lambda^r}{r!} + \cdots\cdots\right]$. Student refers to this result, simply, as the exponential expansion.

Another original derivation of the Poisson distribution was presented by H. Bateman (1910) in a supplementary note to a paper by Rutherford and Geiger.

The two physicists were conducting experiments on the emission of α-particles from a film of polonium. Bateman deduced that the probability that x α-particles strike the screen in a given time interval t is $e^{-\lambda t}\frac{(\lambda t)^x}{x!}$; where λt is the true average number striking the screen in the interval t. The experiments showed a close agreement between theory and observation, and led the physicists to conclude that the α-particles are emitted at random.

The particular case of Bateman's result where x = 0 had been obtained some twenty-five years earlier by Rev. W.A. Whitworth (1886). In his book we find Proposition 51: "If an event happens at random on an average once in time t, the chance of its not happening in a given period τ is $e^{-\frac{\tau}{t}}$". Whitworth's Proposition met with rather rough treatment at the hands of a reviewer in the October issue of The Academy, 1886. The above result and others dealing with random occurrences in time and space are investigated in detail by G. Morant (1921) and tested on the basis of experimental data.

A.G. McKendrick, a Major in the Indian Medical Service, was concerned with the mathematical theory for the phenomenon of phagocytosis. This is the collision and subsequent inclusion of micro-organisms into the substance of the white blood cells. His theoretical consideration of this phenomenon led him directly to the Poisson series (1914).

However, McKendrick was apparently quite unaware of the material that had been published previously on the Poisson distribution.

In 1914 H.E. Soper published tables of the function $e^{-\lambda} \cdot \frac{\lambda^x}{x!}$ for $\lambda = 0.1(0.1)15.0$ and for $x = 0, 1, 2, \ldots$ up to such an integer that gave a figure in the sixth decimal place, the number of places tabulated.

In the same year Lucy Whitaker (1914) published a critical appraisal of the material that had been presented thus far on the Poisson distribution. She examined each instance that had been put forward by Bortkiewicz and Mortara as indicating the operation of the "Law of Small Numbers", and also examined the experimental data in Student's paper.

Whitaker showed that the application of Poisson's result to various data presented by Bortkiewicz and Mortara could not really be justified. For the conditions under which the result is obtained (namely p, the probability that an event will occur in a single trial is very small, positive, and the number of trials very large, positive) are seldom, if at all, demonstrated by the various data. Further, the essential requirement that p remain constant throughout the series of trials is not satisfied. This is indicated after a moment's reflection on the case cited above, for example, of the number of Army Corps deaths over a period of twenty years.

Thus the whole theory of the "Law of Small Numbers" as presented by Bortkiewicz and Mortara appears somewhat obscure. Sight is lost of the fact that the "Law" is nothing more than a limit under certain conditions for binomial probabilities. However, the importance of these early works is that they served to draw attention to the widespread practical application of the Poisson distribution. A subsequent paper by Bortkiewicz (1918) provided a far more detailed mathematical basis for his theory; but to quote Professor Keynes (1921) "...the mathematical argument is right enough, and often brilliant. But what it is all really about, what it all really amounts to, and what the premises are, it becomes increasingly perplexing to decide."

Upon examination of the data on the haemacytometer counts of yeast cells, Whitaker observed that, in some instances, Student obtained a closer fit using the terms of a negative binomial expansion. That is to say, the series of observed frequencies could be more closely fitted by a series of the form $[q + (-p)]^{-N}$ than by a series of Poisson probabilities. The question that naturally arose was how to interpret, for the experiment under consideration, a negative probability for the occurrence of an event in a single trial.

Negative binomial expansions had previously been encountered, but very little was known about them. Karl

Pearson (1905) writes, "A binomial series with negative power or with negative p, q is capable .... of perfectly rational interpretation. But in the present state of our knowledge it would be idle to specify any particular interpretation as the correct one."

Under certain hypotheses a negative binomial may be expected, as in the theory presented by Yule (1910) for the proportion of a population dying after the m^{th} exposure to a disease.

It is easily seen how a negative binomial might arise;[19]. In the usual notation, we have for the binomial with mean $\mu_1' = \mu = Np$ and variance $\mu_2 = Npq$ that $q = \frac{\mu_2}{\mu_1'}$. So that $\mu_2 > \mu_1'$ implies that $q > 1$ and $p = 1 - q$ is negative. But since μ_1' is positive, this implies that N be negative.

We have seen that from theoretical considerations the series of observed frequency of cells in the haemacytometer might reasonably be expected to be given by a Poisson series. An explanation for the fact that this is not necessarily the case in practice was given by Student (1919).

Student showed that series in which the variance is greater than the mean arise from the probability p of an event in a single trial not remaining constant from trial to trial, (the data would then be described as being heterogeneous). Also, he showed that if the presence of one cell in a square of the haemacytometer increased the probability that there would be another, then a negative binomial expansion will give closer agreement with the observed data than will a Poisson series.

The description of heterogeneous material has been the subject of much interest and research. Considerations suggested by the experiments with the haemacytometer, and later by the observed distribution of larvae in experimental plots, has led to many important discoveries. In particular the formulation by J. Neyman (1939) of a new class of "contagious" distributions is mentioned.

CHAPTER ONE: STRUCTURE AND PROPERTIES

1.1 Poisson random variable

Discussion of any random variable may begin by first postulating a Probability Space. Briefly, this implies the existence of a basic sample space Z consisting of sample points z, and having a probability measure m. A point z may be thought of as a possible outcome of a trial, experiment or operation performed under a given set of conditions;[105]. m is a real, non-negative function whose domain of definition is Z; $m(z) = 0$ except on a countable set and $\sum_{z \in Z} m(z) = 1$. A subset S of the sample space Z is called an event, and the probability of the event S is $\text{Prob}(S) = \sum_{z \in S} m(z)$.

Let x(z) be a real, single valued function whose domain of definition is Z. Let T be the range of x(z). If T is the set of real numbers $\{x'\}$ it is seen that the function x(z) induces a partition of Z into mutually exclusive events $S_{x'}$ whose union is Z;[5] where $S_{x'} = \{z;\ x(z) = x'\}$

If the set of real numbers $T = \{x'\}$ is a countable set and such that any finite interval contains at most a finite number of the x' ; and if $\text{Prob}(S_{x'}) > 0$ for all $x' \in T$, then x(z) is called a discrete random variable. We note that

$$\sum_{x' \in T} \text{Prob}(S_{x'}) = \sum_{x' \in T} \sum_{z \in S_{x'}} m(z) = \sum_{z \in Z} m(z) = 1 \qquad (101)$$

Now $\text{Prob}(S_{x'})$ is precisely the probability that the random variable $x(z)$ will assume the value $x' \varepsilon T$, and we write $\text{Prob}(S_{x'}) = \text{Prob}[\, x(z) = x' \,]$.

We shall call $\text{Prob}[\, x(z) = x' \,]$ the probability function (p.f.) of the random variable $x(z)$. The p.f. of a Poisson random variable $x(z)$ is defined as follows.

$$\text{Prob}[\, x(z) = x' \,] = p(x';\lambda) = e^{-\lambda} \frac{\lambda^{x'}}{x'!} \tag{102}$$

where λ is any positive, real number; and the range T of $x(z)$ is the set of non-negative integers.

The notation $x(z)$ and x' is cumbersome, and for simplicity we set aside the symbol x to indicate that value assumed by the Poisson random variable under discussion. Thus we re-write (102) as

$$p(x;\lambda) = e^{-\lambda} \frac{\lambda^{x}}{x!};\; x = 0, 1, 2 \ldots \quad ; \lambda > 0 \tag{103}$$

A random variable having this p.f. is said to have the Poisson distribution $P_o(\lambda)$; for $p(x;\lambda)$ describes the distribution of probability mass over the sample (event) space Z.

Since T is the set of non-negative integers (and is thus countable, with any finite interval containing at most a finite number of elements) and $\lambda > 0 \Rightarrow p(x;\lambda) > 0$ for all $x \varepsilon T$, the Poisson random variable is discrete. We note that

$$\sum_{x \varepsilon T} p(x;\lambda) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^{x}}{x!} = e^{-\lambda} . e^{\lambda} = 1 \tag{104}$$

using the fact that the Maclaurin's series expansion of the function e^{λ} is $1 + \lambda + \frac{\lambda^2}{2!} + \ldots$ and that this series converges to the function, actually for $-\infty < \lambda < \infty$;[39].

1.2 Mode of the distribution $P_o(\lambda)$

Let $\lambda = [\lambda] + f$ where $[\lambda]$ is the largest integer contained in λ and $0 \leq f < 1$.

(a) if $[\lambda] = 0$, then $\lambda < 1$ and $\frac{e^{-\lambda}}{x!} \cdot \lambda^x$ evidently decreases monotonically with increasing x. The maximum term is $p(0;\lambda)$ i.e. the mode is zero.

(b) if $[\lambda] > 0$, form the recurrence relation

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

and consider the two cases:-

(i) $f = 0$ (λ integer).

If $x < \lambda - 1$ then $\phi(x) > 1$ and $p(x;\lambda)$ increases monotonically with increasing x to the modal value.

If $x = \lambda - 1$, then $\phi(\lambda-1)$ is unity, $p(\lambda-1;\lambda) = p(\lambda;\lambda)$ and the distribution $P_o(\lambda)$ is evidently bi-modal with modes at $x = \lambda - 1, \lambda$ for,

if $x > \lambda - 1$ then $\phi(x) < 1$ and $p(x;\lambda)$ decreases monotonically with x.

(ii) $f > 0$. We see that $p(x;\lambda)$ increases monotonically with increasing x to the modal value $p([\lambda];\lambda)$, and then decreases monotonically.

Thus, in all cases, the mode of the distribution $P_o(\lambda)$ is found as the largest integer less than or equal to λ (see chart four, page 126).

1.3 A note on the median

A median of any discrete distribution requires special definition. The following observations are due to Lidstone (1942).

Consider the equation

$$p(0;\lambda) + p(1;\lambda) + \ldots. + p(x-1;\lambda) + c.p(x;\lambda) = \frac{1}{2} \quad ; 0 < c < 1$$

which is

$$e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots + \frac{\lambda^{x-1}}{(x-1)!} + c.\frac{\lambda^x}{x!}\right] = \frac{1}{2} \qquad (105)$$

and define the median as the number x - 1 + c. Thus the sum of the first x terms of the sequence $\{p(i;\lambda)\}$, i = 0, 1, 2 ... will represent less than $\frac{1}{2}$ of the total probability mass of $P_0(\lambda)$, while the sum of the first x + 1 terms will represent more than $\frac{1}{2}$ of this mass.

In a letter to Hardy dated 16th January, 1913, Ramanujan enunciated, without proof, the result that if y is integral

$$\frac{e^y}{2} = 1 + y + \frac{y^2}{2!} + \cdots + \frac{y^{y-1}}{(y-1)!} + t.\frac{y^y}{y!}$$

where $t = \frac{1}{3} + \frac{4}{135(y+v)}$ and $\frac{8}{45} < v < \frac{2}{21}$

Watson (1929) showed this proposition to be true, subject to the truth of two unproved but reasonable hypotheses. While Szegö (1928) proved the proposition in the form

$$\frac{e^y}{2} = 1 + y + \frac{y^2}{2!} + \cdots + \frac{y^{y-1}}{(y-1)!} + K.\frac{y^y}{y!} \quad ; \frac{1}{3} < K < \frac{1}{2}$$

From (105) it follows at once that if λ is integral, the median as defined above is $\lambda - 1 + c$; or $\lambda - 1 <$ median $< \lambda$. Lidstone feels that it is reasonable to assume that if λ is not integral the median will still be approximately $\lambda - 1 + c$. Then, with $\lambda = [\lambda] + f$ as in section 1.2, we will have $[\lambda] + f - 1 <$ median $< [\lambda] + f$. If $f < 1 - c$ then $[\lambda] - 1 <$ median $< [\lambda]$, while $f > 1 - c$ then $[\lambda] <$ median $< [\lambda] + 1$. From his concluding statements Lidstone has apparently confirmed these results by calculation from tables of the Poisson distribution.

1.4 The cumulative distribution function (c.d.f.)

Let the probability that the Poisson random variable assumes a value less than or equal to x be denoted by $F(x;\lambda)$.

i.e. $$F(x;\lambda) = \sum_{i=0}^{x} p(i;\lambda) = e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots + \frac{\lambda^x}{x!}\right] \qquad (106)$$

Then we easily obtain

$$\frac{d}{d\lambda}F(x;\lambda) = -e^{-\lambda}\frac{\lambda^x}{x!} = -p(x;\lambda) \qquad (107)$$

Integrating both sides of (107) from $\lambda = 0$ to $\lambda = \lambda$, and noting that at the lower limit $F(x;0) = 1$, we have immediately

$$F(x;\lambda) = 1 - \int_0^{\lambda} e^{-\lambda}\frac{\lambda^x}{x!}\,d\lambda \qquad (108)$$

$$= \int_{\lambda}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}\,d\lambda \equiv \int_{\lambda}^{\infty} e^{-t}\frac{t^x}{x!}\,dt \qquad (109)$$

Equation (109) was given by Szegö (1928) and previously on

page viii of the Introduction to the Tables of the Incomplete Γ-Function (1922) ;[46].
Equation (108) will be of particular significance in section 4.2.2 where this result is discussed further.

Now it is well known that if y is a continuous random variable with known probability density function (p.d.f.), f(y), then the random variable $u = \int_{-\infty}^{y} f(y)\,dy$ has the uniform distribution on the unit interval. A natural question we might ask then, is whether or not we can make a similar statement about $F(x;\lambda)$.

David and Johnson (1950) have discussed the general case of the distribution of a transformed variable, say $v = \sum_{-\infty}^{y} h(y)$, where h(y) is the p.f. of a discrete variable y. They have considered two cases. First, when the p.f. of y is known; and second, when parameters of the p.f. have to be estimated from observed data. The first case had previously been investigated arithmetically by Lancaster (1949). It will suffice to say here that in both cases the transformed variable v does not have the uniform distribution. David and Johnson investigate the departure from uniformity of the distribution of $F(x;\lambda)$, as measured by the moment ratios $\sqrt{\beta_1}$, β_2 (see section 1.9).

We now define a function which is related directly to the c.d.f. Let the probability that the Poisson random variable assumes a value greater than or equal to x be denoted by $P(x;\lambda)$, then

$$P(x;\lambda) = 1 - F(x-1;\lambda) = \sum_{i=x}^{\infty} p(i;\lambda) \tag{110}$$

We will have frequent need of the derivative of $P(x;\lambda)$ w.r.t. λ. However, as $P(x;\lambda)$ represents an infinite sum, a brief discussion on the legality of performing such a differentiation will be appropriate. If a function of the form $f(t) = \sum_{i=0}^{\infty} c_i t^i$ (where t real, and the c_i are real coefficients) converges for all t in (- R, R) for some R > 0 (R may be $+\infty$) we say that f is expanded in a power series about the point t = 0. For convenience we state here a theorem from mathematical analysis.

<u>Theorem</u>: Suppose the series $\sum_{i=0}^{\infty} c_i t^i$ converges for $|t| < R$, and define $f(t) = \sum_{i=0}^{\infty} c_i t^i$, $|t| < R$. Then $\sum_{i=0}^{\infty} c_i t^i$ converges uniformly on $[-R + \varepsilon, R - \varepsilon]$ no matter which $\varepsilon > 0$ is chosen. The function f is continuous and differentiable in (- R, R) and $\frac{d}{dt} f(t) = \sum_{i=1}^{\infty} i c_i t^i$, $|t| < R$;[82].

Recalling the statement at the end of section 1.1 we know that the series $1 + \lambda + \frac{\lambda^2}{2!} + \ldots$ does converge for $-\infty < \lambda < \infty$ Thus, of course, the series $e^{-\lambda}[\frac{\lambda^x}{x!} + \frac{\lambda^{x+1}}{(x+1)!} + \ldots.]$ converges, which is the series in (110). We may now differentiate $P(x;\lambda)$ with clear conscience and obtain, easily,

$$\frac{d}{d\lambda} P(x;\lambda) = p(x-1;\lambda) \tag{111}$$

which will be true for all x = 0, 1, 2 ... if we define $p(-1;\lambda)$

to be identically zero.

The functions $p(x;\lambda)$, $F(x;\lambda)$ and $P(x;\lambda)$ are now extensively tabulated. Two major sets of tables are those of Molina (1942) and Kitagawa (1952). Some corrections to Kitagawa's tables are given by the Sextons (1959). Perhaps the most extensive tables available are those prepared by the General Electric Company (1962); where all three of the above functions are tabulated to eight decimal places for values of λ from .0000001 through 205 in varying increments.

1.5 Expected value and moments

The r^{th} moment about the origin of the distribution $P_o(\lambda)$ is $\mu_r' = \mathcal{E}(X^r) = \sum_{x=0}^{\infty} x^r . p(x;\lambda)$; r = 0, 1, 2 ...(112) and provided this series is absolutely convergent μ_r' is said to exist. Now,

$$\mu_r' = e^{-\lambda}\left[\lambda \frac{1^r}{1!} + \lambda^2 \frac{2^r}{2!} + \cdots + \frac{\lambda^i i^r}{i!} + \cdots \right] \text{ and } \lambda > 0$$

so that applying the ratio test for convergence to the series of positive terms in the brackets we have

$$\frac{v_{i+1}}{v_i} = \frac{\lambda^{i+1}}{(i+1)!}(i+1)^r . \frac{i!}{\lambda^i . i^r} = \frac{\lambda}{i+1}\left(1 + \frac{1}{i}\right)^r$$

and the ratio $\frac{v_{i+1}}{v_i}$ does tend to a limit less than unity as $i \to \infty$ which implies that the series is convergent, or has a value, for all finite values of λ. Equivalently, since each $v_i > 0$, the series is absolutely convergent. Thus all the moments about the origin of the distribution $P_o(\lambda)$ are seen to exist. With r = 1 we obtain the mean μ'_1, or expected value of x. Only the mean will be found in the manner indicated above, for other methods of obtaining moments will be introduced shortly. Now,

$$\mu'_1 = \mathcal{E}(x) = e^{-\lambda}\left[\lambda + \frac{\lambda^2}{2!}2 + \cdots\right] = \lambda e^{-\lambda} e^{\lambda} = \lambda \qquad (113)$$

indicating the significance of the parameter λ of the Poisson distribution.

Higher moments may be found from a simple recurrence relation. We have $\mu'_r = \sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!}$, an absolutely convergent series; and by the theorem of the preceding section we obtain

$$\frac{d}{d\lambda}\mu'_r = \sum_{x=0}^{\infty}\left\{x^{r+1} \cdot e^{-\lambda}\frac{\lambda^{x-1}}{x!} - x^r \cdot e^{-\lambda}\frac{\lambda^x}{x!}\right\}$$

multiplying both sides by λ yields the recurrence relation

$$\mu'_{r+1} = \lambda\mu'_r + \lambda\frac{d}{d\lambda}\mu'_r \qquad (114)$$

The r^{th} central moment of the distribution $P_o(\lambda)$ is defined as

$$\begin{aligned}
\mu_r &= \mathcal{E}[x - \lambda]^r \qquad ; r = 1, 2, 3 \ldots. \\
&= \mathcal{E}\left[x^r + \binom{r}{1}x^{r-1}(-\lambda) + \cdots + \binom{r}{i}x^{r-i}(-\lambda)^i + \cdots + (-\lambda)^r\right] \\
&= \mu'_r - \binom{r}{1}\mu'_{r-1}\cdot\lambda + \cdots + (-1)^i\binom{r}{i}\mu'_{r-i}\lambda^i + \cdots + (-1)^r\lambda^r \\
&= \sum_{i=0}^{r}(-1)^i\binom{r}{i}\mu'_{r-i}\cdot\lambda^i \qquad (115)
\end{aligned}$$

Since all the μ'_r are known to exist, equation (115) implies the existence of the central moments. Then, by differentiating w.r.t. λ both sides of the equation

$$\mu_r = \sum_{x=0}^{\infty} (x - \lambda)^r . e^{-\lambda} \frac{\lambda^x}{x!}$$

we easily derive the recurrence relation for the central moments of the distribution $P_0(\lambda)$

$$\mu_{r+1} = \lambda \left[r \mu_{r-1} + \frac{d}{d\lambda}\mu_r \right] \quad ; r = 1, 2, 3 \ldots \qquad (116)$$

Now the variance of x is μ_2 and will usually be written V(x). From (113) and (115) $\mu_2 = \mu_2' - \lambda^2$ and using (114) with r = 1

$$\mu_2 = V(x) = [\lambda^2 + \lambda] - \lambda^2 = \lambda \qquad (117)$$

1.6 <u>Moment generating function</u> (m.g.f.)

Suppose a_0, a_1, a_2.... is a bounded sequence of real numbers, and that $S(t) = a_0 + a_1 t + a_2 t^2 + \ldots.$ t real, converges in some interval $-R < t < R$. (A comparison with the geometric series shows that S(t) converges at least for $-1 < t < 1$;[24]). Then, in the sense that the coefficient of t^r is a_r, S(t) is called the generating function of the sequence $\{ a_r \}$.

The expected value of the function $u(y) = e^{ty}$ of a random variable y defines the m.g.f. of y (or of the distribution of y) denoted by $M_y(t)$. For the Poisson random variable we have

$$M_x(t) = \mathcal{E}(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} ; \; -R < t < R$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{t})^{x}}{x!}$$ and recognizing the summation as the Maclaurin's series for e^s with $s = \lambda e^t$, we have

$$M_x(t) = e^{\lambda(e^{t}-1)}; \quad -R < t < R \tag{118}$$

Moreover, since $\lambda > 0$ the above Maclaurins series is absolutely convergent, and the expected value represented by $M_x(t)$ exists.

To show that $M_x(t)$ does indeed generate the moments of the distribution $P_{\cdot}(\lambda)$ we write

$$M_x(t) = \sum_{x=0}^{\infty} \left\{ \sum_{r=0}^{\infty} \frac{(tx)^{r}}{r!} \cdot e^{-\lambda} \frac{\lambda^{x}}{x!} \right\}$$

$$= \sum_{r=0}^{\infty} \frac{t^{r}}{r!} \left\{ \sum_{x=0}^{\infty} x^{r} \cdot e^{-\lambda} \frac{\lambda^{x}}{x!} \right\}$$ $\because$ of absolute convergence

$$= \sum_{r=0}^{\infty} \frac{t^{r}}{r!} \mu_r'$$

and each μ_r' is known to exist so that $\{\mu_r'\}$ is a bounded sequence of reals.

Thus when $M_x(t)$ is expressed as a power series in t, $-R < t < R$, the coefficient of $\frac{t^r}{r!}$ is μ_r'. It is easily seen that

$$\left. \frac{d^{r}}{dt^{r}} M_x(t) \right|_{t=0} = \mu_r' \tag{119}$$

the first two derivatives are

$$\frac{d}{dt} M_x(t) = e^{-\lambda} \lambda e^{t} e^{\lambda e^{t}}$$

$$\frac{d^2}{dt^2} M_x(t) = e^{-\lambda} \lambda e^t e^{\lambda e^t} (1 + \lambda e^t)$$

and hence, as given by (113) and (117)

$$\mathcal{E}(x) = \mu_1' = \lambda$$

$$\mu_2' = \lambda(1+\lambda)$$

and $\quad V(x) = \mu_2 = \lambda(1+\lambda) - \lambda^2 = \lambda$

The existence of the m.g.f. of a random variable y has important consequences. Firstly, the distribution of y is uniquely determined by its m.g.f. when it exists. Secondly, if the m.g.f. of y approaches the m.g.f. of another random variable z, then the distribution of y must approach the distribution of z ;[35].

A well known property of m.g.f.'s is that, with b an arbitrary constant

$$M_{x+b}(t) = \mathcal{E}\left(e^{(x+b)t}\right) = e^{bt} \cdot M_x(t) \tag{120}$$

With b = $-\lambda$ it is easily shown that $M_{x-\lambda}(t)$ generates the central moments μ_r of the distribution $P_0(\lambda)$. For the existence of the function $M_x(t)$, $-R < t < R$, implies the existence of $e^{-\lambda t} \cdot M_x(t)$ for $-R < t < R$. We have,

$$M_{x-\lambda}(t) = e^{-\lambda t} \cdot e^{\lambda(e^t - 1)} = e^{\lambda(e^t - t - 1)} \tag{121}$$

and μ_r is obtained by differentiating $M_{x-\lambda}(t)$ r times w.r.t. t and evaluating at t = 0.

1.7 Probability, and factorial moment generating functions

If we replace t by log t * in the m.g.f. of the distribution $P_o(\lambda)$ we obtain the function $\Pi_x(t)$ say, where

$$\Pi_x(t) = M_x(\log t) = \mathcal{E}\left(e^{x \log t}\right) = \mathcal{E}\left(t^x\right) ; -R < t < R \quad (122)$$

Now $$\mathcal{E}\left(t^x\right) = \sum_{x=0}^{\infty} t^x \, e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!}$$

and the series $\sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!}$ converges absolutely, and to the function $e^{\lambda t}$. Thus $\Pi_x(t)$ exists and is equal to $e^{\lambda(t-1)}$

From (122) we may write

$$\Pi_x(t) = \sum_{x=0}^{\infty} t^x . p(x;\lambda) = p(0;\lambda) + t.p(1;\lambda) + t^2.p(2;\lambda) + \ldots$$

and $\left\{p(i;\lambda)\right\}$, i = 0, 1, 2 ... is a bounded sequence of reals. Thus by differentiating $\Pi_x(t)$ i times w.r.t. t and evaluating at i = 0 we obtain i! $p(i;\lambda)$ i.e. $\Pi_x(t)$ generates the probability function of x.

The r^{th} factorial moment is defined as

$$\mu'_{[r]} = \mathcal{E}\left(x^{[r]}\right) = \mathcal{E}\left(x(x-1)\cdots(x-r+1)\right) \quad ; r = 1, 2, 3 \ldots (123)$$

and if we replace t in $\Pi_x(t)$ by 1 + t we obtain the function $F_x(t)$ say, which generates the factorial moments

$$F_x(t) = \Pi_x(1 + t) = e^{\lambda t} \quad (124)$$

and clearly exists for $-R < t < R$.

---

* We will always mean the natural logarithm.

To show that $F_x(t)$ does generate the factorial moments of the distribution $P_0(\lambda)$ we write

$$F_x(t) = \mathcal{E}(1+t)^x = \sum_{x=0}^{\infty} p(x;\lambda).(1+t)^x$$

$$= \sum_{x=0}^{\infty} p(x;\lambda) \sum_{r=0}^{x} \binom{x}{r} t^r$$

$$= \sum_{x=0}^{\infty} p(x;\lambda) \sum_{r=0}^{x} x^{[r]} \cdot \frac{t^r}{r!}$$

$$= \sum_{r=0}^{x} \frac{t^r}{r!} \sum_{x=0}^{\infty} x^{[r]} \cdot p(x;\lambda)$$

and since $x^{[r]}$ is zero for $r > x$ we write

$$F_x(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \cdot \mu'_{[r]}$$; the factorial moment generating function (f.m.g.f.). The first two factorial moments, obtained from (124) are

$$\mu'_{[1]} = \mathcal{E}(x) = \lambda e^{\lambda t}\Big|_{t=0} = \lambda \quad ; \quad \mu'_{[2]} = \mathcal{E}\big(x(x-1)\big) = \cdots = \lambda^2$$

giving the same moments as before. Obviously $\mu'_{[r]} = \lambda^r$;r=1,2,...

1.8 Cumulant generating function (c.g.f.)

Another infinite set of constants that are useful for characterising a probability distribution are the cumulants (the name suggested by Fisher for the semi-invariants of Thiele).

The cumulants, denoted by k_r ; r = 1, 2, 3 ... are defined by the identity in t

$$e^{\left[k_1 t + k_2 \frac{t^2}{2!} + \cdots + k_r \frac{t^r}{r!} + \cdots\right]} = 1 + \mu_1' t + \cdots + \mu_r' \frac{t^r}{r!} + \cdots \qquad (125)$$

from which any moment μ_r' can be obtained as a polynomial in k_1 , k_2 , ... , k_r and conversely, any k_r is a polynomial in μ_r' and lower order moments. Evidently k_r exists if the moments of order r and lower exist.

Now the right hand side of (125) is the series representation of the m.g.f. If the m.g.f. exists we may take logarithms of both sides of (125) and write

$$k_1 t + k_2 \frac{t^2}{2!} + \cdots + k_r \frac{t^r}{r!} + \cdots = \log M_X(t) = K_X(t) \text{ say}, < \infty$$

Thus, if the logarithm of the m.g.f. can be expanded in a convergent series of powers of t, $-R < t < R$, $K_X(t)$ can be considered as a generating function for the cumulants of the distribution, and

$$k_r = \frac{d^r}{dt^r} K_X(t)\Big|_{t=0} \qquad (126)$$

For the distribution $P_0(\lambda)$

$$K_X(t) = \log M_X(t) = \lambda(e^t - 1) = \lambda \sum_{r=1}^{\infty} \frac{t^r}{r!} \qquad (127)$$

i.e. a convergent power series expansion in t, convergent for $-\infty < t < \infty$. And from (126) it is clear that all the cumulants are equal to λ . However, we may write (127) as

$$\sum_{r=1}^{\infty} k_r \frac{t^r}{r!} = \lambda \sum_{r=1}^{\infty} \frac{t^r}{r!} = \lambda(e^t - 1) \qquad (128)$$

which expresses the fact that we have two power series in t converging to the same function $\lambda(e^{t} - 1)$ in the interval $-\infty < t < \infty$. This implies that the two power series be identical i.e. that they have the same coefficients ;[82] , and hence $k_r = \lambda$; r = 1, 2, 3 ... Thus $K_x(t)$ and hence $M_x(t)$ are uniquely determined, which implies that the distribution $P_0(\lambda)$ is the only probability distribution with such cumulants.

We may define the factorial cumulant generating function (f.c.g.f.), say $J_x(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} k_{[r]}$ as the logarithm of the f.m.g.f. and obtain the factorial cumulants $k_{[r]}$ as polynomials in terms of $\mu'_{[r]}$ and lower order factorial moments, from the identity in t ;[20]

$$J_x(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} k_{[r]} = \log F_x(t) = \log\left(1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} \mu'_{[r]}\right)$$

$$= \frac{t}{1!}\mu'_{[1]} + \frac{t^2}{2!}\left\{\mu'_{[2]} - \mu'^{2}_{[1]}\right\} + \frac{t^3}{3!}\left\{\mu'_{[3]} - 3\mu'_{[2]}\mu'_{[1]} + 2\mu'^{3}_{[1]}\right\}$$

$$+ \frac{t^4}{4!}\left\{\mu'_{[4]} - 4\mu'_{[3]}\mu'_{[1]} + 12\mu'_{[2]}\mu'^{2}_{[1]} - 6\mu'^{4}_{[1]} - 3\mu'^{2}_{[2]}\right\} + \cdots$$

From (124) we have that $\log F_x(t) = \lambda t$ and hence, for the distribution $P_0(\lambda)$ only the first factorial cumulant $k_{[1]}$ is not zero, and is equal to λ.

1.9 Moment ratios

We list below the first few moments μ_r' obtained from (125) in terms of the cumulants; and since for the distribution $P_0(\lambda)$ all the cumulants are equal to λ , we have

$$\mu_1' = k_1 = \lambda$$
$$\mu_2' = k_2 + k_1^2 = \lambda + \lambda^2$$
$$\mu_3' = k_3 + 3k_2k_1 + k_1^3 = \lambda + 3\lambda^2 + \lambda^3$$
$$\mu_4' = k_4 + 4k_3k_1 + 3k_2^2 + 6k_2k_1^2 + k_1^4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4$$

From (121) we may obtain the central moments

$$\mu_1 = 0$$
$$\mu_2 = \lambda$$
$$\mu_3 = \lambda$$
$$\mu_4 = \lambda(1 + 3\lambda)$$

Karl Pearson has defined the moment ratios α_i of a probability distribution by the equations

$$\alpha_i = \frac{\mu_i}{\mu_2^{i/2}} \quad ; \ i = 3, 4, \ldots$$

$\alpha_3 = \dfrac{\mu_3}{\mu_2^{3/2}}$ is often referred to as $\sqrt{\beta_1}$ and is used to describe the departure of the probability distribution from symmetry about the ordinate at the mean i.e. the "skewness".

$\alpha_4 = \dfrac{\mu_4}{\mu_2^2}$ is often referred to as β_2 and is used to describe flatness (or alternatively peakedness). A low value of β_2 is associated with flatness and a high value with peakedness;[20].

For the distribution $P_o(\lambda)$ we have

$$\sqrt{\beta_1} = \frac{\lambda}{\lambda^{3/2}} = \frac{1}{\sqrt{\lambda}} \quad \text{and } \sqrt{\beta_1} \to 0 \quad \text{as } \lambda \to \infty$$

suggesting that the distribution tends to symmetry about the mean with increasing λ (see section 1.15)

$$\beta_2 = \frac{\lambda(1+3\lambda)}{\lambda^2} = 3 + \frac{1}{\lambda} \quad \text{and as } \lambda \to \infty,\ \beta_2 \to 3$$

which is the value of β_2 possessed by the standard Normal distribution.

1.10 A limiting case of the binomial distribution

We have seen that the Poisson distribution was obtained as a limit to the binomial distribution under certain conditions, which are now stated more precisely. The probability that a binomial random variable with distribution $B_i(n;p)$ assumes the value r is $\binom{n}{r} p^r (1-p)^{n-r}$

We require the limiting value of this expression as $n \to \infty$ and $p \searrow 0$ subject to the condition that np remains finite and equal to λ say, > 0. Derivation of this limiting expression can be achieved in a variety of ways. Perhaps the most straightforward approach is to use Stirling's (1730) formula for large factorials;[101]. Now

$$\frac{n!}{(n-r)!\,r!} p^r (1-p)^{n-r} = \left[\frac{\lambda^r}{r!}\left(1-\frac{\lambda}{n}\right)^n\right] \times \frac{n!}{(n-r)!\,n^r\left(1-\frac{\lambda}{n}\right)^r}$$

and $$\lim_{n\to\infty}\left[\frac{\lambda^r}{r!}\left(1-\frac{\lambda}{n}\right)^n\right] = \frac{\lambda^r}{r!}e^{-\lambda}$$

Then

$$\lim_{n\to\infty}\frac{n!}{(n-r)!\,n^r\left(1-\frac{\lambda}{n}\right)^r} = \lim_{n\to\infty}\frac{\sqrt{2\pi n}\,.\,n^n\,.\,e^{-n}}{\sqrt{2\pi(n-r)}\,.\,(n-r)^{n-r}\,.\,e^{-(n-r)}n^r\left(1-\frac{\lambda}{n}\right)^r}$$

$$= \lim_{n\to\infty}\frac{1}{e^r\left(1-\frac{r}{n}\right)^{n-r+\frac{1}{2}}\left(1-\frac{\lambda}{n}\right)^r}$$

$$= \frac{1}{e^r e^{-r}} = 1\,,$$ since r is finite.

Thus $\binom{n}{r}p^r(1-p)^{n-r} \longrightarrow \frac{\lambda^r}{r!}e^{-\lambda}$ under the stated limiting conditions and the terms of the binomial become the successive terms $e^{-\lambda}[\,1\,,\ \frac{\lambda}{1!}\,,\ \ldots .\ \frac{\lambda^r}{r!}\,,\ \ldots .\,]$ i.e. the distribution $P_o(\lambda)$.

Suppose now that we consider, under the same conditions as above, the limit of the m.g.f. $M_y(t) = (1 - p + pe^t)^n$ of the binomial random variable y with distribution $B_i(n;p)$.

Then $$\lim_{n\to\infty}(1 - p + pe^t)^n = \lim_{n\to\infty}\left[1 + \frac{\lambda}{n}(e^t-1)\right]^n$$
$$= e^{\lambda(e^t-1)}$$
$$= M_x(t) \text{ where } x \sim P_o(\lambda).$$

If we now invoke the result stated in section 1.6, this shows that the binomial distribution approaches the Poisson distribution under the given limiting conditions.

As a practical matter it is generally considered justifiable to apply the Poisson approximation to the binomial distribution when $\lambda < 0.1$;[32].

1.11 Reproductivity of Poisson distribution

We may easily establish the following

Result (i) If x_1 and x_2 are distributed independently with $P_o(\lambda_1)$ and $P_o(\lambda_2)$ respectively; then $x_1 + x_2$ is distributed with $P_o(\lambda_1+\lambda_2)$.

Proof: The probability that $x_1 + x_2 = r$ is equal to the joint probability that x_1 assumes the value i and x_2 the value r - i, summed from i = 0 to r.

i.e. $\text{Prob}[\, x_1 + x_2 = r\,] = \sum_{i=0}^{r} \frac{e^{-\lambda_1}\lambda_1^{i}}{i!} \cdot \frac{e^{-\lambda_2}\lambda_2^{r-i}}{(r-i)!}$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^{r} \binom{r}{i} \frac{\lambda_1^{i}\lambda_2^{r-i}}{r!}$$

$$= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^r}{r!} \qquad (129)$$

and the result follows. This property is expressed more simply by saying that the distribution $P_o(\lambda)$ is reproductive w.r.t. λ ;[105].

The converse of this result, that if the sum of two independent random variables has the Poisson distribution then each summand is Poisson distributed, has been proved by Raikov (1937).

Reproductivity can also be shown by using the m.g.f. Thus with x_1 and x_2 as above we obtain immediately

$M_{x_1+x_2}(t) = M_{x_1}(t).M_{x_2}(t)$ because of independence

$= e^{-(\lambda_1+\lambda_2)\{e^{t}-1\}}$ and the result follows.

The condition that $M_{X_1}(t;\lambda_1).M_{X_2}(t;\lambda_2) = M_{X_1+X_2}(t;\lambda_1+\lambda_2)$ can be regarded as characterising the property of reproductivity.

We observe that either of the above two methods can be used to extend the result to the sum of N independent Poisson random variables. Specifically, we use the m.g.f. to establish the following

<u>Result</u> (ii) If $x_1, x_2, \ldots x_N$ are distributed independently with $P_o(\lambda_1)$, $P_o(\lambda_2)$ .... $P_o(\lambda_N)$ respectively; then the random variable $z = \sum_{i=1}^{N} x_i$ has the distribution $P_o(\lambda_1+\lambda_2+ \ldots +\lambda_N)$.

Proof:

$$M_{\sum_{i=1}^{N} x_i}(t) = \prod_{i=1}^{N} M_{x_i}(t) = \prod_{i=1}^{N} e^{\lambda_i(e^t-1)}$$
$$= e^{(\lambda_1+\cdots+\lambda_N)(e^t-1)}$$

and the result follows.

1.12 <u>Difference of two Poisson variables</u>

Now the distribution of the sum of two independent Poisson variables was easily obtained in the previous section and the question that naturally arises is what can be said about the distribution of their difference. This question has been investigated by Irwin (1937) and Johnson (1959).

Suppose two random variables are distributed independently and both with the distribution $P_o(\lambda)$. The probability that the first variable assumes the value x_1 and the second x_2 is $e^{-2\lambda}\dfrac{\lambda^{x_1+x_2}}{x_1!\,x_2!}$; $x_1 = 0, 1, 2, \ldots$; $x_2 = 0, 1, 2 \ldots$

Let $x_2 = x_1 + s$. Then the joint p.f. of x_1 and s is

$f(x_1, s ; \lambda) = e^{-2\lambda}\dfrac{\lambda^{2x_1+s}}{x_1!\,(x_1+s)!}$ and the marginal p.f. for s is

$$g(s;\lambda) = \sum_{x_1=0}^{\infty} f(x_1, s; \lambda) = e^{-2\lambda}\, I_s(2\lambda) \qquad (130)$$

where s can take all integral values from $-\infty$ to ∞, and $I_s(2\lambda)$ is Bessel's modified function of the first kind of order s and argument 2λ. Irwin has shown that $g(s;\lambda)$ tends, as $\lambda \to \infty$, to the Normal distribution with mean zero and variance 2λ.

From (130) we see that the m.g.f. of s is

$$M_s(t) = \mathcal{E}(e^{st}) = e^{-2\lambda}\sum_{s=-\infty}^{\infty} e^{st}\left\{\frac{\lambda^s}{s!} + \frac{\lambda^{s+2}}{1!(s+2)!} + \frac{\lambda^{s+4}}{2!(s+4)!} + \cdots\right\}$$

$$= e^{-2\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\left\{e^{it} + \binom{i}{1}e^{(i-2)t} + \binom{i}{2}e^{(i-4)t} + \cdots + e^{-it}\right\}$$

$$= e^{-2\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\left\{e^{t} + e^{-t}\right\}^i$$

$$= e^{-2\lambda}\cdot e^{\lambda(e^t + e^{-t})}$$

Now the c.g.f. may be found as

$$K_s(t) = \log M_s(t) = -2\lambda + \lambda(e^t + e^{-t})$$

$$= 2\lambda\left[\frac{t^2}{2!} + \frac{t^4}{4!} + \ldots.\right]$$

Hence all the even cumulants of the distribution of s are equal to 2λ, and all the odd cumulants are zero.

Johnson has shown a connection between the distribution of s and the non-central χ^2 distribution.

1.13 On a relation between the Poisson and multinomial distributions

A close relation between the Poisson and the binomial, and between the Poisson and multinomial distributions will be shown in the following sequence of results. In this section we will write, to simplify the notation, a p.f. such as $p(x;\lambda)$ in the form $p(x)$.

Result (i) If x is a random variable with distribution $P_o(\lambda)$ and if the conditional random variable $y|x$ has the binomial distribution $B_i(x;p)$, then the unconditional distribution of y is $P_o(\lambda p)$.

Proof: The joint p.f. for x and y, say f(x , y), can be written as the product of the given conditional p.f. and the marginal p.f. of x.

i.e. $$f(x,y) = \binom{x}{y} p^y (1-p)^{x-y} \cdot \frac{e^{-\lambda}\lambda^x}{x!} \quad ; \quad \begin{array}{l} y = 0, 1, \dots x \\ x = 0, 1, 2 \dots. \end{array}$$

Hence the marginal p.f. for y is

$$g(y) = \sum_{x=0}^{\infty} f(x,y) = \frac{e^{-\lambda}}{y!}\left(\frac{p}{1-p}\right)^y \sum_{x=0}^{\infty} \frac{\lambda^x (1-p)^x}{(x-y)!}$$

$$= e^{-\lambda} \cdot \frac{(\lambda p)^y}{y!} \cdot e^{\lambda(1-p)}$$

$$= e^{-\lambda p} \cdot \frac{(\lambda p)^y}{y!} \quad ; \; y = 0, 1, 2 \dots.$$

Result (ii) If x_1 and x_2 are independently distributed random variables with $P_o(\lambda_1)$ and $P_o(\lambda_2)$ respectively, then the conditional distribution of x_1 given that the value of $x_1 + x_2$ is r say, is the binomial distribution $B_i(r; \frac{\lambda_1}{\lambda_1+\lambda_2})$.

Proof: Using (129) we have immediately that

$$p(x_1 \mid x_1 + x_2 = r) = \frac{e^{-\lambda_1}\dfrac{\lambda_1^{x_1}}{x_1!} \cdot e^{-\lambda_2}\dfrac{\lambda_2^{r-x_1}}{(r-x_1)!}}{e^{-(\lambda_1+\lambda_2)} \cdot \dfrac{(\lambda_1+\lambda_2)^r}{r!}}$$

$$= \binom{r}{x_1}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{r-x_1}$$

$$= \binom{r}{x_1}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1}\left[1 - \frac{\lambda_1}{\lambda_1+\lambda_2}\right]^{r-x_1}$$

Moran (1952) has shown that this result characterises the Poisson distribution among all distributions with range the set of non-negative integers.

Result (iii) follows from (ii). If x_1 and x_2 are independently distributed random variables with the same distribution $P_0(\lambda)$, then the conditional random variable $x_1 \mid x_1 + x_2$ has the binomial distribution $B_r(r; \frac{1}{2})$, where $r = x_1 + x_2$

Result (iv) If $x_1, x_2, \ldots x_N$ are independently distributed random variables, each with the distribution $P_0(\lambda)$, then the conditional random variable $x_1, x_2, \ldots x_N \mid x_1 + x_2 + \ldots + x_N$ has the multinomial distribution $M(r; \frac{1}{N}, \frac{1}{N}, \ldots \frac{1}{N})$, where $r = x_1 + x_2 + \ldots x_N$

Proof:

$$p(x_1, x_2, \ldots x_N \mid x_1 + \cdots + x_N) = \frac{e^{-\lambda}\dfrac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda}\dfrac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda}\dfrac{\lambda^{x_N}}{x_N!}}{e^{-N\lambda}\dfrac{(N\lambda)^r}{r!}}$$

$$= \frac{r!}{x_1! \cdots x_N!} \; \frac{\lambda^{x_1 + \cdots + x_N}}{(N\lambda)^{x_1 + \cdots + x_N}}$$

$$= \frac{r!}{x_1! \cdots x_N!}\left(\frac{1}{N}\right)^{x_1} \cdots \left(\frac{1}{N}\right)^{x_N}$$

where it is understood that $x_N = r - x_1 - x_2 \ldots - x_{N-1}$.
Thus the p.f. of the conditional random variable
$x_1, x_2 \ldots x_N \mid x_1 + x_2 + \ldots + x_N$ is given by the general term in the expansion of the multinomial $\left[\frac{1}{N} + \frac{1}{N} + \cdots + \frac{1}{N}\right]^{x_1 + x_2 + \cdots + x_N}$
i.e. the conditional random variable has the distribution
$M(r; \frac{1}{N}, \ldots \frac{1}{N})$.

<u>Result</u> (v) If $x_1, x_2, \ldots x_N$ are independently distributed random variables with distributions $P_0(\lambda_1), P_0(\lambda_2), \ldots P_0(\lambda_N)$ respectively, and $\mu = \lambda_1 + \lambda_2 + \ldots + \lambda_N$, then the conditional random variable $x_1, x_2, \ldots x_N \mid x_1 + x_2 + \ldots + x_N$ has the distribution $M(r; \frac{\lambda_1}{\mu}, \ldots \frac{\lambda_N}{\mu})$, where $r = x + x + \ldots + x_N$

Proof: We replace $N\lambda$ in (iv) by μ to obtain

$$p(x_1, x_2, \ldots x_N \mid x_1 + x_2 + \ldots + x_N) = \frac{r!}{x_1! \cdots x_N!} \left(\frac{\lambda_1}{\mu}\right)^{x_1} \cdots \left(\frac{\lambda_N}{\mu}\right)^{x_N}$$

where it is understood that $x_N = r - x_1 - x_2 - \ldots - x_{N-1}$

$$\text{and } \frac{\lambda_N}{\mu} = 1 - \frac{\lambda_1}{\mu} - \ldots - \frac{\lambda_{N-1}}{\mu}$$

<u>Result</u> (vi) Any multinomial distribution $\frac{r!}{x_1! \cdots x_N!} p_1^{x_1} p_2^{x_2} \ldots p_N^{x_N}$ can be written as the p.f. of the conditional random variable $x_1, x_2, \ldots x_N \mid x_1 + x_2 + \ldots + x_N$, where $x_1, x_2, \ldots x_N$ are independently distributed Poisson variables with means rp_i, $i = 1, 2, \ldots N$; subject to the condition that the sum of the variables is equal to r ;[78].

Proof: We need only note that in (v) the mean μ of the distribution of the sum $x_1 + x_2 + \ldots + x_N$ becomes $rp_1 + rp_2 + \ldots + rp_N = r$, and hence $\frac{\lambda_i}{\mu} = \frac{\lambda_i}{r} = p_i$, $i = 1, 2, \ldots N$

1.14 On a relation between the Poisson and negative binomial distributions

Result (i) Suppose u is a continuous random variable such that for some positive integer k, ku has the gamma distribution G(k). Suppose $y|u$ is a discrete conditional random variable which has the distribution $P_o(u)$. Then the unconditional distribution of y is a negative binomial distribution; [105].

Proof: The p.d.f. of the random variable t = ku is f(t) say, where $f(t) = e^{-t} \dfrac{t^{k-1}}{\Gamma(k)}$; $0 < t < \infty$

Hence the p.d.f. of u is g(u) say, where ,

$$g(u) = e^{-ku} . u^{k-1} \frac{k^k}{\Gamma(k)} \qquad ; \; 0 < u < \infty$$

Now the joint p.d.f. for y and u is $p(y|u).g(u)$ and we obtain the marginal distribution of y, say h(y), by integrating out u from this joint p.d.f.

i.e. $$h(y) = \int_{u=0}^{\infty} e^{-u} \frac{u^y}{y!} e^{-ku} \frac{u^{k-1}}{\Gamma(k)} k^k \, du$$

put s = (1 + k)u

$$h(y) = \int_{s=0}^{\infty} \frac{e^{-s} k^k \left(\frac{s}{1+k}\right)^{y+k-1}}{\Gamma(k)\, y!\, (1+k)} ds$$

$$= \frac{\Gamma(y+k) . k^k}{\Gamma(k)\, y!\, (1+k)^{y+k}} \qquad ; \; y = 0, 1, 2 \ldots$$

$$= \binom{y+k-1}{k-1} \left(\frac{k}{1+k}\right)^k \left(\frac{1}{1+k}\right)^y$$

now put $p = \frac{k}{1+k}$ and we have

$$h(y) = \binom{y+k-1}{k-1} p^k (1-p)^y \quad ; y = o, 1, 2 \ldots \qquad (131)$$

Or, we may put z = y + k to obtain

$$h(z) = \binom{z-1}{k-1} p^k (1-p)^{z-k} \quad ; z = k, k+1, \ldots.$$

and observe that h(z) is the $[z - k + 1]^{th}$ term in the expression $p.^k[1 - (1 - p)]^{-k}$ when $[1 - (1 - p)]^{-k}$ is expanded in a power series in (1 - p).

Result (ii) If y has the negative binomial distribution given in (131), then the limiting distribution of y as $k \to \infty$ and $(1-p) \to 0$ so that $k(1-p) \to \lambda$ say, finite and > 0, is $P_o(\lambda)$.

Proof: From (131) we have

$$h(y) = \binom{y+k-1}{k-1}\left[1 - \frac{\lambda}{k}\right]^k \left(\frac{\lambda}{k}\right)^y = \left[\frac{\lambda^y}{y!}\left(1 - \frac{\lambda}{k}\right)^k\right] \cdot \frac{(y+k-1)!}{(k-1)!\, k^y}$$

and $$\lim_{k \to \infty} \left[\frac{\lambda^y}{y!}\left(1 - \frac{\lambda}{k}\right)^k\right] = e^{-\lambda} \cdot \frac{\lambda^y}{y!}$$

The remaining terms in h(y) are

$$(y+k-1)(y+k-2)\ldots.(y+k-\overline{y-1}).k \;/\; k^y$$

$$= (1+\tfrac{y-1}{k})(1+\tfrac{y-2}{k})\ldots.(1+\tfrac{1}{k}).1$$

and the result follows immediately.

1.15 Standardised Poisson variable

If y is any random variable with mean μ and variance σ^2 then the new variable $u = \frac{y-\mu}{\sigma}$ is called the standardised random variable corresponding to y, and the distribution of u is said to be in its standard form. Clearly $\mathcal{E}(u) = 0$, and therefore $V(u) = \frac{1}{\sigma^2} \cdot \mathcal{E}(y-\mu)^2 = \frac{\sigma^2}{\sigma^2} = 1$

We observe that the m.g.f. (if it exists) of any standardised variable u will be

$$M_u(t) = \mathcal{E}\left(e^{\frac{y-\mu}{\sigma}\cdot t}\right) = e^{-\mu t/\sigma} . M_y\left(\frac{t}{\sigma}\right) \qquad (132)$$

The standardised variable u corresponding to the Poisson variable x with distribution $P_o(\lambda)$ will be given by $u = \frac{x-\lambda}{\sqrt{\lambda}}$. From (118) and (132) the m.g.f. of u is given by

$$M_u(t) = e^{-\frac{\lambda t}{\sqrt{\lambda}}} . M_x\left(\frac{t}{\sqrt{\lambda}}\right) = e^{-\sqrt{\lambda}\, t} . e^{\lambda\left(e^{\frac{t}{\sqrt{\lambda}}} - 1\right)}$$

$$= e^{\lambda\left[e^{\frac{t}{\sqrt{\lambda}}} - 1 - \frac{t}{\sqrt{\lambda}}\right]} \qquad (133)$$

<u>Theorem</u>: The distribution of the standardised sum of N independently distributed Poisson variables approaches the standard Normal distribution as $N \to \infty$ provided that the sum of the means of the N variables tends to infinity with N ;[19].

Proof: Let $x_1, x_2, \ldots x_N$ be independent random variables with distributions $P_o(\lambda_1)$, $P_o(\lambda_2) \ldots . P_o(\lambda_N)$ respectively, and

$\mu = \lambda_1 + \lambda_2 + \cdots + \lambda_N$

Now the random variable $X = x_1 + x_2 + \ldots + x_N$ has the distribution $P_o(\mu)$, by Result (ii) section 1.11 It follows that $U = \frac{X - \mu}{\sqrt{\mu}}$ is the standardised variable corresponding to the Poisson variable X, and from (133) we have

$$M_U(t) = e^{\mu\left[e^{\frac{t}{\sqrt{\mu}}} - 1 - \frac{t}{\sqrt{\mu}}\right]} \tag{134}$$

We now consider the following two cases.

Case (i): As $N \to \infty$ it may happen that μ tends to a limit b, $b < \infty$.

Then $$\lim_{N\to\infty} M_U(t) = e^{b\left[e^{\frac{t}{\sqrt{b}}} - 1 - \frac{t}{\sqrt{b}}\right]} \tag{135}$$

and (135) is recognised as being the m.g.f. of the standardised variable corresponding to a Poisson variable with mean b. Hence the theorem cannot be true in this case.

Case (ii): As $N \to \infty$ we have that $\mu \to \infty$. Now from (134),

$$\begin{aligned}
\log M_U(t) &= \mu\left[e^{\frac{t}{\sqrt{\mu}}} - 1 - \frac{t}{\sqrt{\mu}}\right] \\
&= \mu\left[1 + \frac{t}{\sqrt{\mu}} + \sum_{i=2}^{\infty}\left(\frac{t}{\sqrt{\mu}}\right)^i \frac{1}{i!} - \left(1 + \frac{t}{\sqrt{\mu}}\right)\right] \\
&= \mu\left[\frac{t^2}{2\mu} + \left\{\frac{t^3}{\mu^{3/2}\,3!} + \frac{t^4}{\mu^2\,4!} + \frac{t^5}{\mu^{5/2}\,5!} + \cdots\right\}\right] \\
&= \frac{t^2}{2} + \frac{t^3}{3!\sqrt{\mu}}\left\{1 + \left(\frac{t}{\sqrt{\mu}}\right)\frac{1}{4} + \left(\frac{t}{\sqrt{\mu}}\right)^2 \frac{1}{4.5} + \cdots\right\} \\
&= \frac{t^2}{2} + \frac{t^3}{3!\sqrt{\mu}}\, e^{\delta\frac{t}{\sqrt{\mu}}} \qquad ; \; 0 < \delta < 1
\end{aligned}$$

Then $$\lim_{N\to\infty} \log M_U(t) = \frac{t^2}{2}$$

and using the fact that the limit of a logarithm equals the logarithm of the limit, provided that these limits exist;[28], we have $\lim_{N\to\infty} M_U(t) = e^{t^2/2}$ which is the m.g.f. of the standard Normal distribution, and the theorem follows.

1.16 Poisson distribution as a special case of a class of discrete distributions

The generalized power series distribution (gpsd) has been defined as follows (Patil, 1959).

Let T be a subset of the set of non-negative integers. Define $f(\theta) = \sum_{y \varepsilon T} a(y)\, \theta^y$, where $a(y) > 0,\ 0 \leqslant \theta < \infty$ (136)

such that $f(\theta)$ is finite and differentiable.

Then a random variable with p.f. $p(y;\theta)$ where

$$p(y;\theta) = \frac{a(y)\ \theta^y}{f(\theta)} \quad ; \ y \ \varepsilon \ T \tag{137}$$

is said to have the gpsd with range T and generating function $f(\theta)$.

The Poisson distribution can be obtained as a special case of the gpsd by taking $f(\theta) = e^{\theta}$ (which is finite and differentiable, $0 \leqslant \theta < \infty$). Then

$$e^{\theta} = \sum_{y=0}^{\infty} \frac{\theta^y}{y!} = \sum_{y \varepsilon T} a(y)\ \theta^y$$

and we have two power series converging to the same function; so that $a(y) = \frac{1}{y!}$ and T is the set of non-negative integers.

Thus $p(y;\theta) = e^{-\theta} \frac{\theta^y}{y!}$; $y = 0, 1, 2, \ldots$

i.e. the random variable y has the distribution $P_o(\theta)$. Evidently the (complete) Poisson distribution is also a special case of the power series distribution (psd) as defined by Noack (1950). We shall not be concerned with a development of the structural properties of the gpsd. But we remark that the structural properties of the Poisson distribution could be derived as special cases of the corresponding properties of the gpsd. However, knowledge of the mean and variance, and of a moment recurrence relation for the gpsd will be found most useful in later work. These quantities are therefore derived below. Let y have the gpsd given in (137).

Then the mean μ of the gpsd is

$$\mathcal{E}(y) = \sum_{y \varepsilon T} y\, p(y;\theta) = \sum_{y \varepsilon T} y\, a(y)\theta^y \;/\; f(\theta)$$

now $$\frac{d}{d\theta} f(\theta) = \sum_{y \varepsilon T} y\, a(y)\theta^{y-1}$$

so that the mean $$\mu = \frac{\theta \frac{d}{d\theta} f(\theta)}{f(\theta)} \tag{138}$$

Consider the quantity

$$\frac{d\mu}{d\theta} = \frac{f(\theta)\left[\frac{d}{d\theta} f(\theta) + \theta \frac{d^2}{d\theta^2} f(\theta)\right] - \theta\left[\frac{d}{d\theta} f(\theta)\right]^2}{[f(\theta)]^2} \tag{139}$$

now $$\frac{d^2}{d\theta^2} f(\theta) = \sum_{y \varepsilon T} y(y-1)\, a(y)\theta^{y-2}$$ and therefore

$$\frac{\theta^2 \frac{d^2}{d\theta^2} f(\theta)}{f(\theta)} = \sum_{y \varepsilon T} y(y-1)\, p(y;\theta) = \mathcal{E}[y(y-1)]$$

Hence we have from (139) that

$$\theta \frac{d\mu}{d\theta} = \frac{\theta \frac{d}{d\theta} f(\theta) + \theta^2 \frac{d^2}{d\theta^2} f(\theta)}{f(\theta)} - \left[\frac{\theta \frac{d}{d\theta} f(\theta)}{f(\theta)}\right]^2$$

$$= \mathcal{E}(y) + \mathcal{E}[y(y-1)] - [\mathcal{E}(y)]^2 = V(y)$$

i.e. the variance of the gpsd is $\theta \frac{d\mu}{d\theta}$ (140)

The r^{th} moment about the origin, μ_r', is

$$\mu_r' = \mathcal{E}(y^r) = \sum_{y \varepsilon T} y^r\, p(y;\theta) = \sum_{y \varepsilon T} y^r \frac{a(y)\theta^y}{f(\theta)}$$

hence we have the equation

$$f(\theta).\mu_r' = \sum_{y \varepsilon T} y^r.a(y)\theta^y \tag{141}$$

and differentiating both sides of (141) w.r.t. θ yields

$$\frac{d}{d\theta} f(\theta).\mu_r' + f(\theta)\frac{d}{d\theta}\mu_r' = \sum_{y \varepsilon T} y^{r+1}.a(y)\theta^{y-1}$$

i.e. $$\frac{\theta\frac{d}{d\theta}f(\theta)}{f(\theta)}.\mu_r' + \theta\frac{d}{d\theta}\mu_r' = \sum_{y \varepsilon T} y^{r+1}.\frac{a(y)\theta^y}{f(\theta)}$$

and hence the recurrence relation between moments

$$\mu_{r+1}' = \theta\frac{d}{d\theta}\mu_r' + \mu.\mu_r' \tag{142}$$

The fact that (138), (140) and (142) include the results obtained in previous sections for the Poisson distribution is easily shown, and will serve to illustrate the remark made above.

For the Poisson distribution we have $f(\theta) = e^\theta$

and hence the mean $\theta\dfrac{\frac{d}{d\theta}f(\theta)}{f(\theta)} = \theta.\dfrac{e^\theta}{e^\theta} = \theta$

and the variance $\theta\dfrac{d\mu}{d\theta} = \theta.1 = \theta$ also.

While (114) is precisely (142), with λ written for θ, since λ is the mean μ.

CHAPTER TWO: THE TRUNCATED AND CENSORED POISSON DISTRIBUTIONS

2.1 Truncation versus censorship

It will be helpful to begin the discussion by considering two different circumstances that commonly present themselves when sampling a Poisson population. At the same time this will allow the statistical terminology to be introduced.

(i) In the first case to be discussed, the population from which the sample is drawn is not complete. For some reason we are only able to sample a part of the population.

(ii) In the second case the sample is drawn from the complete population; but for some reason the individual values of observations above (or below) a given value are not specified.

Cases (i) and (ii) give rise to the truncated and censored Poisson distributions respectively; as so called and discussed by Hald (1952).

As an illustration of case (i) suppose the population to consist of a lot of manufactured items, and suppose the number of defects per item is a Poisson random variable x. After quality inspection of the lot all those items having $x > d$ defects are removed. If the consumer now takes a random sample of the remainder of the lot he is only able to sample the truncated Poisson population (distribution) consisting of those items for which $x \leq d$. Thus the values that the restricted variable can assume are say T^*, where $T^* = \{0, 1, 2 \ldots d\}$. A random observation of the restricted variable is said to have a singly truncated Poisson distribution, truncated on the right at d. Alternatively, the distribution is said to be truncated away from $A = \{d+1, d+2, \ldots\}$.

We may similarly speak of truncation on the left. An example of this instance has been given by Finney and Varley (1955) based upon certain biological data collected by Varley for a study of population balance in the gall-fly.

Other forms of truncation are easily visualised. Thus we may have the restricted variable taking values only in the set $J = \{c, c+1, \ldots d\}$, an instance that is commonly referred to as double truncation. Again, the restricted variable may assume only the values $\{0, 1, \ldots c-1, d+1, d+2, \ldots\}$. i.e. the complementary set J^c.

To consider case (ii) let us recall the experiment that was concerned with counting cells in the haemacytometer. The notation that will be used is given in the table below.

number of cells per square, x	0	1	2	r	d	$>d$	:	Total
number of squares containing x cells	n_0	n_1	n_2....	n_r....	n_d	n_t		N

It is easy to count the number of squares that contain no cells. Counting those that contain only one, two or three is also a simple matter. However, it is not easy to distinguish between squares with high counts; especially if the cells happen to be in Brownian motion!;[94]. As a practical matter then, it is convenient to pool all those squares for which $x > d$, say. The number of squares thus grouped together will be known, though the number of individual cells present in this group will not be known. Thus it will not be possible to compute the mean of the distribution directly;[6].

Assume that x is a Poisson random variable. As a consequence of the experimental procedure observations of the variable have been restricted. For we may observe only that x takes the values 0, 1, 2 .... d and that x is greater than d. A random observation of the variable x in this case is said to have a censored Poisson distribution, singly censored on the right at d.

We may similarly speak of a distribution singly censored on the left and doubly censored, that is, on the left and right with the totals in each tail known separately. Again another type arises when we know the total number of counts in both tails together, but not the totals for each tail separately.

The types of censoring discussed above are generally described as being classical. It is assumed that in repeated sampling the total sample size N is fixed, while the n_t counts in the censored section is an observed variable, i.e. is known but not fixed from sample to sample; [34]. This becomes clear when we recall that the haemacytometer contains $N = 400$ squares, which of course remains fixed.

More general types of censoring than the classical type arise when certain of the observed frequencies are pooled, so that only the pooled totals of these frequencies are known. Or, the number n_t of counts in the censored section may be a fixed quantity and the sample size N a variable. Time will not permit consideration of all the various types of truncation and censorship.

2.2 A special conditional probability

Let the random variable y have the gpsd given by (137). Let T^* be a non-null subset of the range T of the gpsd; and suppose that y be truncated to the set T^*.

What is of interest here is the probability of obtaining an observation y' say, where y' must necessarily be a value in T^*. If the truncated random variable be denoted by y^* we are seeking the p.f. of y^*.

Now $\mathrm{Prob}[\, y^* = y' \,]$, $y' \varepsilon T^*$, is equal to the conditional probability that the random variable y takes the value y' given that y cannot assume a value in the complementary set $T-T^*$.

$$\text{i.e. } \mathrm{Prob}[\, y^* = y' \,] = \frac{p(y';\theta)}{1 - \sum\limits_{y \varepsilon T-T^*} p(y;\theta)} = \frac{p(y';\theta)}{\sum\limits_{y' \varepsilon T^*} p(y';\theta)} \qquad (201)$$

Equation (201) embodies the same notational difficulties encountered in section 1.1. We will drop primes and write the p.f. of a random variable having the truncated gpsd with range T* and parameter θ, as

$$p^*(y;\theta) = \frac{p(y;\theta)}{\sum_{y \varepsilon T^*} p(y;\theta)} \quad ; y \varepsilon T^* \tag{202}$$

where p(y;θ) is the p.f. of the complete gpsd and is given by (137). Thus

$$p^*(y;\theta) = \frac{a(y)\theta^y \;/\; f(\theta)}{\sum_{y \varepsilon T^*} a(y)\theta^y \;/\; f(\theta)} = \frac{a(y)\theta^y}{\sum_{y \varepsilon T^*} a(y)\theta^y} \tag{203}$$

$$= \frac{a(y)\theta^y}{f^*(\theta)} \text{, where } f^*(\theta) = \sum_{y \varepsilon T^*} a(y)\theta^y \tag{204}$$

clearly $p^*(y;\theta) > 0$ and we have $\sum_{y \varepsilon T^*} p^*(y;\theta) = 1.$

We observe that the "truncated" gpsd with p.f. p*(y;θ) given by (204) is in fact a gpsd by definition. Hence the properties that hold for a gpsd continue to hold for its truncated form;[62]. In particular the mean and variance of the truncated gpsd can be found from (138) and (140) respectively; and the recurrence formula (142) is valid in the truncated distribution.

2.3 Poisson distribution singly truncated on the right at d

The Poisson random variable x with distribution $P_o(\lambda)$ is truncated to the set $T^* = \{0, 1, 2, \ldots d\}$.

From (202) we obtain immediately the p.f. of the truncated variable as

$$p^*(x;\lambda) = \frac{e^{-\lambda} \lambda^x}{x!\ F(d;\lambda)} \quad ; \lambda > 0 \quad ; x = 0, 1, 2 \ldots . d \tag{205}$$

where $F(x;\lambda)$ is defined in (106).

From (204) we have that $f^*(\lambda) = \sum_{x=0}^{d} \frac{\lambda^x}{x!}$ and hence the distribution (205) has mean μ^* say, and from (138)

$$\mu^* = \mu^*(d;\lambda) = \frac{\lambda \frac{d}{d\lambda} \sum_{x=0}^{d} \frac{\lambda^x}{x!}}{\sum_{x=0}^{d} \frac{\lambda^x}{x!}} = \frac{\lambda \sum_{x=0}^{d-1} \frac{\lambda^x}{x!}}{\sum_{x=0}^{d} \frac{\lambda^x}{x!}} \tag{206}$$

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

$$= \lambda \frac{F(d-1;\lambda)}{F(d;\lambda)} \tag{207}$$

From (140) the variance, say $V^*(x) = \mu_2^*(d;\lambda) = \lambda \frac{d}{d\lambda} \mu^*$ and from (206) we have

$$V^*(x) = \lambda \frac{\left[\sum_{0}^{d} \frac{\lambda^x}{x!} \left\{ \sum_{0}^{d-1} \frac{\lambda^x}{x!} + \lambda \sum_{0}^{d-1} x \frac{\lambda^{x-1}}{x!} \right\} - \lambda \sum_{0}^{d-1} \frac{\lambda^x}{x!} \sum_{0}^{d} x \frac{\lambda^{x-1}}{x!} \right]}{\left\{ \sum_{0}^{d} \frac{\lambda^x}{x!} \right\}^2}$$

$$= \lambda \frac{F(d-1;\lambda)}{F(d;\lambda)} \left[1 + \lambda \frac{F(d-2;\lambda)}{F(d-1;\lambda)} - \lambda \frac{F(d-1;\lambda)}{F(d;\lambda)} \right] \tag{208}$$

$$= \mu^*(d;\lambda)\ [\ 1 + \mu^*(d-1;\lambda) - \mu^*(d;\lambda)] \tag{209}$$

Thus evaluation of the mean and variance can be achieved by using tables of Poisson sums; or more simply by using tables presented by Patil (1959), who has tabulated the function $\mu^*(d;\lambda)$ for the arguments $d = 4(1)10$ and $\lambda = 0.0(0.1)4.9$

Second and higher moments about the origin can be found using the recurrence relation (142). For example, to find the second moment we have, with $r = 1$

$$\mu_2'^*(d;\lambda) = \lambda \frac{d}{d\lambda}\mu^*(d;\lambda) + [\mu^*(d;\lambda)]^2$$

which from (209) is

$$= \mu^*(d;\lambda)[\ 1 + \mu^*(d-1;\lambda) - \mu^*(d;\lambda)] + [\mu^*(d;\lambda)]^2$$

$$= \mu^*(d;\lambda)[\ 1 + \mu^*(d-1;\lambda)] \tag{210}$$

The third and fourth moments about the origin become

$$\mu_3'^*(d;\lambda) = \mu^*(d;\lambda)[\ 1 + 2\mu^*(d-1;\lambda) + \mu_2'^*(d-1;\lambda)]$$

$$\mu_4'^*(d;\lambda) = \mu^*(d;\lambda)[\ 1 + 3\mu^*(d-1;\lambda) + 3\mu_2'^*(d-1;\lambda) + \mu_3'^*(d-1;\lambda)]$$

2.4 <u>Poisson distribution singly truncated on the left at c</u>

The Poisson random variable x now assumes values only in the set $\{c, c+1, c+2, \ldots.\}$ and (202) yields the p.f. of the truncated variable as, say $p_*(x;\lambda)$ where

$$p_*(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!\ P(c;\lambda)} \quad ; \quad \lambda > 0\ ;\ x = c,\ c+1,\ \ldots. \tag{211}$$

where $P(x;\lambda)$ is defined in (110).

From (204) we have $f_*(\lambda) = \sum_{x=c}^{\infty} \frac{\lambda^x}{x!}$, and proceeding the same manner as in section 2.3 we obtain the mean μ_* say

$$\mu_* = \mu_*(c;\lambda) = \frac{\lambda \frac{d}{d\lambda} \sum_{x=c}^{\infty} \frac{\lambda^x}{x!}}{\sum_{x=c}^{\infty} \frac{\lambda^x}{x!}} = \lambda \frac{\sum_{x=c-1}^{\infty} \frac{\lambda^x}{x!}}{\sum_{x=c}^{\infty} \frac{\lambda^x}{x!}} \tag{212}$$

$$= \lambda \frac{P(c-1;\lambda)}{P(c;\lambda)} \tag{213}$$

and the variance, say $V_*(x) = \mu_{*2}(c;\lambda)$ which becomes

$$V_*(x) = \frac{\lambda \left[\sum_{c}^{\infty} \frac{\lambda^x}{x!} \left\{ \sum_{c-1}^{\infty} \frac{\lambda^x}{x!} + \lambda \sum_{c-1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right\} - \lambda \sum_{c-1}^{\infty} \frac{\lambda^x}{x!} \sum_{c}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right]}{\left\{ \sum_{c}^{\infty} \frac{\lambda^x}{x!} \right\}^2}$$

$$= \lambda \frac{P(c;\lambda)[\,P(c-1;\lambda) + \lambda P(c-2;\lambda)] - \lambda P(c-1;\lambda).P(c-1;\lambda)}{[P(c;\lambda)]^2}$$

$$= \lambda \frac{P(c-1;\lambda)}{P(c;\lambda)} \left[1 + \lambda \frac{P(c-2;\lambda)}{P(c-1;\lambda)} - \lambda \frac{P(c-1;\lambda)}{P(c;\lambda)} \right] \tag{214}$$

and from (213) we have that

$$V_*(x) = \mu_*(c;\lambda)[\,1 + \mu_*(c-1;\lambda) - \mu_*(c;\lambda)] \tag{215}$$

Thus the mean and variance may be computed from (213) and (214) with the aid of tables of the function $P(x;\lambda)$, for example the General Electric Company tables (1962). Or, one can find the mean directly, and the variance from (215) using the tables of Patil (1959), who has tabulated $\mu_*(c;\lambda)$ for the arguments $c = 1(1)10$ and $\lambda = 0.0(0.1)9.9$

Second and higher moments about the origin can be found from the recurrence relation (142). Thus we have

$$\mu'_{*2}(c;\lambda) = \lambda \frac{d}{d\lambda} \mu_*(c;\lambda) + [\mu_*(c;\lambda)]^2$$
$$= \mu_*(c;\lambda)[1 + \mu_*(c-1;\lambda)] \qquad (216)$$

using (140) and (215).

2.4.1. An important special case of the Poisson distribution truncated on the left occurs with c = 1; which is often referred to as the case of "missing zero counts".

From (211) we obtain the p.f.

$$p_*(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!\,[1 - e^{-\lambda}]} \quad ; \quad \lambda > 0 \; ; \; x = 1, 2, 3 \ldots . \qquad (217)$$

By putting c = 1 in (213), (215), and (216) we have the corresponding constants for this special case. However, let us consider the m.g.f. (existence is obvious) of the distribution (217).

$$M_x(t) = \mathcal{E}\left(e^{tx}\right) = \sum_{x=1}^{\infty} \frac{e^{tx}\lambda^x e^{-\lambda}}{x!\,[1 - e^{-\lambda}]}$$

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

$$= \frac{e^{\lambda e^t} - 1}{e^{\lambda} - 1} \qquad (218)$$

$$\mu_*(1;\lambda) = \frac{d}{dt} M_x(t)\Big|_{t=0} = \frac{\lambda e^{\lambda}}{e^{\lambda} - 1} = \frac{\lambda}{1 - e^{-\lambda}} \qquad (219)$$

$$\mu'_{*2}(1;\lambda) = \frac{d^2}{dt^2} M_x(t)\Big|_{t=0} = \frac{\lambda}{e^{\lambda}-1}\cdot[\, e^{\lambda} + \lambda e^{\lambda}\,] = \frac{\lambda(1+\lambda)}{1 - e^{-\lambda}} \qquad (220)$$

Hence the variance $\mu_{*2}(1;\lambda)$ obtained from (219) and (220) is

$$\mu_{*2}(1;\lambda) = \frac{\lambda(\,1 + \lambda\,)}{1 - e^{-\lambda}} - \frac{\lambda^2}{(1 - e^{-\lambda})^2} = \lambda\,\frac{[\,1 - (1+\lambda)e^{-\lambda}\,]}{(\,1 - e^{-\lambda}\,)^2} \qquad (221)$$

and the variance can be written in the alternative form, from (215) with c = 1

$$\mu_{*2}(1;\lambda) = \mu_*(1;\lambda)[\,1 + \lambda - \mu_*(1;\lambda)]$$

which is seen to be equivalent to (221).

2.5 The doubly truncated Poisson distribution

Moore (1954) has observed that the doubly truncated Poisson distribution occurs sometimes in botanical work. If we consider truncation on the left at c and on the right at d, then (202) yields for the doubly truncated random variable the p.f. $p_D(x;\lambda)$ say, where

$$p_D(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!\,[\,P(c;\lambda) - P(d+1;\lambda)]} \quad ; \quad \lambda > 0\ ;\ c \leq x \leq d \qquad (222)$$

and the function $P(x;\lambda)$ is defined in (110).

To obtain the mean μ_D say, of the distribution (222) we may proceed in the same manner as in the two previous sections. Then,

$$\mu_D = \mu_D(c,d;\lambda) = \frac{\lambda \frac{d}{d\lambda} \sum_{x=c}^{d} \frac{\lambda^x}{x!}}{\sum_{x=c}^{d} \frac{\lambda^x}{x!}}$$

$$= \frac{\lambda \sum_{x=c-1}^{d-1} \frac{\lambda^x}{x!}}{\sum_{x=c}^{d} \frac{\lambda^x}{x!}} \tag{223}$$

and multiplying numerator and denominator of (223) by $e^{-\lambda}$

$$\mu_D = \mu_D(c,d;\lambda) = \lambda \cdot \frac{P(c-1;\lambda) - P(d;\lambda)}{P(c;\lambda) - P(d+1;\lambda)} \tag{224}$$

Now the variance, say $V_D(x) = \lambda \frac{d}{d\lambda}\mu_D$, and we may obtain the required derivative from (223); or from (224) using the result given by (111), with about the same amount of work in either case. Using (223) we obtain

$$V_D(x) = \frac{\lambda\left[\sum_{c}^{d} \frac{\lambda^x}{x!}\left\{\sum_{c-1}^{d-1} \frac{\lambda^x}{x!} + \lambda \sum_{c-1}^{d-1} \frac{\lambda^{x-1}}{(x-1)!}\right\} - \lambda \sum_{c-1}^{d-1} \frac{\lambda^x}{x!} \sum_{c}^{d} \frac{\lambda^{x-1}}{(x-1)!}\right]}{\left\{\sum_{c}^{d} \frac{\lambda^x}{x!}\right\}^2}$$

$$= \frac{\lambda \sum_{c-1}^{d-1} \frac{\lambda^x}{x!}}{\sum_{c}^{d} \frac{\lambda^x}{x!}} \left[1 + \lambda \frac{\sum_{c-2}^{d-2} \frac{\lambda^x}{x!}}{\sum_{c-1}^{d-1} \frac{\lambda^x}{x!}} - \lambda \frac{\sum_{c-1}^{d-1} \frac{\lambda^x}{x!}}{\sum_{c}^{d} \frac{\lambda^x}{x!}} \right]$$

$$= \lambda \frac{P(c-1;\lambda) - P(d;\lambda)}{P(c;\lambda) - P(d+1;\lambda)}\left[1 + \lambda \frac{P(c-2;\lambda) - P(d-1;\lambda)}{P(c-1;\lambda) - P(d;\lambda)} - \lambda \cdot \frac{P(c-1;\lambda) - P(d;\lambda)}{P(c;\lambda) - P(d+1;\lambda)}\right]$$

and from (224) we obtain finally

$$V_D(x) = \mu_D(c,d;\lambda)[\,1 + \mu_D(c-1,d-1;\lambda) - \mu_D(c,d;\lambda)] \tag{225}$$

2.6 Censored Poisson distributions

We should recall from section 2.1 that in the censored situation the sample is drawn from a complete Poisson population. After the sample has been drawn, it is the values actually recorded by the statistician that are not specified completely (due to pooling or other reasons). Or, expressed more crudely, all the information is present in the sample; but because of faulty apparatus, poor experimental technique, cost or other reasons, not all of this information is extracted.

For the singly censored case, on the right at d say, the probability that an observation of the censored Poisson random variable has the value x will be

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0,\ 1,\ 2 \ldots .\ d$$

and $P(d+1;\lambda)$ for $x > d$.

Similarly, for the singly censored case on the left at c say, the probability that an observation of the censored variable has the value x will be

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = c,\ c+1,\ \ldots .$$

and $F(c-1;\lambda)$ for $x < c$.

The probabilities appropriate for the doubly censored case are now obvious.

CHAPTER THREE: POINT ESTIMATION

3.1 Introduction

In this chapter we are concerned with the problem of providing a value for the true value of the parameter of a population (Poisson) on the basis of the information obtained by sampling from this population. We begin with a brief outline of the principal methods that have been used for point estimation. References to some of the many important results in estimation theory are given.

3.1.1 The method of moments;- is the oldest general method and was introduced by Karl Pearson. The method consists of equating the sample moments m'_r to the corresponding population moments μ'_r which are functions of the unknown parameters. One considers as many moments as necessary in order to solve the equations for the unknown parameters. Now, for any sample moment m'_r we have $\mathcal{E}(m'_r) = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathcal{E}(x_i^r) = \mu'_r$

By a theorem of Khintchine it follows that, as soon as the population moment μ'_r exists, the sample moment m'_r converges in probability to μ'_r as $N \to \infty$. Thus, in large samples m'_r may be regarded as an "estimate" of μ'_r ;[15].

3.1.2 The method of maximum likelihood:- was used in particular cases by Gauss; but it was R.A. Fisher who introduced and developed it as a general method of point estimation.

If $f(x_1, \ldots x_N;\theta)$ is the joint p.d.f. of a random sample of size N **drawn** from the distribution with p.d.f. $f(x;\theta)$ and unknown parameter $\theta \varepsilon \Omega$, then the maximum likelihood estimate (m.l.e.) of θ is the number $\hat{\theta}(x_1, \ldots x_N) = \hat{\theta}$ which maximises $f(x_1, \ldots x_N;\theta)$.

When the sample values are given the **joint** p.d.f. of the random sample becomes a function of θ, called the likelihood function $L(\theta;x_1, \ldots x_N) = L$ and

$$L(\theta;x_1, \ldots x_N) = f(x_1;\theta)\ldots f(x_N;\theta)$$

The value of θ which maximises L can often be found by calculus techniques. Since log L attains its maximum for the same value of θ as does L, it is often convenient to solve the "likelihood equation"

$$\frac{\partial}{\partial\theta}\log L = 0 \qquad \text{for the m.l.e. } \hat{\theta}$$

<u>Asymptotic variance of $\hat{\theta}$</u>: It was Fisher who first gave the result that under certain regularity conditions on $f(x;\theta)$, the distribution of $\hat{\theta}$ approaches the Normal distribution with mean θ and variance $\frac{1}{N\tau^2}$ as N becomes large, where

$\tau^2 = V[\frac{\partial}{\partial\theta}\log f(x;\theta)]$. Fisher also proved that no other estimate, Normally distributed and unbiased for large N, can have smaller variance than the m.l.e.;[27].

We may obtain an alternative form for $N\tau^2$ as follows (assuming differentiation under the integral sign is permitted, and the range of x is independent of θ).

Since the integral over the range of x is unity we obtain

$$\int \frac{\partial f}{\partial \theta} dx = \int \left(\frac{1}{f} \frac{\partial f}{\partial \theta} \right) f\, dx = \int \left(\frac{\partial}{\partial \theta} \log f \right) f\, dx = 0 \tag{301}$$

$$\text{so that} \quad \mathcal{E}\left(\frac{\partial}{\partial \theta} \log f \right) = 0 \tag{302}$$

$$\text{Hence} \quad \tau^2 = V\left(\frac{\partial}{\partial \theta} \log f \right) = \mathcal{E}\left(\frac{\partial}{\partial \theta} \log f \right)^2 \tag{303}$$

Differentiating again w.r.t. θ the last equation in (301) gives

$$\int \left(\frac{\partial^2}{\partial \theta^2} \log f \right) f\, dx + \int \left(\frac{\partial}{\partial \theta} \log f \right)^2 f\, dx = 0$$

which implies that

$$- \mathcal{E}\left(\frac{\partial^2}{\partial \theta^2} \log f \right) = \mathcal{E}\left(\frac{\partial}{\partial \theta} \log f \right)^2 = \tau^2, \text{ from (303)}$$

Since $\log L = \sum_{i=1}^{N} \log f(x_i;\theta)$, then $\frac{\partial^2}{\partial \theta^2} \log L = \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta^2} \log f$

and $- \mathcal{E}\left(\frac{\partial^2}{\partial \theta^2} \log L \right) = - N.\, \mathcal{E}\left(\frac{\partial^2}{\partial \theta^2} \log f \right) = N \tau^2$

The asymptotic variance of $\hat{\theta}$ is given by

$$\frac{1}{- \mathcal{E}\left(\frac{\partial^2}{\partial \theta^2} \log L \right)} \tag{304}$$

3.1.3 Sufficient statistic: single parameter $\theta \in \Omega$.

In 3.2.5. we shall have need of the following well known theorem. A necessary and sufficient condition for the statistic $\tilde{t}(x_1, \ldots x_N) = \tilde{t}$ to be sufficient for θ is given by the Fisher-Neyman Criterion: The statistic $\tilde{t}$ is sufficient for θ iff the joint p.d.f. of the random sample can be factored as

$$f(x_1;\theta)\ldots f(x_N;\theta) = g(\tilde{t};\theta).H(x_1, \ldots x_N) \tag{305}$$

where $H(x_1, \ldots x_N)$ does not depend on θ;[36].

Sufficiency of the condition was given by Fisher, and necessity by Neyman. A general proof which considers the case where θ is a vector quantity has been given by Halmos and Savage (1949).

An important property possessed by a sufficient statistic $\tilde{t}$ is expressed in the following

Result: If u is a one-to-one function of θ, then $u(\tilde{t})$ is also a sufficient statistic for θ; and $\tilde{t}$ is a sufficient statistic for $u(\theta)$;[50].

3.1.4. Cramer-Rao Inequality. Suppose $(x_1, \ldots x_N)$ is a random sample from a distribution having p.d.f. $f(x;\theta)$, $\theta \in \Omega$, and $t(x_1, \ldots x_N) = t_N$ is any function of the observations. It has been shown by Cramer (1946), and Rao (1945) that, under certain conditions,

$$V(t_N) \geqslant \frac{[\, g'(\theta)\,]^2}{N.\mathcal{E}\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2} \quad \text{where } g(\theta) = \mathcal{E}(t_N)$$

Generally the estimator t_N will have a certain bias $b(\theta)$ depending on θ, and we may write $g(\theta) = \theta + b(\theta)$. If t_N is an unbiased estimator for θ, $b(\theta) = 0$ and the inequality reduces to

$$V(t_N) \geqslant \frac{1}{N.\,\mathcal{E}\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2} \tag{306}$$

and equality holds iff there exists a constant K, which may depend on θ and N, such that

$$\sum_{i=1}^{N} \frac{\partial}{\partial\theta} \log f(x_i;\theta) = K[\ t_N - \theta\ ] \text{ with probability one;[55]} \tag{307}$$

3.1.5. Distributions admitting sufficient statistics.

Suppose $(x_1, \ldots x_N)$ is a random sample from the distribution with p.d.f. $f(x;\theta)$, $\theta \in \Omega$, such that a sufficient statistic $\tilde{t}$ for θ exists. It has been shown that, under certain regularity conditions, the joint p.d.f. $f(x_1, \ldots x_N;\theta)$ must be of the form

$$e^{p(\theta).Q(\tilde{t}) + S(x_1, \ldots x_N) + q(\theta)}$$

where $p(\theta)$ and $q(\theta)$ are functions of θ only; and $S(x_1, \ldots x_N)$, $Q(\tilde{t})$ are functions of $(x_1, \ldots x_N)$ only;[105]. The class of probability distributions thus characterised is referred to as the Koopman-Pitman (K-P) class. If for some member of this class the range of x is independent of θ then this member is referred to as a regular case of the K-P class.

Now it is known that in a regular case of the K-P class with parameter θ, the p.d.f. of a sufficient statistic for θ is complete;[36]. By referring to Result (ii) section 1.11 it is immediately seen that the Poisson distribution $P_o(\lambda)$ is a regular case of the K-P class. In section 3.2.5. a sufficient statistic for λ will be found, and it will be proved that the p.d.f. of the sufficient statistic is complete.

3.1.6. Minimum variance unbiased estimation.

A key result in the theory of estimation is the Rao-Blackwell Theorem: Suppose $\tilde{t}(x_1, \ldots x_N)$ is a sufficient statistic for θ and $t'(x_1, .. x_N) = t'$ is any other unbiased estimator for θ. Let $\mathcal{E}(t'|\tilde{t}) = h(\tilde{t})$. Then $h(\tilde{t})$ is an unbiased estimator for θ whose variance cannot exceed that of t';[105].

It follows from the theorem that when a sufficient statistic exists we need to look only at unbiased estimators that are functions of the sufficient statistic in order to locate that unbiased estimator whose variance is less than the variance of every other unbiased estimator. However, there may well exist many unbiased estimators based on the sufficient statistic.

Another important result is the following. Suppose a sufficient statistic $\tilde{t}$ for θ exists. If the p.d.f. of $\tilde{t}$ is complete, then there is only ONE unbiased estimator for θ which is based on the sufficient statistic;[50]. This unique unbiased estimator for θ will be called the minimum variance unbiased (MVU) estimator i.e. the MVU estimator is that unbiased estimator which is based on a sufficient statistic having a complete p.d.f.;[27]. Some authors refer to the MVU estimator as the uniformly minimum variance unbiased (UMVU) estimator; but the former terminology will be used.

3.2 The complete Poisson distribution

3.2.1. Notation. Let a random sample $(x_1, \ldots x_N)$ of size N be drawn from the distribution $P_0(\lambda)$, which has p.f. $p(x;\lambda)$. The r^{th} sample moment is given by $m'_r = \sum_{i=1}^{N} x_i^r / N$

Now each x_i , $i = 1, 2 \ldots N$, is a non-negative integer, and it may be convenient to tabulate the observations as follows.

possible value for x	0	1	2	h	h+1	...	: Total
observed frequency	n_0	n_1	n_2....	n_h	0	...	N

In this case $m'_r = \sum_{x=0}^{h} \frac{x^r n_x}{N}$ which is also equal to $\sum_{x=0}^{\infty} \frac{x^r n_x}{N}$

Each of the N random observations may be considered as being the result of one of N independent trials which can have one of the outcomes $x = 0, 1, 2 \ldots$, with probability $p(x;\lambda)$.

The components of the infinite dimensional random variable $(n_0, n_1, n_2, \ldots)$ denote the number of trials which resulted in the outcomes o, 1, 2 ... respectively. These components are linearly dependent since their sum is N. The random variable $(n_0, n_1, \ldots)$ has the multinomial distribution with p.f. given by the general term in the multinomial expansion $[p(0;\lambda)+p(1;\lambda)+\ldots+p(h;\lambda)+p(h+1;\lambda)+\ldots]^N$

The marginal distribution of the h+1 dimensional random variable $(n_0, n_1, \ldots n_h)$ will also be multinomial, with p.f. given by the general term in the expansion

$[p(0;\lambda)+p(1;\lambda)+\ldots +p(h;\lambda)+P(h+1;\lambda)]^N$ where $P(h+1;\lambda)$

is defined by (110). This general term is

$$\frac{N!}{n_0!n_1!\ldots n_h!n_H!}\{p(0;\lambda)\}^{n_0}\ldots\{p(h;\lambda)\}^{n_h}.\{P(h+1;\lambda)\}^{n_H}$$

where n_H is the number of trials resulting in outcomes $x > h$, and is zero. Thus we may write the p.f. as

$$p(n_0, n_1, \ldots n_h) = \frac{N!}{n_0!\ldots n_h!}\{p(0;\lambda)\}^{n_0}\ldots\{p(h;\lambda)\}^{n_h} \qquad (309)$$

where $p(h;\lambda) = [\ 1 - P(h+1;\lambda)] - p(0;\lambda) - \ldots - p(h-1;\lambda)$

and $\quad n_h \quad = [\ N - n_H\ ] - n_0 - \ldots\ldots - n_{h-1}$

We may obtain easily the joint factorial moment

$$\mu'_{[r_0]\cdots[r_{h-1}]} = N^{[r_0+\cdots+r_{h-1}]}\{p(0;\lambda)\}^{r_0}\ldots\ldots\{p(h-1;\lambda)\}^{r_{h-1}}$$

and the expectations

$$\left.\begin{aligned}
\mu'_{[0]\cdots[r_x=1]\cdots[0]} &= \mathcal{E}(n_x) &&= N.p(x;\lambda) \\
\mu'_{[0]\cdots[r_x=1]\cdots[r_y=1]\cdots} &= \mathcal{E}(n_x n_y) &&= N(N-1).p(x;\lambda)p(y;\lambda)\ ;x\neq y \\
\text{and also}\qquad & \mathcal{E}(n_x^2) &&= N.p(x;\lambda)+N(N-1)[p(x;\lambda)]^2
\end{aligned}\right\} \qquad (310)$$

3.2.2. An estimate of λ is obtained by using the observed frequencies of two different values of x, say n_b and n_{b+r}, as estimates of their expected values $N.p(b;\lambda)$ and $N.p(b+r;\lambda)$

If this is done we obtain the equation

$$\lambda^r = \frac{(b+r)!\ \ n_{b+r}}{b!\ \ \ n_b} \qquad (311)$$

which can be solved for the estimate of λ .

This result can be obtained as a special case of a general method of estimation called the ratio method by Patil (1961). The method is applicable to the problem of estimating the parameter θ of a gpsd; and the truncated and censored forms provided that their range contains a subset of consecutive integers.

Consider the gpsd (137) with range T finite and $T = \{c, c+1, \ldots\ldots d\}$

Let $g_r(y) = \dfrac{a(y-r)}{a(y)}$, $y \in T$ and r being an integer such that $y-r \in T$.

Then, for arbitrary u and v with $c+r \leq u \leq v \leq d$ we have

$$\sum_{y=u}^{v} g_r(y).p(y;\theta) = \sum_{y=u}^{v} a(y-r)\frac{\theta^y}{f(\theta)} = \theta^r \sum_{y=u-r}^{v-r} a(y)\frac{\theta^y}{f(\theta)}$$

and hence the identity

$$\theta^r = \frac{\sum_{y=u}^{v} g_r(y).p(y;\theta)}{\sum_{y=u-r}^{v-r} p(y;\theta)}$$

If we now take n_y as an estimate of its expected value $N.p(y;\theta)$, the statistic

$$\frac{\sum_{y=u}^{v} g_r(y).n_y}{\sum_{y=u-r}^{v-r} n_y} \tag{312}$$

may be taken as an estimate of θ^r for admissable values of $r = 1, 2 \ldots\ldots$ These estimates are referred to as "ratio estimates"; [62].

Considering the distribution $P_0(\lambda)$ we have $r \leq u \leq v \leq \infty$, and choose $u = v = b+r$.

The statistic (312) becomes

$$g_r(b+r).\frac{n_{b+r}}{n_b} = \frac{a(\overline{b+r} - r).n_{b+r}}{a(b+r) \quad n_b} = \frac{(b+r)! \; n_{b+r}}{b! \quad n_b}$$

as an estimate of λ^r, which is seen to be the result in (311).

If in (312) we take $r = 1$ and choose $u = c+1$, $v = d$ then we obtain the statistic

$$\sum_{y=c+1}^{d} \frac{a(y-1)}{a(y)} n_y \Big/ \sum_{y=c}^{d-1} n_y \tag{313}$$

which we will call THE ratio estimate, R, for the parameter θ.

Patil (1961) has shown that R is not in general an unbiased nor an efficient estimator for θ. However, for the case where d is infinite we have

$$\mathcal{E}(R) = \frac{1}{N} \cdot \sum_{y=c+1}^{\infty} \frac{a(y-1)}{a(y)} \mathcal{E}(n_y)$$

$$= \sum_{y=c+1}^{\infty} \frac{a(y-1)}{a(y)} \frac{a(y)\theta^y}{f(\theta)}$$

and $$\mathcal{E}(R) = \theta. \sum_{y=c}^{\infty} a(y)\theta^y \,/f(\theta) = \theta \tag{314}$$

Now,

$$\mathcal{E}(R^2) = \frac{\mathcal{E}\left[\sum_{c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}.n_y\right\}^2\right] + \mathcal{E}\left[\sum_{x}^{\infty}\sum_{\neq y}^{\infty}\frac{a(x-1)}{a(x)}.n_x \; \frac{a(y-1)}{a(y)}.n_y\right]}{N^2}$$

$$= \frac{N\sum_{c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}\right\}^{2} p(y;\theta) \quad + N(N-1)\sum_{c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}\right\}^{2}\left[p(y;\theta)\right]^{2}}{N^2}$$

$$+ \frac{N(N-1)\sum_{x}^{\infty}\sum_{\neq\, y}^{\infty} \frac{a(x-1)}{a(x)}\ \frac{a(y-1)}{a(y)}\ p(x;\theta).p(y;\theta)}{N^2}$$

$$= \frac{N(N-1)\left[\sum_{c+1}^{\infty}\frac{a(y-1)}{a(y)}\ p(y;\theta)\right]^{2} + N\sum_{c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}\right\}^{2} p(y;\theta)}{N^2} \qquad (315)$$

Therefore the variance of R, V(R) becomes

$$V(R) = \mathcal{E}(R^2) - [\mathcal{E}(R)]^2$$

$$= \left[N(N-1)\theta^2 + N\sum_{c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}\right\}^{2} p(y;\theta) \quad - N^2\theta^2\right] \quad / \quad N^2$$

$$= \frac{1}{N}\left[\sum_{c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}\right\}^{2} p(y;\theta) \quad - \theta^2\right] \qquad (316)$$

Using (315) it is easily shown that the statistic

$$S = \frac{\sum_{y=c+1}^{\infty}\left\{\frac{a(y-1)}{a(y)}\right\}^{2} n_y \quad - N.R^2}{N(N-1)} \qquad (317)$$

is an unbiased estimator for V(R).

Thus, for the distribution $P_o(\lambda)$ we have that the ratio estimate for the parameter λ is

$$R = \frac{\sum_{x=1}^{\infty}\frac{x!}{(x-1)!}n_x}{\sum_{x=0}^{\infty} n_x} = \frac{\sum_{x=0}^{\infty} x\ n_x}{N} = m_1' = \bar{x} \qquad (318)$$

(for the summation is unaffected by the addition of the zero counts); and $\mathcal{E}(R) = \lambda$. From (316)

$$V(R) = \frac{1}{N} \cdot \left[\sum_{x=1}^{\infty} x^2 \, p(x; \ ) - \lambda^2 \right]$$

$$= \frac{1}{N} \left[\mu_2' - \lambda^2 \right] = \frac{\lambda}{N} \tag{319}$$

and from (317) an unbiased estimator for V(R) is

$$S = \frac{\sum_{x=0}^{\infty} x^2 n_x - N. \bar{x}^2}{N(N-1)} \tag{320}$$

3.2.3. Since the distribution $P_o(\lambda)$ has only one parameter the method of moments suggests that we equate

$m_1' = \frac{\sum_{x=0}^{\infty} x \, n_x}{N} = \bar{x}$ to the first population moment (which is λ) to obtain as an estimate of λ

$$\tilde{\lambda} = \bar{x} \tag{321}$$

However, making use of the fact that the variance of the distribution $P_o(\lambda)$ is λ, we have that

$\lambda = \frac{\mu_2'}{\mu_1'} - 1$, and using m_r' , r = 1, 2 as estimates of their expected values we obtain the estimate

$$\tilde{\lambda} = \frac{m_2'}{m_1'} - 1 \tag{322}$$

The estimates for the Poisson parameter, given by (321) and (322) can be obtained as special cases of a more general method of estimation.

Suppose the range T of the gpsd (137) is $\{c, c+1, \dots. d\}$ where d may be finite or infinite. Then the mean μ is given by

$$\mu = \sum_{y=c}^{d} y.p(y;\theta) = c\ p(c;\theta) + \sum_{y=c+1}^{d} y.p(y;\theta)$$

$$= c\ p(c;\theta) + \sum_{y=c}^{d-1} (y+1).p(y+1;\theta)$$

$$= c\ p(c;\theta) + \theta \sum_{y=c}^{d-1} (y+1) \frac{a(y+1)}{a(y)}\ p(y;\theta) \quad (323)$$

And the second moment about the origin is given by

$$\mu_2' = \sum_{y=c}^{d} y^2.p(y;\theta) = \sum_{y=c}^{d} \left[y(y-1) + y \right].p(y;\theta)$$

$$= c(c-1)\ p(c;\theta) + \sum_{y=c+1}^{d} y(y-1).p(y;\theta) + \mu$$

$$= c(c-1)\ p(c;\theta) + \theta \sum_{y=c}^{d-1} (y+1)y.\frac{a(y+1)}{a(y)}.p(y;\theta) + \mu \quad (324)$$

Eliminating c $p(c;\theta)$ from (323) and (324) we obtain, $c \neq 0$

$$\frac{\mu_2' - \mu - \theta \sum_{y=c}^{d-1} y(y+1) \frac{a(y+1)}{a(y)}\ p(y;\theta)}{\mu - \theta \sum_{y=c}^{d-1} (y+1) \frac{a(y+1)}{a(y)}\ p(y;\theta)} = c - 1$$

which when solved for θ yields the identity; $c \neq 0$

$$\theta = \frac{\mu_2' - c\,\mu}{\sum_{c}^{d-1} y(y+1) \frac{a(y+1)}{a(y)} p(y;\theta) - (c-1) \sum_{c}^{d-1} (y+1) \frac{a(y+1)}{a(y)} p(y;\theta)} \quad (325)$$

While if c is zero we have from (323) the identity

$$\theta = \frac{\mu}{\sum_{y=c}^{d-1} (y+1) \dfrac{a(y+1)}{a(y)} p(y;\theta)} \tag{326}$$

Then, if the four quantities

$$N.m_1' = \sum_{y=c}^{d} y\, n_y \qquad ; \quad N.m_2' = \sum_{y=c}^{d} y^2.n_y$$

$$g_1 = \sum_{y=c}^{d-1} (y+1) \frac{a(y+1)}{a(y)} n_y \text{ and } g_2 = \sum_{y=c}^{d-1} y(y+1)\frac{a(y+1)}{a(y)} n_y \tag{327}$$

are computed from the sample and used as estimates of their expected values, we obtain the following estimator W for θ; where from (325) and (326)

$$W = \frac{\sum_{y=c}^{d-1} y^2 n_y - \sum_{y=c}^{d-1} y\, n_y}{g_2 - (c-1)\, g_1} \quad ; \text{ when } c \neq 0, \text{ and} \tag{328}$$

$$W = \frac{\sum_{y=c}^{d-1} y\, n_y}{g_1} \quad ; \text{ when } c = 0 \tag{329}$$

This general method of estimation has been called the "two-moments method" by Patil (1962,a) and the estimator W the two-moments estimator for θ.

Now suppose the random variable x has the distribution $P_0(\lambda)$. Then $c = 0$, $d = \infty$ and the quantities g_1 and g_2 of (327) reduce to

$$g_1 = \sum_{x=0}^{\infty} n_x = N$$

$$g_2 = \sum_{x=0}^{\infty} x\, n_x = N.m_1'$$

The two-moments method will provide an estimator W for the parameter λ, where from (329)

$$W = \frac{\sum_{x=0}^{\infty} x\, n_x}{N} = \bar{x} \tag{330}$$

But this is the same estimator for λ that was provided in (321) by the method of moments. Now observe that for the distribution $P_0(\lambda)$ equation (324) becomes, since c = 0

$$\mu_2' = \lambda \sum_{x=0}^{\infty} x\, p(x;\lambda) + \mu \quad \text{and} \quad \lambda = \frac{\mu_2'}{\mu} - 1$$

so that an estimator for λ is provided by

$\tilde{\lambda} = \dfrac{m_2'}{m_1'} - 1$, which is the same as (322) and is based upon the first two moments.

We therefore suggest that the name "two-moments estimator" be reserved for the estimator given in (322), rather than the estimator W in (330).

3.2.4. Let $x_1, x_2, \ldots x_N$ be a random sample of size N from the distribution $P_0(\lambda)$. Then the likelihood function becomes

$$L(\lambda; x_1, \ldots x_N) = \frac{e^{-N\lambda}\, \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!}$$

and $\log L = \text{constant} + \sum_{i=1}^{N} x_i . \log \lambda - N\lambda$

and the solution $\hat{\lambda}$ of the likelihood equation is

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x} \tag{331}$$

Now it has been shown by Patil (1962,c) that the method of maximum likelihood and the method of moments give the same estimate in the case of a gpsd involving a single parameter. For, if $y_1, y_2, \ldots y_N$ be a random sample from the gpsd (137), then the logarithm of the likelihood function becomes

$$\log L = \text{constant} + \sum_{i=1}^{N} y_i \log \theta - N.\log f(\theta)$$

$$\frac{\partial}{\partial\theta}\log L = \sum_{i=1}^{N} y_i \cdot \frac{1}{\theta} - \frac{N \frac{df}{d\theta}}{f(\theta)}$$

$$= \frac{N}{\theta}[\bar{y} - \mu] \quad , \text{ using (138)}$$

Thus equating $\frac{\partial}{\partial\theta}\log L$ to zero to obtain the m.l.e. of θ, say $\hat{\theta}$, is equivalent to equating the sample mean to the population mean.

The asymptotic variance of $\hat{\theta}$ is given by equation (304) and

$$\frac{1}{-\mathcal{E}\left[\frac{\partial^2}{\partial\theta^2}\log L\right]} = \frac{1}{-\mathcal{E}\left[\frac{\partial}{\partial\theta}\left\{\frac{N\bar{y}}{\theta} - \frac{N}{\theta}\cdot\mu\right\}\right]}$$

$$= \frac{-1}{\mathcal{E}\left[-\frac{N\bar{y}}{\theta^2} - \frac{N}{\theta}\frac{d\mu}{d\theta} + \frac{N}{\theta^2}\cdot\mu\right]}$$

$$= \frac{-1}{\mathcal{E}\left[-\frac{N}{\theta}\frac{d\mu}{d\theta} - \frac{N}{\theta^2}(\bar{y}-\mu)\right]} = \frac{1}{\frac{N}{\theta}\cdot\frac{d\mu}{d\theta}}$$

Then from (140) we obtain

$$V(\hat{\theta}) = \frac{\theta^2 \cdot 1}{N\,\mu_2(\theta)} \tag{332}$$

Thus, for the distribution $P_o(\lambda)$ we may obtain

$$V(\hat{\lambda}) = \frac{\lambda^2}{N} \cdot \frac{1}{\lambda} = \frac{\lambda}{N} \tag{333}$$

3.2.5. Minimum variance unbiased estimation has been discussed by Roy and Mitra (1957) and by Guttman (1958) for the psd; and by Patil (1962,b) for the gpsd. Roy and Mitra obtained the MVU estimator for the parameter θ of the psd. Patil has derived a necessary and sufficient condition for the parameter of the gpsd to be MVU estimable. This condition is expressed in terms of the number theoretic structure of the range of the gpsd.

Suppose T is the range of the gpsd (137), and $\{a\}$ denotes the set consisting of only one number a. Patil has shown that an unbiased estimator for θ exists iff $T + \{1\} \subseteq T$. It is an immediate consequence of this result that an unbiased estimator does not exist for the parameter of the Poisson distribution truncated on the right, which will be proved in section 3.3.2.

For the distribution $P_o(\lambda)$ with p.f. $p(x;\lambda)$, we have

$$\log p(x;\lambda) = -\lambda + x \log\lambda - \log x! \qquad \text{so that}$$

$$\begin{aligned} \tau^2 = \mathcal{E}\left[\frac{\partial}{\partial\lambda}\log p(x;\lambda)\right]^2 &= \sum_{x=0}^{\infty}\left[\frac{x}{\lambda} - 1\right]^2 . p(x;\lambda) \\ &= \frac{1}{\lambda^2}\cdot\sum_{x=0}^{\infty}[x-\lambda]^2\, p(x;\lambda) \\ &= \frac{1}{\lambda^2}\cdot V(x) \end{aligned} \tag{334}$$

Thus if $t_N(x_1, \ldots x_N) = t_N$ be any unbiased estimator for λ based on a random sample of size N, we have from the Cramer-Rao inequality that the minimum possible variance for t_N is

$$\frac{1}{N\tau^2} = \frac{\lambda}{N}$$

To see that this lower bound is attainable by some unbiased estimator t_N, we have that

$$\sum_{i=1}^{N} \frac{\partial}{\partial\lambda} \log p(x;\lambda) = \frac{N}{\lambda} \cdot \left[\frac{1}{N} \cdot \sum_{i=1}^{N} x_i - \lambda \right] \tag{335}$$

which according to (307) shows that the minimum is attainable and that the statistic $t_N = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i = \bar{x}$ is an efficient estimator for λ, for any given sample size. Then $V(t_N) = \frac{\lambda}{N}$; which is the same result as given by (333).

If $x_1, x_2, \ldots x_N$ be a random sample of size N from the distribution $P_o(\lambda)$ then the joint p.f. of the sample is

$$p(x_1;\lambda)\ldots\ldots p(x_N;\lambda) = \frac{e^{-N\lambda} \lambda^{x_1+\cdots+x_N}}{x_1! \ldots x_N!} \quad ; \; x_i = 0, 1, 2 \ldots \quad i = 1, 2, \ldots N \tag{336}$$

We have from section 1.11 that the statistic $z = x_1 + x_2 + \ldots + x_N$ has the distribution $P_o(N\lambda)$. Now (336) may be factored as

$$p(x_1;\lambda)\ldots p(x_N;\lambda) = \left[\frac{e^{-N\lambda}(N\lambda)^z}{z!} \right] \cdot \frac{(x_1 + x_2 + \ldots + x_N)!}{N^{x_1+\cdots+x_N} \prod_{i=1}^{N} x_i!}$$

$$= p(z;N\lambda) . H(x_1, \ldots x_N)$$

so that z is a sufficient statistic for λ (section 3.1.3).

[This is immediately seen to be so by recognizing the joint p.f. in (336) as a regular case of the K-P class (section 3.1.5) and hence, by inspection, the statistic $z = x_1 + x_2 + .. + x_N$ is sufficient for λ .]

If $u_1(z)$ and $u_2(z)$ are unbiased estimators for λ then the expected value of $u(z) = u_1(z) - u_2(z)$ will be zero for all $\lambda \varepsilon \Omega = (0,\infty)$. Thus

$$\mathcal{E}[u(z)] = \sum_{z=0}^{\infty} u(z) \frac{e^{-N\lambda}(N\lambda)^z}{z!} = 0$$

i.e. $e^{-N\lambda}[\, u(0) + u(1)\, N\lambda + u(2)\, \frac{(N\lambda)^2}{2!} + \ldots . \,] = 0$

and $e^{-N\lambda} > 0$ since $\lambda > 0$. We then have a power series in $N\lambda$ converging to 0 so that the coefficients $u(0) = u(1) = \ldots . = 0$, and hence $u_1(z) = u_2(z)$ with probability one. Thus the distribution of z is seen to be complete and there is only one unbiased estimator for λ based on z (section 3.1.6).

Since $\mathcal{E}(z) = N\lambda$ the MVU estimator for λ is easily obtained as $\frac{z}{N} = \bar{x}$ (337)

3.3 Single truncation on the right

3.3.1. Ratio estimate. Consider a random sample of size N drawn from the distribution with p.f. given by (205) i.e. we have truncation on the right at d. We will use the same notation as in 3.2.1. The ratio estimate (313) becomes R* say

$$\text{where } R^* = \frac{\sum_{x=1}^{d} x\, n_x}{\sum_{x=0}^{d-1} n_x} = \frac{\sum_{x=0}^{d} x\, n_x}{\sum_{x=0}^{d-1} n_x} \qquad (338)$$

Now it was Moore (1952) who first considered ratio estimation.* In his paper he noted that for any d, the Poisson distribution has the following property

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

and therefore λ may be written as

$$\lambda = \frac{\sum_{x=0}^{d} x\, p(x;\lambda)}{\sum_{x=0}^{d-1} p(x;\lambda)} \qquad (339)$$

Subsequently, Murakami and Co. (March, 1954) were considering estimation for the case of truncation on the right. It is obvious that by dividing both numerator and

---

* The statistical term "censored" was introduced by Hald to distinguish the two cases discussed in 2.1. Until this terminology became generally accepted the distinction was not made clear in the literature, and both cases are described by "truncation". Moore (1952) was actually considering ratio estimation for the same problem that had been approached by Tippett (1932) from the point of view of maximum likelihood i.e. the distribution singly censored on the right. (3.6.3)

denominator of (339) by $F(d;\lambda)$ we may write

$$\lambda = \frac{\sum_{x=0}^{d} x\, p^*(x;\lambda)}{\sum_{x=0}^{d-1} p^*(x;\lambda)} \tag{340}$$

where $p^*(x;\lambda)$ is defined in (205). When the observed frequencies are used as estimates of their expected values (340) immediately leads to the estimator R^*.

R^* is the estimator for λ given by Murakami and Co. for the case of truncation on the right and in their paper they refer to this estimator as "Moore's estimator" for obvious reasons. However, in the same year Moore published another paper (September, 1954) in which he noted that the type of estimator he had previously proposed for the censored case is applicable to any of the left, right or doubly truncated cases.

In order to find $\mathcal{E}(R^*)$ we may write (338) in the form

$$R^* = \frac{\sum_0^d x\, n_x}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} \left[1 - \frac{\mathcal{E}\left(\sum_0^{d-1} n_x\right) - \sum_0^{d-1} n_x}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} \right]^{-1}$$

$$= \frac{\sum_0^d x\, n_x}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} \left[1 + \frac{\mathcal{E}\left(\sum_0^{d-1} n_x\right) - \sum_0^{d-1} n_x}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} + \left\{ \frac{\mathcal{E}\left(\sum_0^{d-1} n_x\right) - \sum_0^{d-1} n_x}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} \right\}^2 + \cdots \right]$$

neglecting the squared and higher order terms

$$R^* \doteq \frac{\sum_0^d x\, n_x}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} \left[1 - \frac{\sum_0^{d-1} n_x - \mathcal{E}\left(\sum_0^{d-1} n_x\right)}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} \right]$$

and

$$\mathcal{E}(R^*) \doteq \frac{\mathcal{E}\left(\sum_0^d x n_x\right)}{\mathcal{E}\left(\sum_0^{d-1} n_x\right)} - \left[\frac{\mathcal{E}\left(\sum_0^d x n_x \sum_0^{d-1} n_x\right) - \mathcal{E}\left(\sum_0^d x n_x\right)\mathcal{E}\left(\sum_0^{d-1} n_x\right)}{\left\{\mathcal{E}\left(\sum_0^{d-1} n_x\right)\right\}^2}\right]$$

now $\sum_0^d x\ n_x . \sum_0^{d-1} n_x = (\ 1.n_1 + 2n_2 + \ldots + dn_d)(\ n_0 + n_1 + \ldots + n_{d-1})$

$$= \sum_0^{d-1} x\ n_x^2 + \sum_{x \neq y}^{d} \sum^{d-1} x\ n_x\ n_y$$

and using the expected values given by (310) as they apply to the truncated distribution, we have immediately

$$\mathcal{E}(R^*) \doteq \frac{\sum_0^d x\ p^*(x;\lambda)}{\sum_0^{d-1} p^*(x;\lambda)} - \frac{N \sum_0^{d-1} x\ p^*(x;\lambda) + N(N-1) \sum_0^{d-1} x[p^*(x;\lambda)]^2}{N^2 \left\{\sum_0^{d-1} p^*(x;\lambda)\right\}^2}$$

$$- \frac{N(N-1) \sum_{x \neq y}^{d} \sum^{d-1} xp^*(x;\lambda)p^*(y;\lambda) + N^2 \sum_0^d xp^*(x;\lambda) \sum_0^{d-1} p^*(x;\lambda)}{N^2 \left\{\sum_0^{d-1} p^*(x;\lambda)\right\}^2}$$

and from (340) we may write

$$\mathcal{E}(R^*) \doteq \lambda - \frac{\sum_0^{d-1} xp^*(x;\lambda) - \sum_0^d x\ p^*(x;\lambda) . \sum_0^{d-1} p^*(x;\lambda)}{N \left\{\sum_0^{d-1} p^*(x;\lambda)\right\}^2} \qquad (341)$$

which is the expression given by Murakami and Co.

Using (340) again we have

$$\mathcal{E}(R^*) \doteq \lambda\left(1 + \frac{1}{N}\right) - \frac{\sum_0^{d-1} x\ p^*(x;\lambda)}{N \left\{\sum_0^{d-1} p^*(x;\lambda)\right\}^2} \qquad (342)$$

Thus R* is not an unbiased estimator for λ ; but owing to the presence of N in the denominator we see that the amount of bias is very small.

3.3.2. It will be fruitless to make some adjustment to R* in order to obtain an unbiased estimator for λ. For it can be proved that there does not exist an unbiased estimator for the case of truncation on the right. This fact is established in the following

<u>Result:</u> For the distribution with p.f. given by (205) where d is finite, there does not exist an unbiased estimator for λ.

Proof: Let $x_1, x_2, \ldots x_N$ be a random sample of size N from the distribution (205). Let $t_N(x_1, \ldots x_N) = t_N$ be any function of the observations alone, such that $\mathcal{E}(t_N) \equiv \lambda$. The joint p.f. of the sample is $h(x_1, \ldots x_N;\lambda)$ say, where

$$h(x_1, \ldots x_N;\lambda) = \frac{e^{-N\lambda}\ \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!\ [F(d;\lambda)]^N}$$

and the expected value of t_N will be $\sum t_N . h(x_1, \ldots x_N;\lambda)$, where the summation extends over the range 0 to d of each of the x_i, $i = 1, 2, \ldots N$. Thus

$$\mathcal{E}(t_N) = \sum_{x_1=0}^{d} \cdots \sum_{x_N=0}^{d} t_N \frac{e^{-N\lambda}\ \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!\ [F(d;\lambda)]^N} \equiv \lambda$$

by hypothesis. That is,

$$\sum_{x_1=0}^{d} \cdots \sum_{x_N=0}^{d} \frac{t_N}{\prod_{i=1}^{N} x_i!} \lambda^{\sum_{i=1}^{N} x_i} \equiv \lambda \left\{ \sum_{j=0}^{d} \frac{\lambda^j}{j!} \right\}^N \tag{343}$$

which is an identity between two polynomials in the argument λ. Now the coefficient of λ^{Nd+1} in the right hand side of (343) is $[\frac{1}{d!}]^N$, while in the left hand side this coefficient is zero. However, for finite d and N, $[\frac{1}{d!}]^N$ is not zero -- a contradiction that establishes the result.

The non-existence of an unbiased estimator for λ in the case of truncation on the right has been observed by Tate and Goen (1958), who indicated the above argument based on the identity (343).

3.3.3. We have seen in (342) that the ratio estimate R* is biased, though the amount of bias is small. In order to investigate the efficiency of R* we require V(R*).

An approximate expression for V(R*) has been given by Murakami and Co. (1954). Noting that R* is defined as the ratio of two random variables, they apply the approximation formula

$$V[\frac{\xi}{\eta}] \doteq \left\{\frac{\mathcal{E}(\xi)}{\mathcal{E}(\eta)}\right\}^2 \cdot \left[\frac{V(\xi)}{\{\mathcal{E}(\xi)\}^2} + \frac{V(\eta)}{\{\mathcal{E}(\eta)\}^2} - 2\frac{\mathrm{Cov}(\xi,\eta)}{\mathcal{E}(\xi)\,\mathcal{E}(\eta)}\right] \quad (344)$$

The quantities not yet derived and required for (344) may be obtained in a similar manner to that demonstrated in the derivation of (342). These quantities are obtained by Murakami and may be written

$$V[\sum_{x=0}^{d-1} n_x] = N\,[\,1 - p^*(d;\lambda) - \{1 - p^*(d;\lambda)\}^2\,]$$

$$V[\sum_{x=0}^{d} x\,n_x] = N.V^*(x) \;;\quad \text{where } V^*(x) \text{ is given by (209).}$$

$$\text{Cov}\left(\sum_{0}^{d} x n_x, \sum_{0}^{d-1} n_x\right) = N[\mu^*(d;\lambda) - dp^*(d;\lambda) - \mu^*(d;\lambda)\{1 - p^*(d;\lambda)\}]$$

The approximate expression for V(R*) reduces to

$$V(R^*) \doteq \frac{\lambda^2}{N}\cdot\left\{\frac{V^*(x)}{[\mu^*(d;\lambda)]^2} + \frac{p^*(d;\lambda)[\,2d - \mu^*(d;\lambda)]}{\mu^*(d;\lambda)[\,1 - p^*(d;\lambda)]}\right\}$$

Murakami and Co. have graphed the relative efficiency of the ratio estimate R* to the maximum likelihood estimate $\hat{\lambda}$ for the values of d = 1(1)10 for λ going from 0 to 10. The m.l.e. is shown to have smaller variance than the ratio estimate. However, the authors remark that the ease with which R* is calculated may be thought to outweigh the advantages of the m.l.e.

An approximate expression for V(R*) has also been given by Patil (1959) who obtains

$$V(R^*) \doteq \frac{\sum_{x=0}^{d} x^2 p^*(x;\lambda) - \lambda^2 \sum_{x=0}^{d-1} p^*(x;\lambda) + 2\lambda^2 p^*(d-1;\lambda)}{N\left\{\sum_{x=0}^{d-1} p^*(x;\lambda)\right\}^2}$$

which may be written, using (210)

$$V(R^*) = \frac{\mu^*(d;\lambda)[\,1 + \mu^*(d-1;\lambda)] - \lambda^2[\,1 - p^*(d;\lambda) - 2p^*(d-1;\lambda)]}{N\,[\,1 - p^*(d;\lambda)]^2}$$

Patil has considered the efficiency of R* for values of d = 5 with λ = .25, 0.5(0.5)2.5 and also for d = 10 with λ = 0.5(0.5)4.5. R* is shown to be highly efficient, with efficiency never less than 82% for the arguments considered. Patil also shows that R* has almost negligible bias.

3.3.4. Let x_1, x_2, ... x_N be a random sample of size N from the distribution (205). The likelihood function is

$$L(\lambda;x_1, \ldots x_N) = \frac{e^{-N\lambda} \; \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i! \; [F(d;\lambda)]^N}$$

Write $N\bar{x}$ in place of $\sum_{i=1}^{N} x_i$ i.e. $\bar{x}$ is the sample mean; and take logarithms to obtain

$$\log L = -N\lambda + N\bar{x}.\log\lambda - \log\prod_{i=1}^{N} x_i! - N\log F(d;\lambda)$$

$$\frac{1}{N}\cdot\frac{\partial \log L}{\partial\lambda} = \frac{\bar{x}}{\lambda} - 1 + \frac{p(d;\lambda)}{F(d;\lambda)}$$

using equation (107). The m.l.e. of λ is given by the solution of the equation

$$\bar{x} = \lambda\left[1 - \frac{p(d;\lambda)}{F(d;\lambda)}\right] = \lambda\frac{F(d-1;\lambda)}{F(d;\lambda)} \tag{345}$$

which from (207), $\bar{x} = \mu^*(d;\lambda)$ (346)

Thus the solution $\hat{\lambda}$ can be obtained using tables of Poisson sums and (345). However, Cohen (1961) has provided a table of $\lambda\frac{F(d-1;\lambda)}{F(d;\lambda)}$ for the arguments d = 1(1)16 and λ = 0.1(0.1)4.0 Thus in practice, with d specified and $\bar{x}$ known from the sample the m.l.e. $\hat{\lambda}$ is readily calculated. Again, Patil (1959) has presented a table of values of $\mu^*(d;\lambda)$ for the arguments d = 4(1)10 and λ = 0.0(0.1)4.9 so that (346) may be employed to obtain $\hat{\lambda}$.

Now equation (345) was given by Cohen (1954) and in the same year by Murakami and Co. These latter authors presented nomograms which enable one to read off the value of $\hat{\lambda}$ directly, corresponding to the sample value $\bar{x}$, for a known value of d. Now,

$$\frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{N\bar{x}}{\lambda^2} + N\frac{F(d;\lambda)\frac{d}{d\lambda}p(d;\lambda) + [p(d;\lambda)]^2}{[F(d;\lambda)]^2}$$

$$\therefore -\mathcal{E}\left(\frac{\partial^2 \log L}{\partial \lambda^2}\right)$$

$$= \frac{N}{\lambda^2}.\lambda\frac{F(d-1;\lambda)}{F(d;\lambda)} - \frac{N.F(d;\lambda)[\,p(d-1;\lambda) - p(d;\lambda)]}{[F(d;\lambda)]^2} - N\frac{[p(d;\lambda)]^2}{[F(d;\lambda)]^2}$$

$$= \frac{N}{\lambda\{F(d;\lambda)\}^2}\cdot\Big[F(d;\lambda)[F(d-1;\lambda) - \lambda\,p(d-1;\lambda)] + \lambda F(d;\lambda)p(d;\lambda) - [p(d;\lambda)]^2\Big]$$

$$= \frac{N}{\lambda\{F(d;\lambda)\}^2}\cdot\Big[F(d;\lambda)[F(d-1;\lambda) - \lambda\,p(d-1;\lambda)] + \lambda\,F(d-1;\lambda)p(d;\lambda)\Big]$$

and hence the asymptotic variance of $\hat{\lambda}$,

$$V(\hat{\lambda}) = \frac{\lambda}{N}\cdot\frac{[F(d;\lambda)]^2}{F(d;\lambda)[\,F(d-1;\lambda) - \lambda\,p(d-1;\lambda)] + \lambda\,F(d-1;\lambda)p(d;\lambda)}$$

$$= \frac{\lambda}{N}\cdot\phi_d(\lambda)$$ say, which agrees with the expression given by Cohen (1961), who tabulates $\phi_d(\lambda)$ for d = 2(1)14 and selected values of λ from .001 to 15.0.

For d = 1, $\phi_1(\lambda)$ becomes $[\,\lambda + 1\,]^2$.

A neat expression for the asymptotic variance can be derived as follows. Now,

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

so that

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

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

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

$$= \frac{V^*(x)}{\lambda} \quad \text{from (208)}$$

Thus the asymptotic variance can be expressed as $\dfrac{\lambda^2}{N \,.\, V^*(x)}$

and if we now use (209) then Patil's (1959) table of $\mu^*(d;\lambda)$ is available for the calculation of $V(\hat{\lambda})$.

We observe that this expression for $V(\hat{\lambda})$ is in agreement with that given by (332) for the gpsd, with λ in place of θ.

3.4 Single truncation on the left

3.4.1. Ratio estimate. A random sample is drawn from the distribution with p.f. $p_*(x;\lambda)$ given by (211) i.e. the truncation point is on the left at c. Then the ratio estimate given by (313) becomes R_* say, where

$$R_* = \frac{\sum_{x=c+1}^{\infty} x\, n_x}{\sum_{x=c}^{\infty} n_x} = \frac{\sum_{x=c+1}^{\infty} x\, n_x}{N} \tag{347}$$

which is the estimator proposed by Moore (1954).

Now from (314) we know that R_* is an unbiased estimator for λ. While from (316) and (317) we have

$$V(R_*) = \frac{1}{N} \cdot \left[\sum_{x=c+1}^{\infty} x^2 p_*(x;\lambda) - \lambda^2 \right] \tag{348}$$

and an unbiased estimator for $V(R_*)$ is given by

$$S_* = \frac{\sum_{c+1}^{\infty} x^2 n_x - N.R_*^2}{N(N-1)}$$

Moore (1954) gave the approximate formula

$$V(R\ ) \doteqdot \frac{\sum_{c+1}^{\infty} x^2 p(x;\lambda) - \left\{ \sum_{c+1}^{\infty} x\ p(x;\lambda) \right\}^2}{N \left\{ \sum_{c}^{\infty} p(x;\lambda) \right\}^2} \tag{349}$$

which we may write, using (110) and (211), as

$$V(R_*) \doteqdot \frac{1}{N} \cdot \left[\frac{\sum_{c+1}^{\infty} x^2 p_*(x;\lambda)}{P(c;\lambda)} - \left\{ \sum_{c+1}^{\infty} x\ p_*(x;\lambda) \right\}^2 \right]$$

$$= \frac{1}{N} \cdot \left[\frac{\sum_{c+1}^{\infty} x^2 p_*(x;\lambda)}{P(c;\lambda)} - \left\{ \frac{\lambda\, P(c-1;\lambda)}{P(c;\lambda)} - c\ p_*(c;\lambda) \right\}^2 \right]$$

using (213); and by comparing this expression with the true value of $V(R_*)$ given in (345) we see the error terms involved in the approximate formula (349). Moore provided a table of values of $N.V(R_*)$ calculated from (349), for the arguments $c = 1, 2, 3$ and $\lambda = 1.0(0.5)4.0(1.0)6.0$

For the important special case where $c = 1$, the case of "missing zero-counts", we obtain

$$R_* = \frac{\sum_{x=2}^{\infty} x\, n_x}{N} \tag{350}$$

which is the unbiased estimator for λ first proposed by Plackett (1953). He derived this estimator by commencing with an estimator of the form

$\theta_* = \dfrac{\sum_{x=1}^{\infty} s_x\, n_x}{N}$ and evaluating the unknowns s_x by the requirement that θ_* is to be unbiased. Thus we obtain

$$\mathcal{E}(\theta_*) = \sum_{x=1}^{\infty} s_x \frac{e^{-\lambda}\lambda^x}{x!\,[\,1 - e^{-\lambda}\,]} = \lambda$$

that is $\displaystyle\sum_{x=1}^{\infty} s_x \frac{\lambda^x}{x!} = \lambda\,[\,e^{\lambda} - 1\,] = \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-1)!}$

By comparing coefficients of λ^x we have

$s_1 = 0,\ s_2 = 2,\ s_3 = 3 \ldots\ldots s_x = x$ for $x \geqslant 2$ and

the desired estimate is

$\theta_* = \dfrac{\sum_{x=2}^{\infty} x\, n_x}{N}$ as given in (350) by R_*, which incidentally may be written in the alternative form $\bar{x} - \frac{n_1}{N}$ Again, when $c = 1$ we obtain from (348)

$$V(R_*) = \frac{1}{N} \cdot \left[\sum_{x=2}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!\ P(1;\lambda)} - \lambda^2 \right] \tag{351}$$

(note $x = 1$ is not included in the summation).

Now $\sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!\ P(1;\lambda)}$ is the second moment about the origin of the distribution (217) which we have obtained in (220) as

$\mu'_{*2}(1;\lambda) = \frac{\lambda(1+\lambda)}{1 - e^{-\lambda}}$. It follows that (351) becomes

$$V(R_*) = \frac{1}{N} \cdot \left[\frac{\lambda(1+\lambda)}{1-e^{-\lambda}} - \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} - \lambda^2 \right]$$

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

$$= \frac{1}{N} \cdot \left[\lambda + \frac{\lambda^2}{e^{\lambda} - 1} \right]$$

which agrees with the expression given by Plackett. By using the same technique as demonstrated above he derived an unbiased estimate for $V(R_*)$, when $c = 1$, as

$$\frac{\sum_{x=2}^{\infty} x\, n_x + 2\, n_2}{N^2}.$$

Plackett also examined the efficiency of the estimator given in (350) and showed that this tends to 100% as $\lambda \to 0$ or $\lambda \to \infty$, and never falls below 95%. Minimum efficiency is attained for $\lambda = 1.355$.

Rider (1953) has considered a slightly different problem. A sample is drawn from a complete Poisson distribution; but the observations for some of the lowest frequency classes are missing (it is the sample that is then said to be truncated on the left). From sample moment considerations Rider easily obtains the estimator M for λ,

where
$$M = \frac{\sum_{x=k}^{\infty} x^2 n_x - k \sum_{x=k}^{\infty} x\, n_x}{\sum_{x=k}^{\infty} x\, n_x - (k-1) \sum_{x=k}^{\infty} n_x}$$

and k is the number of missing lowest frequency classes.

3.4.2. Let $x_1, \ldots x_N$ be a random sample of size N from the distribution with p.f. $p_*(x;\lambda)$ given by (211). Then the likelihood function is

$$L(\lambda; x_1, \ldots x_N) = \frac{e^{-N\lambda} \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i! \; [P(c;\lambda)]^N}$$

Taking logarithms, and writing $N\bar{x}$ in place of $\sum_{i=1}^{N} x_i$ we obtain

$$\log L = -N\lambda + N\bar{x}.\log\lambda - \log \prod_{i=1}^{N} x_i! - N.\log P(c;\lambda)$$

$$\frac{1}{N} \cdot \frac{\partial \log L}{\partial \lambda} = \frac{\bar{x}}{\lambda} - 1 - \frac{p(c-1;\lambda)}{P(c;\lambda)} \tag{352}$$

using (111). The m.l.e. of λ is given by the solution of the equation

$$\bar{x} = \lambda\left[1 + \frac{p(c-1;\lambda)}{P(c;\lambda)}\right] = \lambda \frac{P(c-1;\lambda)}{P(c;\lambda)} \tag{353}$$

or from (213), $$\bar{x} = \mu_*(c;\lambda) \tag{354}$$

Thus the solution $\hat{\lambda}$, can be found from (353) with the aid of tables of the function $P(x;\lambda)$. However, from (354) $\hat{\lambda}$ may be obtained more simply by using Patil's (1959) table of $\mu_*(c;\lambda)$, $c = 1(1)10$ and $\lambda = 0.0(0.1)9.9$

For the special case $c = 1$, the m.l.e. of λ is the solution of the equation

$$\bar{x} = [\ 1 - e^{-\lambda}\ ]^{-1} \tag{355}$$

This result was obtained by David and Johnson (1952) who provided a small table of values of $\hat{\lambda}$ corresponding to $\bar{x} = 1.1(0.1)3.5$ to assist in the solution of (355). For $\bar{x} > 3.5$ a good approximation with an error less than 0.1% is given by $$\hat{\lambda} = \bar{x}\,[\ 1 - e^{-\bar{x}}\ ]$$

Equation (355) was also given by Rider (1953), who provided a table to facilitate obtaining the solution. Again, Cohen (1960) has provided an extensive table of values of $\hat{\lambda}$ corresponding to sample values of $\bar{x}$.

Putting $c = 1$ in (352) and taking the second derivative gives

$$\frac{1}{N}\frac{\partial^2 \log L}{\partial\lambda^2} = -\left[\ \frac{\bar{x}}{\lambda^2} - \frac{e^{-\lambda}}{[\ 1 - e^{-\lambda}\ ]^2}\ \right]$$

so that from (219) we have

$$-\,\mathcal{E}\left(\frac{\partial^2 \log L}{\partial\lambda^2}\right) = N\left[\ \frac{1}{\lambda[\ 1 - e^{-\lambda}]} - \frac{e^{-\lambda}}{[\ 1 - e^{-\lambda}\ ]^2}\ \right]$$

and the asymptotic variance of $\hat{\lambda}$ becomes

$$V(\hat{\lambda}) = \frac{\lambda(1 - e^{-\lambda})^2}{N[1 - (\lambda + 1)e^{-\lambda}]} = \frac{\lambda}{N} \cdot \psi(\lambda), \text{ say.} \tag{356}$$

We may observe that $\psi(\lambda)$ is continuous and monotone decreasing, that $\lim_{\lambda \to 0} \psi(\lambda) = 2$ and $\lim_{\lambda \to \infty} \psi(\lambda) = 1$ (which is easily seen from the graph of $\psi(\lambda)$ plotted against λ.) Therefore, regardless of the value of λ, the asymptotic variance satisfies the inequality $\frac{\lambda}{N} \leq V(\hat{\lambda}) \leq \frac{2\lambda}{N}$

This result is given by Cohen (1960) who provides the graph of $\psi(\lambda)$ against λ and a brief table of $\psi(\lambda)$ in order to calculate $V(\hat{\lambda})$.

Now observe that from equation (221) we may write (356) as follows

$$V(\hat{\lambda}) = \frac{\lambda^2}{N} \cdot \frac{1}{\mu_{*2}(1;\lambda)} = \frac{\lambda^2}{N} \cdot \frac{1}{V_*(x)}$$

where $V_*(x)$ is given, with $c = 1$, by (215) and thus Patil's (1959) tables are available for the calculation of $V(\hat{\lambda})$. This expression for $V(\hat{\lambda})$ is seen to be in the same form as (332) for the gpsd.

An explicit solution of equation (355) has been given by Irwin (1959) in the form of a Lagrange series expansion $\hat{\lambda} = \bar{x} - \sum_{r=1}^{\infty} \left\{ \frac{r^{r-1}}{r!} (\bar{x}e^{-\bar{x}})^r \right\}$

However, for small values of $\bar{x}$ convergence is slow and as a practical matter the numerous tables that are now available provide a more convenient method of solution.

3.4.3. For the distribution $P_0(\lambda)$ singly truncated on the left at c the MVU estimator for λ was obtained by Tate and Goen (1958). Roy and Mitra (1957) have discussed the case c = 1, and Cacoullos (1960) has also contributed to this case, which we now consider.

Let $x_1, x_2, \ldots x_N$ be a random sample of size N from the distribution with p.f. $p_*(x;\lambda)$ given by (217). The joint p.f. of the sample is

$$p_*(x_1;\lambda) \ldots p_*(x_N;\lambda) = \frac{e^{-N\lambda} \quad \lambda^{\sum_{i=1}^{N} x_i}}{[1 - e^{-\lambda}]^N \, x_1! \ldots x_N!} \tag{357}$$

Let the sample sum $z = \sum_{i=1}^{N} x_i$ have p.f. denoted by $h(z;\lambda)$. Tate and Goen obtained $h(z;\lambda)$ by using the characteristic function of z; while Cacoullos derived $h(z;\lambda)$ more simply by combinatorial methods. We have

$$\begin{aligned} h(z;\lambda) &= \text{Prob}[\, x_1 + x_2 + \ldots + x_N = z \,] \\ &= \sum_{x_1, \ldots x_N} p_*(x_1;\lambda) \ldots p_*(x_N;\lambda) \end{aligned}$$

where the summation is over all ordered N-part partitions of z; the parts being N-tuples ($x_1, x_2, \ldots x_N$) of integers such that $x_i \geqslant 1$; i = 1, 2, ... N and $\sum_{i=1}^{N} x_i = z$.

Hence we obtain, from (357)

$$h(z;\lambda) = \sum_{x_1, \ldots x_N} \frac{e^{-N\lambda} \quad \lambda^{z}}{[1 - e^{-\lambda}]^N \prod_{i=1}^{N} x_i!}$$

$$= \frac{\lambda^z}{[e^{\lambda} - 1]^N z!} \cdot \sum_{x_1, \cdots x_N} \frac{z!}{\prod_{i=1}^{N} x_i!} \tag{358}$$

Suppose now we imagine that z distinguishable balls have been tossed randomly into N cells, and the number of balls in the i^{th} cell is denoted by x_i, $i = 1, 2, \ldots N$. The problem of interest here is to find the number of ways in which this can be done subject to the condition that all cells be occupied. Evidently this will be the summation $\sum_{x_1, \cdots x_N} \frac{z!}{\prod_{i=1}^{N} x_i!}$ which appears in (358).

This classical "occupancy" problem has been considered by many writers, and in particular Feller (1957). Now, in the event that the i^{th} cell is empty, $i = 1, \ldots N$, all z balls are placed in the (N-1) remaining cells. This can be done in $(N-1)^z$ different ways. Similarly, in the event that two preassigned cells are empty there are $(N-2)^z$ possible arrangements, etc. The required number is obtained as

$$\sum_{i=0}^{N} (-1)^i \binom{N}{i} (N-i)^z \tag{359}$$

by a direct application of the theorem of probability relating to the realisation of at least one among N events.
Substituting k for N-i in (359) will lead to the result that

$$\sum_{x_1, \cdots x_N} \frac{z!}{\prod_{i=1}^{N} x_i!} = \sum_{k=0}^{N} (-1)^{N-k} \binom{N}{k} k^z \tag{360}$$

and now introduce the so called Stirling numbers of the second kind (Jordan, 1950) denoted by $\mathfrak{S}_z^N$ where

$$\mathfrak{S}_z^N = \frac{1}{N!} \cdot \sum_{k=0}^{N} (-1)^{N-k} \binom{N}{k} k^z \quad ; \; z = N, N+1, \ldots$$
$$= 0 \qquad\qquad ; \; z < N$$

Then from (358) and (360) we have

$$h(z;\lambda) = \frac{\lambda^z}{(e^\lambda - 1)^N} \cdot \frac{N!}{z!} \mathfrak{S}_z^N \quad ; \; z = N, N+1, \ldots \qquad (361)$$

Now in section 3.2.5 the sample sum was shown to be a sufficient statistic for the parameter of the complete Poisson distribution. It follows immediately from a theorem of Tukey (1949) that the sample sum is a sufficient statistic for the parameter of the truncated distribution also. Or, it is easily seen that the joint p.f. in (357) factorises into the p.f. $h(z;\lambda)$ given in (361) and a function of the observations alone. Thus sufficiency of the sample sum z is quickly established by the Fisher-Neyman Criterion (3.1.3). Again, proceeding in the same manner as in 3.2.5 it is easily shown that the distribution of z is complete.

Thus, if an unbiased estimator for λ based on z exists, it will be the unique MVU estimator, denoted by $\tilde{\lambda}$ say. The condition for unbiasedness of $\tilde{\lambda}$ is

$$\sum_{z=N}^{\infty} \tilde{\lambda} \cdot \frac{\lambda^z}{(e^\lambda - 1)^N} \cdot \frac{N!}{z!} \mathfrak{S}_z^N \equiv \lambda \qquad (362)$$

and in view of the fact that ;[93]

$$(e^\lambda - 1)^N = \sum_{z=N}^{\infty} \frac{\lambda^z N!}{z!} \mathfrak{S}_z^N$$

the identity (362) becomes

$$\sum_{z=N}^{\infty} \tilde{\lambda} \cdot \frac{\lambda^{z}}{z!} \mathfrak{S}_{z}^{N} = \sum_{z=N}^{\infty} \frac{\lambda^{z+1}}{z!} \mathfrak{S}_{z}^{N}$$

and a comparison of the coefficients of powers of λ yields

$$\tilde{\lambda} = z \frac{\mathfrak{S}_{z-1}^{N}}{\mathfrak{S}_{z}^{N}} = \frac{z}{N}\left(1 - \frac{\mathfrak{S}_{z-1}^{N-1}}{\mathfrak{S}_{z}^{N}}\right)$$

$\tilde{\lambda}$ may be evaluated by reference to a table of Stirling numbers of the second kind prepared by Miksa [48], who has tabulated $\mathfrak{S}_{z}^{N}$ for $N = 1(1)z$, $z = 1(1)50$.

However Tate and Goen (1958) provide a table of

$$\left(1 - \frac{\mathfrak{S}_{z-1}^{N-1}}{\mathfrak{S}_{z}^{N}}\right) \text{for } N = 2(1)z-1 \text{ , } z = 3(1)50$$

$\tilde{\lambda}$ is the same estimator obtained by Roy and Mitra (1957). These authors did not introduce Stirling numbers; but expressed $\tilde{\lambda}$ in terms of "differences of zero", i.e. $\Delta^{N} 0^{z} = N! \mathfrak{S}_{z}^{N}$. They calculated $\tilde{\lambda}$ for sample size $N = 2(1)10$ and sample sum $z = 2(1)96$.

3.5 Double truncation

3.5.1. Ratio estimate. A random sample is drawn from the distribution with p.f. $p_{\mathcal{D}}(x;\lambda)$ given by (222) i.e. the points of truncation are on the left at c and on the right at d. Then the ratio estimate (313) becomes $R_{\mathcal{D}}$ say, where

$$R_{\mathcal{D}} = \frac{\sum_{x=c+1}^{d} x\, n_x}{\sum_{x=c}^{d-1} n_x} \tag{363}$$

which is the estimator given by Moore (1954). It is clear that R_D is biased and, in fact, that there does not exist an unbiased estimator for λ in the case of double truncation.

3.5.2. If $x_1, x_2, \ldots x_N$ be a random sample of size N from the distribution (222), then the likelihood function is

$$L(\lambda; x_1, \ldots x_N) = \frac{e^{-N\lambda} \qquad \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i! \; [\; P(c;\lambda) - P(d+1;\lambda) \;]^N}$$

from which one obtains

$$\frac{1}{N} \cdot \frac{\partial \log L}{\partial \lambda} = \frac{\bar{x}}{\lambda} - 1 - \frac{p(c-1;\lambda) - p(d;\lambda)}{P(c;\lambda) - P(d+1;\lambda)}$$

and the m.l.e. of λ is the solution of the equation

$$\bar{x} = \lambda \left[1 + \frac{p(c-1;\lambda) - p(d;\lambda)}{P(c;\lambda) - P(d+1;\lambda)} \right]$$

$$= \lambda \left[\frac{P(c-1;\lambda) - P(d;\lambda)}{P(c;\lambda) - P(d+1;\lambda)} \right] \tag{364}$$

$$= \mu_D(c,d;\lambda) \quad \text{from (224)} \tag{365}$$

Thus with c and d known the solution $\hat{\lambda}$ may be found using tables of the function $P(x;\lambda)$ and (364). However, we present two charts below from which the m.l.e. $\hat{\lambda}$ may be obtained directly from the observed value of the sample mean in virtue of (365). The cases considered are c = 1, 2 and d = 2(1)9. Entering the appropriate chart with the observed value of the sample mean as ordinate we read off the m.l.e. $\hat{\lambda}$ as abscissa, corresponding to the known value of d. A table of values of the functions $\mu_D(1,d;\lambda)$ and $\mu_D(2,d;\lambda)$ are provided if interpolatory methods are preferred.

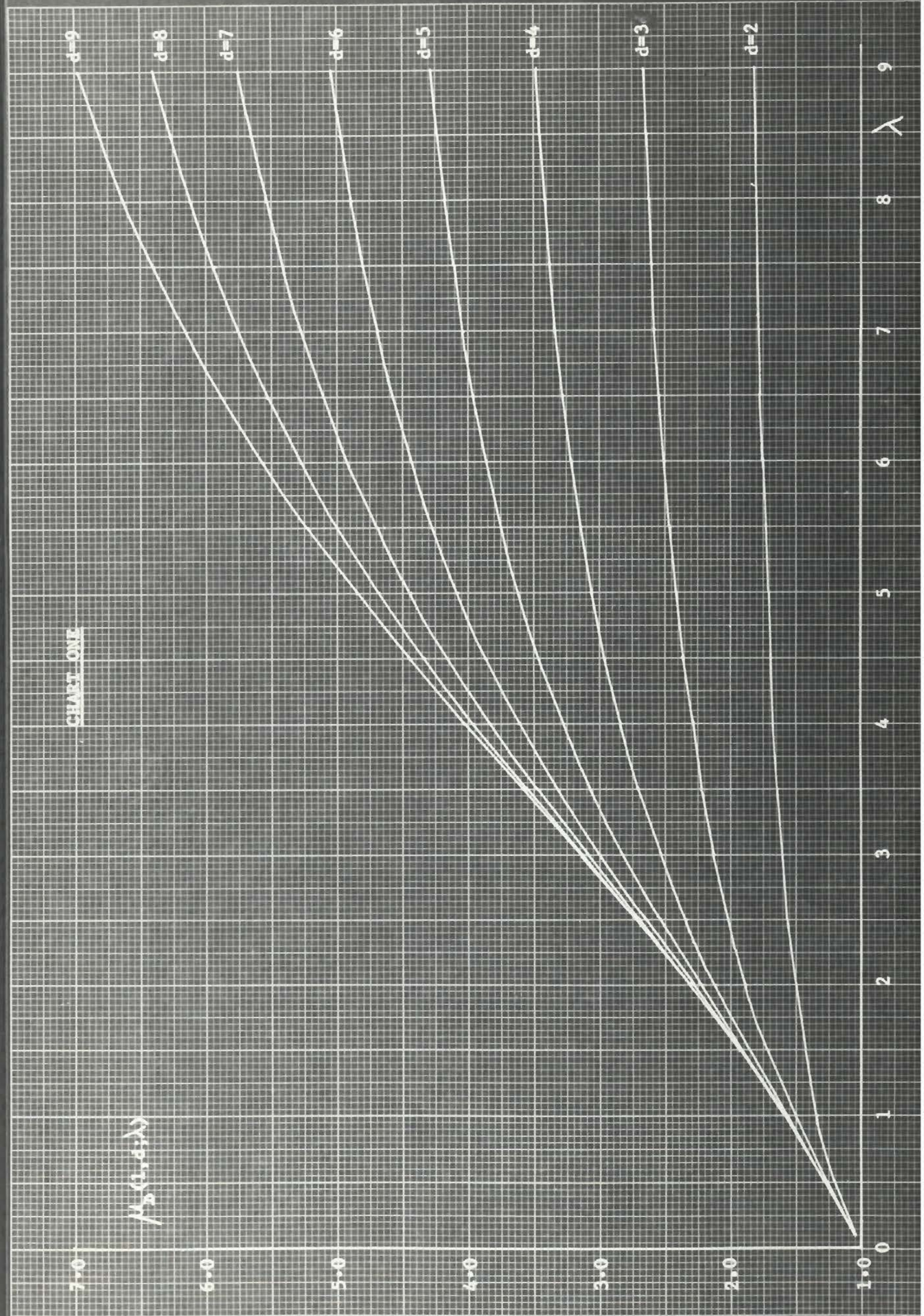
CHART ONE
μ_B(1,d;λ)
d=9
d=8
d=7
d=6
d=5
d=4
d=3
d=2
7·0
6·0
5·0
4·0
3·0
2·0
1·0
0
1
2
3
4
5
6
7
8
9
λ

CHART TWO

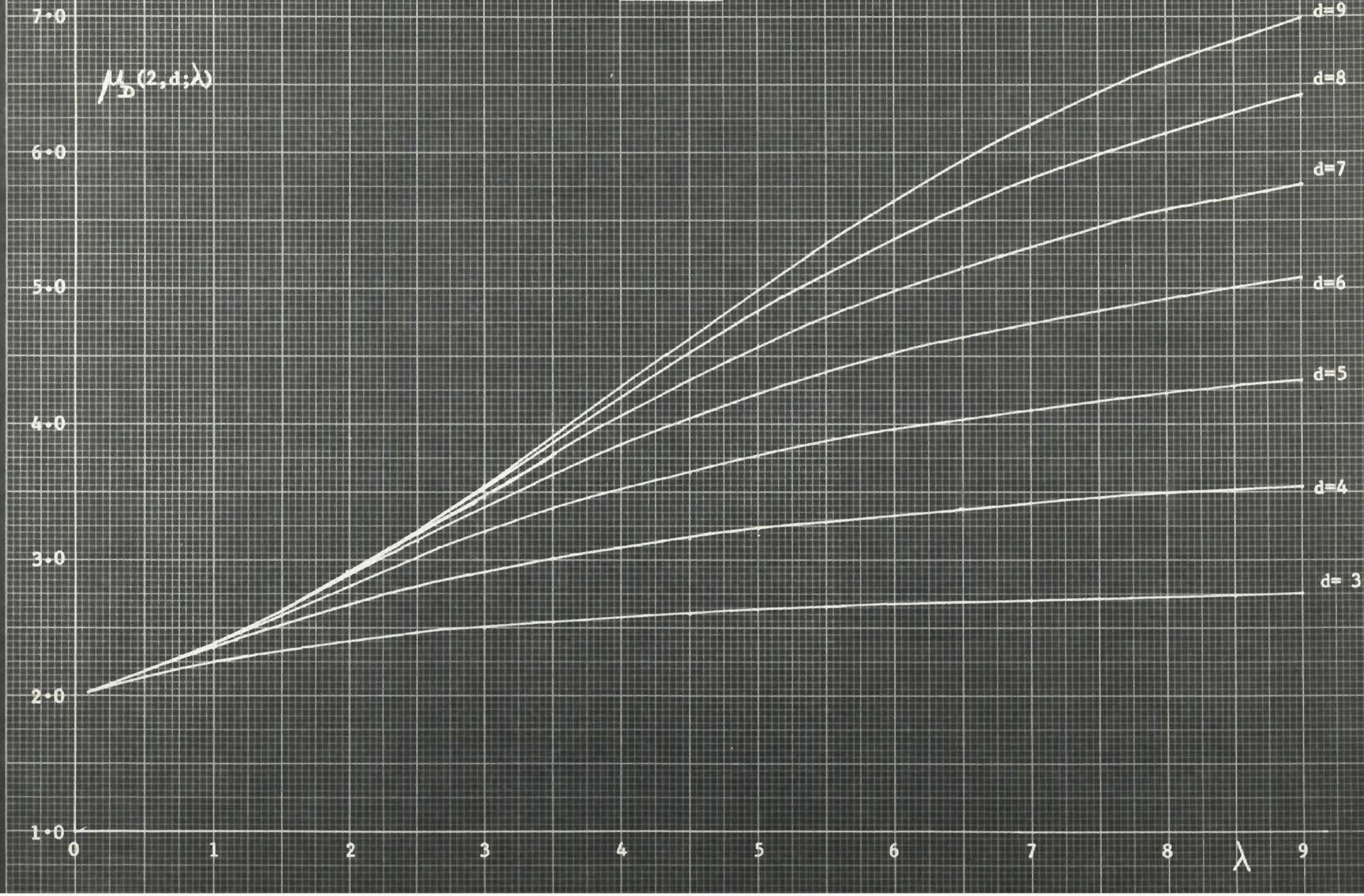
$\mu_D(2,d;\lambda)$
7·0
6·0
5·0
4·0
3·0
2·0
1·0
0
1
2
3
4
5
6
7
8
9
λ
d=9
d=8
d=7
d=6
d=5
d=4
d= 3

TABLE ONE : $\mu_{\nu}(1,d;\lambda)$

d \ λ	0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0
2	1.048	1.200	1.333	1.429	1.500	1.556	1.600	1.636	1.667	1.692	1.714	1.733	1.750	1.765	1.779	1.788	1.802	1.809	1.818
3	1.051	1.258	1.500	1.706	1.875	2.013	2.125	2.217	2.294	2.358	2.413	2.460	2.500	2.535	2.566	2.593	2.617	2.639	2.658
4	1.051	1.269	1.561	1.848	2.111	2.341	2.537	2.702	2.840	2.957	3.055	3.139	3.210	3.272	3.326	3.373	3.414	3.451	3.484
5	1.051	1.271	1.578	1.906	2.234	2.544	2.823	3.069	3.280	3.461	3.615	3.747	3.860	3.956	4.039	4.111	4.175	4.231	4.280
6	1.051	1.271	1.581	1.925	2.287	2.650	2.998	3.319	3.606	3.858	4.077	4.266	4.429	4.569	4.690	4.796	4.887	4.968	5.038
7	1.051	1.271	1.582	1.929	2.306	2.697	3.090	3.469	3.823	4.145	4.432	4.685	4.906	5.098	5.266	5.412	5.539	5.650	5.748
8	1.051	1.271	1.582	1.931	2.311	2.715	3.132	3.549	3.952	4.333	4.684	5.001	5.284	5.535	5.755	5.943	6.119	6.268	6.399
9	1.051	1.271	1.582	1.931	2.312	2.721	3.149	3.586	4.021	4.444	4.846	5.221	5.564	5.874	6.152	6.399	6.618	6.812	6.983

TABLE TWO : $\mu_{\nu}(2,d;\lambda)$

d \ λ	0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0
3	2.032	2.143	2.250	2.333	2.400	2.455	2.500	2.538	2.572	2.600	2.625	2.647	2.667	2.685	2.700	2.715	2.726	2.739	2.750
4	2.033	2.175	2.353	2.519	2.667	2.796	2.909	3.007	3.091	3.164	3.228	3.284	3.333	3.377	3.416	3.450	3.481	3.510	3.535
5	2.034	2.180	2.384	2.599	2.812	3.016	3.203	3.372	3.521	3.653	3.768	3.869	3.958	4.037	4.105	4.166	4.220	4.269	4.312
6	2.034	2.180	2.391	2.626	2.878	3.135	3.387	3.625	3.845	4.043	4.220	4.377	4.515	4.637	4.744	4.838	4.921	4.995	5.060
7	2.034	2.180	2.392	2.633	2.901	3.189	3.486	3.781	4.064	4.329	4.572	4.791	4.986	5.159	5.312	5.447	5.566	5.671	5.764
8	2.034	2.180	2.392	2.635	2.909	3.210	3.532	3.864	4.196	4.518	4.823	5.105	5.361	5.592	5.798	5.980	6.142	6.286	6.413
9	2.034	2.180	2.392	2.635	2.911	3.217	3.550	3.903	4.267	4.631	4.986	5.324	5.640	5.929	6.192	6.428	6.639	6.827	6.994

3.6 Single censorship on the right

3.6.1 The notation that will be used is given below.

possible value for x	0	1	2 ... d		>d	Total
observed frequency	n_0	n_1	n_2	n_d	n_t	N

and $$N = \sum_{x=0}^{d} n_x + n_t = K + n_t \text{ say.}$$

We assume that a sample of size N has been drawn from the distribution $P(\lambda)$ and that individual values have been obtained for the K observations for which $x \leq d$. However, the only information we have about the remaining $N - K = n_t$ observations in the right tail is that these observations exceed d.

3.6.2. Ratio estimate. The estimator proposed by Moore (1952) is

$$R^c = \frac{\sum_{x=0}^{d} x\, n_x}{\sum_{x=0}^{d-1} n_x} \tag{366}$$

which is seen to be the same estimator as R* given in (338) for the case of truncation on the right at d. Thus the number of observations n_t in the right tail is not employed in the estimation of λ by R^c.

Moore (1952) obtained an expression for $\mathcal{E}(R^c)$, which is the same as expression (341) for $\mathcal{E}(R^*)$; but with the quantity $p^*(x;\lambda)$ replaced by its complete counterpart $p(x;\lambda)$.

Now R^c is the ratio of two random variables and let us take the first term only of the formula (344) to obtain an approximate expression for $V(R^c)$. Then,

$$V(R^c) \doteq \frac{V[\sum_0^{\alpha} x\, n_x]}{\left\{\mathcal{E}(\sum_0^{\alpha-1} n_x)\right\}^2} \tag{367}$$

using the expectations of (310) we have

$$\mathcal{E}(\sum_0^{\alpha-1} n_x) = N \sum_0^{\alpha-1} p(x;\lambda) \tag{368}$$

$$\mathcal{E}(\sum_0^{\alpha} x\, n_x) = N \sum_0^{\alpha} x\, p(x;\lambda)$$

$$\begin{aligned}
\mathcal{E}\left[\left(\sum_0^{\alpha} x\, n_x\right)^2\right] &= \mathcal{E}\left[\sum_0^{\alpha}(x\, n_x)^2 + \sum\sum_{x \neq y}^{\alpha\ \alpha} x\, n_x . y\, n_y\right] \\
&= \sum_0^{\alpha} x^2 \left\{N\, p(x;\lambda) + N(N-1)[p(x;\lambda)]^2\right\} \\
&\quad + \sum\sum_{x \neq y}^{\alpha\ \alpha} x\, y\left\{N(N-1)\, p(x;\lambda)\, p(y;\lambda)\right\} \\
&= N(N-1)\left\{\sum_0^{\alpha} x\, p(x;\lambda)\right\}^2 + N\sum_0^{\alpha} x^2 p(x;\lambda)
\end{aligned}$$

Thus

$$\begin{aligned}
V\left(\sum_0^{\alpha} x\, n_x\right) &= \mathcal{E}\left[\left(\sum_0^{\alpha} x\, n_x\right)^2\right] - \left[\mathcal{E}\left(\sum_0^{\alpha} x\, n_x\right)\right]^2 \\
&= N\sum_0^{\alpha} x^2 p(x;\lambda) - N\left\{\sum_0^{\alpha} x\, p(x;\lambda)\right\}^2
\end{aligned} \tag{369}$$

Substituting (368) and (369) in (367) we obtain

$$V(R^c) \doteq \frac{\sum_0^{\alpha} x^2 p(x;\lambda) - \left\{\sum_0^{\alpha} x\, p(x;\lambda)\right\}^2}{N.\left\{\sum_0^{\alpha-1} p(x;\lambda)\right\}^2} \tag{370}$$

and (370) is the approximate expression given by Moore (1952). However, Murakami and Co. (1954) have employed the complete formula (344), and proceeding in exactly the same manner as in the case of truncation on the right they derive a more precise expression for $V(R^c)$. These authors have made a comparison of the standard errors of the m.l.e. (see next section) and R^c for the censored case. They have also graphed the relative efficiency of R^c to the m.l.e. $\hat{\lambda}$, that is the ratio $\frac{V(\hat{\lambda})}{V(R^c)}$, for values of d = 1(1)10 and λ going from zero to 10.

Their results show that for small values of λ, R^c is reasonably efficient; but is considerably inefficient for larger λ and they recommend the use of the m.l.e. rather than R^c. Recalling that R^c does not make use of the censored observations at all, while, as seen in the next section, the m.l.e. does employ this additional information, it is not surprising that R^c is the inferior estimator.

3.6.3. For the distribution $P_0(\lambda)$ singly censored on the right at d we see, from section 3.6.1, that the likelihood function of the random sample $x_1, x_2, \ldots x_K, \ldots x_{K+n_t}$ is

$$L(\lambda; x_1, \ldots x_K, \ldots x_{K+n_t}) = \binom{K+n_t}{K} \frac{e^{-K\lambda} \lambda^{\sum_{i=1}^{K} x_i}}{\prod_{i=1}^{K} x_i!} [P(d+1;\lambda)]^{n_t}$$

and $\log L = \text{constant} + \sum_{i=1}^{K} x_i . \log\lambda - K\lambda + n_t . \log P(d+1;\lambda)$

If we write $K\bar{x} = \sum_{i=1}^{K} x_i$, which incidentally is also $\sum_{x=0}^{d} x\, n_x$, we obtain after differentiation,

$$\frac{1}{K} \cdot \frac{\partial \log L}{\partial \lambda} = \frac{\bar{x}}{\lambda} - 1 + \frac{n_t}{K} \cdot \frac{p(d;\lambda)}{P(d+1;\lambda)} \tag{371}$$

and the m.l.e. for λ is the solution of the equation

$$\bar{x} = \lambda\left[1 - \frac{n_t}{K} \cdot \frac{p(d;\lambda)}{P(d+1;\lambda)}\right] \tag{372}$$

which is seen to be identical with the equation first given by Tippett (1932).

For the situations where there are not more than four individual frequency classes available, Tippett provided nomograms to aid the solution of equation (372).

Bliss (1948) developed an approximation to the maximum likelihood estimate. He also provided tables necessary for the application of his procedure of approximation.

Equation (372) was also presented by Cohen (1954), and by Murakami and Co. (1954). These latter authors constructed nomograms from which $\hat{\lambda}$ may be read off directly, corresponding to $d = 1(1)10$.

However, let us write equation (372) in another form. We have

$$K\bar{x} = K\lambda - n_t \cdot \lambda \frac{p(d;\lambda)}{P(d+1;\lambda)}$$

$$= N\lambda - n_t \cdot \lambda\left[1 + \frac{p(d;\lambda)}{P(d+1;\lambda)}\right]$$

and $$K\bar{x} = N\lambda - n_t \cdot \mu_*(d+1;\lambda)$$

where $\mu_*(d+1;\lambda)$ is the mean of the distribution truncated on the LEFT at d+1, from (213). Thus Patil's (1959) table referred to in section 2.4, is available for the calculation of $\hat{\lambda}$. From (371), and using (111) we have

$$\frac{1}{K}\cdot\frac{\partial^2 \log L}{\partial\lambda^2} = -\frac{\bar{x}}{\lambda^2} + \frac{n_t}{K}\cdot\left\{\frac{P(d+1;\lambda)[p(d-1;\lambda) - p(d;\lambda)] - [p(d;\lambda)]^2}{[P(d+1;\lambda)]^2}\right\}$$

$$= -\frac{\bar{x}}{\lambda^2} + \frac{n_t}{K}\cdot\left\{\frac{P(d+1;\lambda)\ p(d-1;\lambda) \ - \ P(d;\lambda)\ p(d;\lambda)}{[P(d+1;\lambda)]^2}\right\} \qquad (373)$$

The quantity in the curly brackets is recognised as being $\frac{\partial}{\partial\lambda}\left[\frac{P(d;\lambda)}{P(d+1;\lambda)}\right]$, which from (213) is

$$\frac{\partial}{\partial\lambda}\left[\frac{\mu_*(d+1;\lambda)}{\lambda}\right] = \frac{\lambda\ \frac{\partial}{\partial\lambda}\mu_*(d+1;\lambda) \ - \ \mu_*(d+1;\lambda)}{\lambda^2}$$

$$= \left[V_*(x) \ - \ \mu_*(d+1;\lambda)\right]\cdot\frac{1}{\lambda^2}$$

where $V_*(x)$ is the variance of the distribution which is truncated on the LEFT at d+1, [and 'given' in (215)].

Thus equation (373) becomes

$$\frac{\partial^2 \log L}{\partial\lambda^2} = -\frac{\sum_{i=1}^{K} x_i}{\lambda^2} + \frac{n_t}{\lambda^2}\left[V_*(x) \ - \ \mu_*(d+1;\lambda)\right] \qquad (374)$$

Now $\mathcal{E}(n_t) = N.P(d+1;\lambda)$ and we may obtain $\mathcal{E}(\sum_{i=1}^{K} x_i)$ as follows.

$$\mathcal{E}(\sum_{i=1}^{K} x_i) = \mathcal{E}(\sum_{x=0}^{d} x\ n_x) = N\sum_{x=0}^{d} x\ .\ \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= N\lambda\ .\ F(d-1;\lambda)$$

It follows that

$$-\mathcal{E}\left(\frac{\partial^2 \log L}{\partial \lambda^2}\right) = \frac{N\lambda . F(d-1;\lambda)}{\lambda^2} - \frac{N.P(d+1;\lambda)}{\lambda^2}\left[V_*(x) - \mu_*(d+1;\lambda)\right]$$

$$= \frac{N}{\lambda^2}\left[\lambda[\,1 - P(d;\lambda)] - P(d+1;\lambda)\, V_*(x) + \lambda P(d;\lambda)\right]$$

and the asymptotic variance of the m.l.e. $\hat{\lambda}$ is then

$$V(\hat{\lambda}) = \frac{\lambda^2}{N} \cdot \frac{1}{\left[\lambda - P(d+1;\lambda)\, V_*(x)\right]}$$

which may be computed using (215) and Patil's (1959) table.

3.6.4. One estimate that has been proposed for λ is ;[61]

$$Y = \frac{\sum_{i=1}^{K} x_i + (d+1)\, n_t}{N}$$

which is very simple to compute and might prove useful when the magnitude of the bias is small. Y could be a first approximation for an iterative procedure when solving (372) for $\hat{\lambda}$. Now,

$$\mathcal{E}(Y) = \frac{1}{N} \cdot [\, N\lambda . F(d-1;\lambda) + (d+1)\, N.P(d+1;\lambda)]$$

The quantity b = $\dfrac{\mathcal{E}(Y) - \lambda}{\lambda}$ is called the relative bias of the estimate Y and is equal to

$$- \left[\lambda P(d;\lambda) + (d+1)\, P(d+1;\lambda)\right] / \lambda$$

$$= \frac{-\, P(d+1;\lambda)\, [\, \mu_*(d+1;\lambda) - (d+1)\,]}{\lambda}$$

A table of values and charts of the relative bias of Y have been prepared by Patil (1959), for suitably chosen λ and d.

The methods of the previous sections are now clear and only a brief discussion will be given of the remaining cases of censorship.

3.7 Single censorship on the left

A random sample of size N is drawn from the distribution $P_o(\lambda)$. Let n° be the number of observations $x < c$ that are pooled in the censored portion of the sample. Let K be the number of measured observations $x_1, \ldots x_K$ for which each $x_i \geqslant c$; $i = 1, \ldots K$. Then $N = n^\circ + K$. The likelihood function L is

$$L = \binom{n^\circ+K}{K} [F(c-1;\lambda)]^{n^\circ} \frac{e^{-K\lambda} \lambda^{\sum_{i=1}^{K} x_i}}{\prod_{i=1}^{K} x_i!}$$

and

$$\frac{1}{K} \cdot \frac{\partial \log L}{\partial \lambda} = \frac{\bar{x}}{\lambda} - 1 - \frac{n^\circ}{K} \cdot \frac{p(c-1;\lambda)}{F(c-1;\lambda)}$$

where $\bar{x} = \sum_{i=1}^{K} x_i / K$

The m.l.e. for λ is the solution of the equation

$$\bar{x} = \lambda \left[1 + \frac{n^\circ}{K} \cdot \frac{p(c-1;\lambda)}{F(c-1;\lambda)}\right]$$

which, from (207) can be written in the form

$$K\bar{x} = N\lambda - n^\circ \cdot \mu^*(c-1;\lambda) \tag{375}$$

The asymptotic variance of $\hat{\lambda}$ reduces to

$$V(\hat{\lambda}) = \frac{\lambda^2}{N} \cdot \frac{1}{[\ \lambda - F(c-1;\lambda)\ V^*(x)\ ]} \tag{376}$$

where $V^*(x)$ is the variance of the distribution singly truncated on the RIGHT at c - 1. We use equation (209) for the calculation of $V(\hat{\lambda})$.

Note that when c = 1 the number n° of "pooled" observations is really the number of zero counts, and the sample is complete. Then, since $\mu^*(0;\lambda)$ is zero (375) becomes $\sum_i x_i = N\lambda$, where $\sum_i x_i$ includes the zero counts, i.e. $\hat{\lambda}$ is the sample mean as given by (331) for the complete case. Also, when c = 1, (376) reduces to $V(\hat{\lambda}) = \frac{\lambda}{N}$ which agrees with (333).

3.8 The doubly censored case

This case has been investigated by Cohen (1954) from the point of view of maximum likelihood.

Let n°, n_t be the number of left censored and right censored observations respectively. Let K be the number of measured observations for which $c \leq x \leq d$. Then the total sample size $N = n^\circ + K + n_t$

The likelihood function L takes the form

$$L = C.[F(c-1;\lambda)]^{n^\circ} . \frac{e^{-K\lambda} \quad \lambda^{\sum_{i=1}^{K} x_i}}{\prod_{i=1}^{K} x_i!} . [P(d+1;\lambda)]^{n_t}$$

where C is a constant. The m.l.e. for λ is the solution of the equation

$$\bar{x} = \lambda\left[1 + \frac{n^\circ}{K} \cdot \frac{p(c-1;\lambda)}{F(c-1;\lambda)} - \frac{n_t}{K} \cdot \frac{p(d;\lambda)}{P(d+1;\lambda)}\right]$$

which we will write in the form

$$K\bar{x} = N\lambda - n^\circ . \mu^*(c-1;\lambda) - n_t . \mu_*(d+1;\lambda)$$

where all symbols take the same meaning as in previous sections.

CHAPTER FOUR: INTERVAL ESTIMATION

4.1 Introduction

4.1.1. In this chapter we are concerned with the problem of providing a confidence interval for the single parameter of the Poisson distribution. Some fundamentals of the theory of confidence intervals due to Neyman are mentioned below; but for the single parameter case. Reference has been made to Cramér (1946), Kendall (1951) and Mood (1963).

Suppose $(y_1, \ldots y_N)$ is a random sample from the distribution with p.d.f. $f(y;\theta)$. We assume that the single parameter θ takes some constant value in the parameter space Ω, though the actual value taken is unknown.

Let $t(y_1, \ldots y_N) = t$ be an estimate of θ, and denote the p.d.f. of t by $g(t;\theta)$. Now for fixed θ the unit of probability mass associated with the distribution of t may be thought of as lying along the line $t = \theta$ in the plane of t and θ (here we take the axes of t and θ as the horizontal and vertical rectangular coordinate axes respectively).

If the real number α be given, $0 < \alpha < 1$, we can find two reals β and γ such that

$$\text{Prob}[\beta < t < \gamma;\theta] = \int_{\beta}^{\gamma} g(t;\theta)\, dt = 1 - \alpha \tag{401}$$

In fact, many such β and γ can be found for they need only satisfy the equations

$$\int_{-\infty}^{\beta} g(t;\theta)\, dt = \alpha_1 \quad \text{and} \quad \int_{\gamma}^{\infty} g(t;\theta)\, dt = \alpha_2 \tag{402}$$

where $\alpha_1 + \alpha_2 = \alpha$ i.e. the total probability mass in the two "tails" is α.

However, when the probability mass on the line $t = \theta$ is situated in discrete mass points, β and γ cannot always be found so that (401) is satisfied. But certainly we can find β and γ satisfying

$$\text{Prob}[\ \beta < t < \gamma\ ;\theta] \geqslant 1 - \alpha \tag{403}$$

Now β and γ depend on θ and α. With α given and θ taking different values in Ω we see that the points (β,θ) and (γ,θ) map out two curves (assumed to be monotonic in θ) which determine the boundary of a region $R(\alpha)$ in the (t,θ) plane.

But the boundary of the region $R(\alpha)$ is determined equivalently in another way. The estimate t will take different values from sample to sample. For a given value of t, say t' (and the same value of α as before) the boundary points of the region $R(\alpha)$ will be $\underline{\theta}$ and $\bar{\theta}$ say, which depend only on α and t. From (402)

$\underline{\theta}(t;\alpha)$ is the value of θ for which $\displaystyle\int_{t'}^{\infty} g(t;\theta)\,dt = \alpha_2$ (404)

$\bar{\theta}(t;\alpha)$ is the value of θ for which $\displaystyle\int_{-\infty}^{t'} g(t;\theta)\,dt = \alpha_1$ (405)

The events $\beta(\theta;\alpha) < t < \gamma(\theta;\alpha)$ and $\underline{\theta}(t;\alpha) < \theta < \bar{\theta}(t;\alpha)$ are equivalent and have the same associated probability. From (403) we have for the discrete case

$$\text{Prob}[\,\underline{\theta}(t;\alpha) < \theta < \bar{\theta}(t;\alpha)\,;\theta] \geqslant 1 - \alpha \qquad (406)$$

as given by Neyman (1935). Equation (406) is demonstrated for the Poisson distribution by Pearson and Hartley (1954).

Now $\underline{\theta}$ and $\bar{\theta}$ are random variables although θ is not; and we interpret (406) in the following way. For any sample values observed we say the probability that the variable interval $(\underline{\theta},\bar{\theta})$ covers the fixed point θ is at least equal to $1 - \alpha$. Or, we can say that the probability of being correct in stating that θ belongs to $(\underline{\theta},\bar{\theta})$ is at least equal to $1 - \alpha$.

The region $R(\alpha)$ is called the confidence region corresponding to a confidence coefficient $1 - \alpha$; and the interval $(\underline{\theta},\bar{\theta})$ is a $1 - \alpha$ confidence interval for the parameter θ. The quantities $\underline{\theta}$ and $\bar{\theta}$ are the lower and upper confidence limits respectively.

4.1.2. For large samples we can make use of the limiting distribution of a particular function of the observations to obtain approximate confidence limits. Suppose $(y_1, \ldots y_N)$ is a random sample from the distribution with p.d.f. $f(y;\theta)$, $\theta \in \Omega$. Then the logarithm of the likelihood function is $\log L = \sum_{i=1}^{N} \log f(y_i;\theta)$. The random variable $\frac{\partial}{\partial\theta}\log L$ has expected value zero, for

$$\mathcal{E}\left(\frac{\partial}{\partial\theta}\log L\right) = N.\mathcal{E}\left(\frac{\partial}{\partial\theta}\log f\right) = 0 \quad \text{from (302)}$$

Also $V[\frac{\partial}{\partial\theta}\log L] = V[\sum \frac{\partial}{\partial\theta}\log f] = N.V[\frac{\partial}{\partial\theta}\log f] = N.\tau^2$

Let $Q = \dfrac{\frac{\partial}{\partial\theta}\log L}{\sqrt{N\tau^2}}$. Then Q may be regarded as a member of a general class C of functions defined as follows;[42].

Let $h(y;\theta)$ be a function having expected value zero and such that the sum of a number of similar functions obeys the Central Limit Theorem. The class C consists of all those members w, with $w = \dfrac{\sum_{i=1}^{N} h(y_i;\theta)}{\sqrt{N.V[h(y;\theta)]}}$

Wilks (1938) has shown that under certain regularity conditions the distribution of Q approaches the standard Normal distribution as $N \to \infty$. For large samples we may obtain approximate confidence limits for θ using the limiting distribution of Q. Choose $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ in (402) with Q taking the place of t, then in the limit β and γ are the standard Normal fractiles $t_{\alpha/2}$ and $t_{1-\alpha/2}$ i.e. $\Phi(t_{\alpha/2}) = \alpha/2$. If Q is a monotonic function of θ then an inequality in terms of Q can be rearranged into an inequality in terms of θ, and hence we obtain the $1 - \alpha$ confidence interval $(\underline{\theta},\bar{\theta})$ for θ.

Wilks (1938) also proved that the confidence interval $(\underline{\theta},\bar{\theta})$ derived from Q in this manner has a certain optimum property. The optimum property is that whatever value θ may take in Ω , the limits $\underline{\theta}$ and $\bar{\theta}$ are (for large samples) closer together on the average than those computed from any other function belonging to class C, that is to say $\mathcal{E}(\bar{\theta} - \underline{\theta})$ is a minimum.

4.2 Central confidence intervals

4.2.1. Central confidence intervals are obtained by first choosing $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ to determine the points β and γ from (402). For the discrete case we likewise consider the tail probabilities

$$\text{Prob}[\, t \leqslant \beta \,;\theta] \leqslant \frac{\alpha}{2} \quad \text{and } \text{Prob}[\, t \geqslant \gamma \,;\theta] \leqslant \frac{\alpha}{2} \tag{407}$$

Then, if the range of the discrete distribution is the set of non-negative integers, we find β as the largest integer satisfying the first inequality in (407), and γ as the smallest integer satisfying the second inequality. (β,θ) and (γ,θ) plot as two boundary points of the region $R(\alpha)$, as described above. For a given value of t the central confidence interval is then found as the vertical intercept of $R(\alpha)$.

For the Poisson parameter λ upper central confidence limits were first given by Przyborowski and Wilenski (1935) for the single observation x = 0(1)50 and α = .001, .005, .01, .02, .05, .10. Lower and upper central limits have been tabulated by Garwood (1936), Ricker (1937) and Pearson and Hartley (1954).

Ricker derived the region $R(\alpha)$ by using a single observation of the Poisson random variable for the estimate t. For a given value of λ he found the boundary point β of $R(\alpha)$ as the integer satisfying the inequality

$$F(\beta;\lambda) \leq \frac{\alpha}{2} < F(\beta+1;\lambda)$$

and the boundary point γ as the integer satisfying

$$P(\gamma;\lambda) \leq \frac{\alpha}{2} < P(\gamma-1;\lambda)$$

The actual determination of β and γ was done with the aid of Soper's and Whitaker's tables of the Poisson distribution. The points $(\beta;\lambda)$ and (γ,λ) were then plotted for continuously increasing values of λ; starting at $\lambda = 0$ and proceeding by intervals of 0.1. This procedure gave a semi-infinite region $R(\alpha)$ in the upper right hand quadrant of the (x,λ) plane, with boundaries a pair of stepped lines; the steps occurring because β and γ are integers.

Having determined the region $R(\alpha)$ the confidence limits for λ were easily obtained as the end points of the vertical intercepts between the two stepped lines. Ricker's tables give the lower and upper central confidence limits for λ corresponding to $\alpha = .01$ and $.05$ for $x = 0(1)50$.

4.2.2. The procedure adopted by the other authors mentioned above was somewhat different, and in order to understand their approach we need to consider the following relationship between the Poisson, gamma and chi-square distributions.

In section 1.4 we obtained the relations

$$F(k;\lambda) = 1 - \int_0^{\lambda} \frac{e^{-t}.t^k dt}{k!} = \int_{\lambda}^{\infty} \frac{e^{-t}.t^k dt}{\Gamma(k+1)} \qquad (408)$$

Now Pearson's (1922) tables give values of the function

$$I(u,p) = \int_0^{u\sqrt{p+1}} \frac{e^{-t}.t^p\, dt}{\Gamma(p+1)}$$

for the arguments u and p. So that entering these tables with $u = \dfrac{\lambda}{\sqrt{K+1}}$ and writing k for p we have from (408) that

$$F(k;\lambda) = 1 - I\left(\frac{\lambda}{\sqrt{K+1}}, k\right) \tag{409}$$

and hence the Poisson c.d.f. may be evaluated with the aid of these tables. Again from (408)

$$F(k;\lambda) = \int_{\lambda}^{\infty} \frac{e^{-t}.t^{(k+1)-1}}{\Gamma(K+1)}\, dt = \text{Prob}[\ t \geqslant \lambda\ ] \tag{410}$$

where t has the gamma distribution G(k+1). As mentioned by Satterthwaite (1957) equation (410) may be obtained from repeated integration by parts of $\dfrac{1}{\Gamma(k+1)} \cdot \int_{\lambda}^{\infty} e^{-t}.t^{k}\, dt$

Making the transformation $\chi^2 = 2t$ in (410) we immediately see that

$$F(k;\lambda) = \int_{2\lambda}^{\infty} \frac{e^{-\frac{1}{2}\chi^2} \cdot (\chi^2)^{\frac{2(K+1)-2}{2}}}{2^{\frac{2(K+1)}{2}} \cdot \Gamma\left(\frac{2(K+1)}{2}\right)}\, d\chi^2 = \text{Prob}[\chi^2_{2(K+1)} \geqslant 2\lambda\ ] \tag{411}$$

and hence tables of the cumulative chi-square distribution may be used to evaluate the Poisson c.d.f.

It is now clear that the Poisson c.d.f., the incomplete gamma function and the χ^2 integral are all different forms of the same mathematical function; which remark has been made by Pearson and Hartley (1950).

Suppose a single observation of the Poisson variable yielded the value x = k. We may write the variables

$\underline{\theta}(t,\alpha)$ and $\overline{\theta}(t,\alpha)$ appearing in (404) and (405) as $\underline{\lambda}(k,\alpha)$ and $\overline{\lambda}(k,\alpha)$, or more simply $\underline{\lambda}$ and $\overline{\lambda}$ respectively. Corresponding to the desired confidence coefficient $1 - \alpha$, ($\alpha < \frac{1}{2}$) and the observed value k we seek the boundary points $\underline{\lambda}$ and $\overline{\lambda}$ of the confidence region $R(\alpha)$.

From (404) and (405) we have that $\underline{\lambda}$ is found by solving for λ the equation

$$P(k;\lambda) = \frac{\alpha}{2} \tag{412}$$

and $\overline{\lambda}$ is found by solving

$$F(k;\lambda) = \frac{\alpha}{2} \tag{413}$$

It follows from (411) that

$$2\overline{\lambda} = \chi'^2 \text{, where } \int_{\chi'^2}^{\infty} f(\chi^2_{2(K+1)})\, d\chi^2 = \frac{\alpha}{2} \tag{414}$$

Now $P(k;\lambda) = 1 - F(k-1;\lambda)$ and thus

$$2\underline{\lambda} = \chi''^2 \text{, where } \int_{\chi''^2}^{\infty} f(\chi^2_{2K})\, d\chi^2 = 1 - \frac{\alpha}{2} \tag{415}$$

If $\underline{\lambda}$ and $\overline{\lambda}$ are found for observed values of x = 0, 1, 2 .. and their values plotted as ordinates against x as abscissa the boundaries of the region $R(\alpha)$ are obtained.

However, the procedure for determination of the confidence limits $\underline{\lambda}$ and $\overline{\lambda}$ using the tables of the incomplete gamma function is not so elegant. From (409) we see that $\overline{\lambda}$ is the solution of the equation

$$I\left(\frac{\lambda}{\sqrt{K+1}}, k\right) = 1 - \frac{\alpha}{2}$$

and is found by inverse interpolation in the tables.

Similarly, $\underline{\lambda}$ is obtained as the solution of

$$I\left(\frac{\lambda}{\sqrt{K+2}}, k+1\right) = \frac{\alpha}{2}$$

by inverse interpolation.

Garwood's tables give the lower and upper central confidence limits for λ corresponding to α = .01 and .05 for the single observation x = 0(1)20(5)50, calculated to two decimal places. For values of x up to 15 he used Fisher's (1934) χ^2 table and for the remaining values of x the incomplete gamma function tables were used. The values of $\bar{\lambda}$ agree with those obtained previously by Przyborowski and Wilenski. The agreement with the limits given by Ricker is not exact; but only minor differences exist. Pearson and Hartley give $\underline{\lambda}$ and $\bar{\lambda}$ corresponding to α = .002, .01, .02, .05 and .10 for x = 0(1)30(5)50.

4.3 Methods of approximation and large sample methods

4.3.1. The tables of confidence limits referred to above extend only as far as x = 50. The incomplete gamma function tables also stop at p = 50. But for an observation x > 50 confidence limits for λ may be found by approximate methods.

Two well known approximation formulae involving the chi-square distribution are mentioned here. Fisher (1934) suggests that when v, the number of degrees of freedom, is greater than 30 then $\sqrt{2\chi^2} - \sqrt{2v - 1}$ is approximately a standard Normal variate. If t_β denotes the standard Normal fractile i.e. $\Phi(t_\beta) = \beta$, and χ^2_β has a similar meaning; then the approximation becomes

$$\chi^2_\beta \doteq \frac{1}{2}[\, t_\beta + \sqrt{2v - 1}\,]^2 \quad ; v > 30 \qquad (416)$$

However, (416) is not a very accurate approximation for small or large values of β, even when v is large;[32].

Wilson and Hilferty (1931) have suggested that the variable $\left(\frac{\chi^2}{v}\right)^{1/3}$ is approximately Normally distributed with mean $1 - \frac{2}{9v}$ and variance $\frac{2}{9v}$; and hence

$$\chi^2_\beta \doteq v(\, t_\beta \cdot \sqrt{\frac{2}{9v}} + 1 - \frac{2}{9v}\,)^3 \qquad (417)$$

To test the error in the approximations (416) and (417) for the purpose of providing confidence limits for the Poisson parameter, Garwood (1936) calculated these limits for the single observation x = 20(10)50. He compared these results with the true values calculated from the incomplete gamma function tables and showed that (417) is the more accurate approximation.

4.3.2. Tables of the Normal c.d.f. $\Phi(t)$ are readily available, and can be used to approximate the Poisson c.d.f. and hence to provide approximate confidence limits for λ. There are two commonly employed Normal approximation formulae.

From the theorem of section 1.15 it follows immediately that when λ is large

$$F(k;\lambda) \doteq \Phi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) \qquad (418)$$

applying the correction for continuity. As a practical matter the approximation is good for $\lambda > 9$;[32]. A large observed value k of x, say $k > 50$, certainly suggests the approximation is valid if this observation be considered as an estimate of its expected value λ. Corresponding to the confidence coefficient $1 - \alpha$ we have from (413) and (418) that $\bar{\lambda}$ is approximated by the solution of the equation

$$k = \lambda - \frac{1}{2} + t_{\alpha/2} \cdot \sqrt{\lambda} \qquad (419)$$

or the equation $\quad (\sqrt{\lambda})^2 + t_{\alpha/2} \cdot \sqrt{\lambda} - (k + \frac{1}{2}) = 0$

and hence

$$\sqrt{\lambda} = [\, - t_{\alpha/2} + \sqrt{t^2_{\alpha/2} + 4(k + \tfrac{1}{2})}\,] \,/\, 2 \qquad (420)$$

the negative sign for the square root is not allowed since $\sqrt{\lambda} > 0$. Squaring (420) we obtain the solution

$$\bar{\lambda} = (k + \tfrac{1}{2}) + \frac{t^2_{\alpha/2}}{4} - t_{\alpha/2} \cdot \sqrt{(k + \tfrac{1}{2}) + \frac{t^2_{\alpha/2}}{4}} \qquad (421)$$

Since $P(k;\lambda) = 1 - F(k-1;\lambda)$ we have from (412) and (418) that $\underline{\lambda}$ is approximated by the solution of the equation

$$k = \lambda + \frac{1}{2} + t_{1-\alpha/2} \cdot \sqrt{\lambda} \qquad \text{which is}$$

$$\underline{\lambda} = (k - \tfrac{1}{2}) + \frac{t^2_{1-\alpha/2}}{2} - t_{1-\alpha/2} \cdot \sqrt{(k - \tfrac{1}{2}) + \frac{t^2_{1-\alpha/2}}{4}} \qquad (422)$$

Taking $1 - \alpha = .95$ for example, we have $\frac{\alpha}{2} = .025$, $t_{\alpha/2} = -1.960$ and $t_{1-\alpha/2} = +1.960$. The approximate 95% confidence interval for λ when the observed value of x is k becomes

$$(\underline{\lambda}, \bar{\lambda}) = (k \pm \tfrac{1}{2}) + 1.921 \pm 1.960 \sqrt{(k \pm \tfrac{1}{2}) + .960} \qquad (423)$$

where the upper sign corresponds to $\bar{\lambda}$ and the lower sign to $\underline{\lambda}$.

The interval (423) was first suggested by E.S. Pearson and appeared in the paper by Ricker (1937), (though without the continuity correction).

The second commonly employed Normal approximation to the Poisson c.d.f. is arrived at from consideration of the "square root transformation". The following result is due to Curtiss (1943).

Let x have the distribution $P_o(\lambda)$. If δ is an arbitrary constant and if

$$T = \sqrt{x + \delta} \quad , x \geqslant - \delta \qquad (424)$$
$$= 0 \qquad , x < - \delta$$

then the distribution of T tends to the Normal distribution with mean $\sqrt{\lambda + \delta}$ and variance $\frac{1}{4}$ as $\lambda \to \infty$; and $\lim_{\lambda \to \infty} V[\sqrt{x + \delta}] = \frac{1}{4}$. Bartlett (1936) and (1947) has made an investigation of the degree of approximation involved in the equation $\lim_{\lambda \to \infty} V[\sqrt{x + \delta}] = \frac{1}{4}$ for values of λ from 0.5 to 15.0 in the cases $\delta = 0$ and $\frac{1}{2}$. He found that the limiting value of $\frac{1}{4}$ is much more closely approximated by $V[\sqrt{x + \frac{1}{2}}]$ than by $V[\sqrt{x}]$. Taking $\delta = 0$ in (424), the random variable $\dfrac{\sqrt{x} - \sqrt{\lambda}}{\frac{1}{2}}$ has a distribution that approaches the standard Normal distribution as $\lambda \to \infty$. Thus,

$$F(k;\lambda) \doteq \Phi[2 (\sqrt{k + \tfrac{1}{2}} - \sqrt{\lambda})] \qquad (425)$$

applying the correction for continuity. Corresponding to the confidence coefficient $1 - \alpha$ we have from (413) and (425) that $\bar{\lambda}$ is approximated by the solution of the equation

$$2 \left[\sqrt{k + \tfrac{1}{2}} - \sqrt{\lambda} \right] = t_{\alpha/2} \qquad (426)$$

which is

$$\bar{\lambda} = k + \tfrac{1}{2} + \frac{t^2_{\alpha/2}}{4} - t_{\alpha/2} \cdot \sqrt{k + \tfrac{1}{2}} \qquad (427)$$

Similarly, we obtain

$$\underline{\lambda} = k - \tfrac{1}{2} + \frac{t^2_{1-\alpha/2}}{4} - t_{1-\alpha/2} \cdot \sqrt{k - \tfrac{1}{2}} \qquad (428)$$

With $1 - \alpha = .95$, the approximate 95% confidence interval for λ, derived from (425) becomes, when the observed value of x is k

$$(\underline{\lambda}, \bar{\lambda}) = k \pm \tfrac{1}{2} + .960 \pm 1.960 \sqrt{k \pm \tfrac{1}{2}} \qquad (429)$$

where the upper sign corresponds to $\bar{\lambda}$ and the lower sign to $\underline{\lambda}$.

Hald (1952) has indicated that a better approximation to $F(k;\lambda)$ than either of (418) or (425) alone is obtained by taking their mean. Now the mean of (419) and (426) becomes

$$k = \lambda - \tfrac{1}{2} + t_{\alpha/2} \cdot \sqrt{\lambda} + t^2_{\alpha/2} / 8$$

which when solved for the upper limit $\bar{\lambda}$ yields

$$\bar{\lambda} = k + \tfrac{1}{2} + \tfrac{3}{8} t^2_{\alpha/2} - t_{\alpha/2} \cdot \sqrt{(k + \tfrac{1}{2}) + \frac{t^2_{\alpha/2}}{8}} \qquad (430)$$

Similarly, we may obtain

$$\underline{\lambda} = k - \tfrac{1}{2} + \tfrac{3}{8} t^2_{1-\alpha/2} - t_{1-\alpha/2} \cdot \sqrt{(k - \tfrac{1}{2}) + \frac{t^2_{1-\alpha/2}}{8}} \qquad (431)$$

Equations (430) and (431) are given by Crow and Gardner (1959) who acknowledge the remarkable accuracy of these equations as approximations to the central confidence limits.

The confidence limits derived thus far have been based on a single observation of the Poisson random variable x, say x = k. However, we know that if N random observations on x be made, then the distribution of their sum is also Poisson with parameter $N\lambda$. Thus the same limits provide confidence intervals for $N\lambda$, and hence for λ, by putting $\sum_{i=1}^{N} x_i = k$.

4.3.3. Suppose now that (x_1, ... x_N) is a random sample of size N from the distribution $P_o(\lambda)$. Then the random variable $Q = \dfrac{\frac{\partial}{\partial\theta}\log L}{\sqrt{N\tau^2}}$ of section 4.1.2 becomes, in view of (334) and (335), $Q = \sqrt{\dfrac{N}{\lambda}}\,[\,\bar{x} - \lambda\,]$ where $\bar{x}$ is the sample mean; and Q has a limiting Normal distribution. Thus for large samples we have, corresponding to a confidence coefficient $1 - \alpha$, that approximately

$$\text{Prob}\left[\, t_{\alpha/2} < \sqrt{\frac{N}{\lambda}}\,(\,\bar{x} - \lambda\,) < t_{1-\alpha/2} \,\right] = 1 - \alpha \qquad (432)$$

where $t_{\alpha/2}$, $t_{1-\alpha/2}$ are the standard Normal fractiles.

Taking $1 - \alpha = .95$, for example, and rearranging the inequality in (432) we see that $\underline{\lambda}$, the lower confidence limit for λ, is the solution of

$$\lambda + \frac{1.960}{\sqrt{N}}\,\sqrt{\lambda} - \bar{x} = 0$$

which is a quadratic equation in $\sqrt{\lambda}$, giving

$$\sqrt{\lambda} = \left[-\frac{1.960}{\sqrt{N}} + \sqrt{\frac{(1.960)^2}{N} + 4\,\bar{x}} \right] \Big/ 2$$

the negative sign for the square root is not allowed since $\sqrt{\lambda} > 0$. Squaring we obtain

$$\underline{\lambda} = \bar{x} + \frac{1.921}{N} - \frac{1.960}{\sqrt{N}} \cdot \sqrt{\frac{.960}{N} + \bar{x}}$$

similarly,

$$\bar{\lambda} = \bar{x} + \frac{1.921}{N} + \frac{1.960}{\sqrt{N}} \cdot \sqrt{\frac{.960}{N} + \bar{x}} \qquad (433)$$

$\underline{\lambda}$ and $\bar{\lambda}$ in (433) are equal to the 95% confidence limits for λ given by Wilks (1938) and the interval $(\underline{\lambda}, \bar{\lambda})$ possesses the optimum property referred to at the end of 4.1.2.

To the order $\frac{1}{\sqrt{N}}$ the interval $(\underline{\lambda}, \bar{\lambda})$ given in (433) becomes

$$(\underline{\lambda}, \bar{\lambda}) = \bar{x} \pm 1.960\sqrt{\frac{\bar{x}}{N}} \qquad (434)$$

Now note that if we replace the observed value k of x in (423) by $\sum_{i=1}^{N} x_i$ (and ignore the continuity correction) we obtain the interval

$$(N\underline{\lambda}, N\bar{\lambda}) = \sum_{i=1}^{N} x_i + 1.921 \pm 1.960 \cdot \sqrt{.960 + \sum_{i=1}^{N} x_i}$$

$$\text{or } (\underline{\lambda}, \bar{\lambda}) = \bar{x} + \frac{1.921}{N} + \frac{1.960}{\sqrt{N}} \cdot \sqrt{\frac{.960}{N} + \bar{x}}$$

which is the interval (433).

Again, recall from section 3.1.2 the theorem of Fisher which states that the m.l.e. $\hat{\theta}(x_1, \ldots x_N)$ is approximately Normally distributed in large samples with mean θ and variance $\frac{1}{N\tau^2}$. Hence the random variable $\sqrt{N\tau^2}(\hat{\theta} - \theta)$ has approximately the standard Normal distribution for large N. Using this fact, we have from (331) that $\hat{\theta} = \bar{x}$, and so we immediately obtain the same approximate confidence interval (433).

Thus, we have arrived, in the limit, at the same confidence interval for λ by using three different approaches to the problem. The position may be stated as follows.

In determining which of the many possible limits β and γ of equation (402) to use, we shall usually wish to make the confidence interval as small as possible. If we consider only unbiased estimates which are asymptotically Normally distributed, the confidence interval can be made as small as possible by choosing the estimate t with the smallest variance and selecting the limits β and γ such that $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$;[1].

Now the sample mean $\bar{x}$ is unbiased and asymptotically Normally distributed, and for the Poisson distribution it is the m.l.e. and the MVU estimate. Hence we may expect to arrive at the same shortest interval, in the limit, by any one of the above three methods.

4.4 Non-central confidence intervals

4.4.1. We have seen that for an observed value of the Poisson variable x, say x = k, central 1 - α confidence intervals for λ may be obtained by considering the probability of obtaining an observation which is $\geqslant$ k and an observation $\leqslant$ k. Neither of these tail probabilities are allowed to exceed $\frac{\alpha}{2}$. Suppose now that we do not require the tail probabilities to be separately $\leq \frac{\alpha}{2}$; but only that their sum be $\leq \alpha$. With this modification as starting point we may derive a non-central confidence interval for the Poisson parameter along the lines first considered by Sterne (1954) for the binomial distribution.

Suppose a single observation of the Poisson variable is taken as the estimate t of section 4.1.1. Sterne's proposal is to consider the probability of obtaining an observation as probable as or more probable than that observed in order to determine the boundary points of the confidence region $R(\alpha)$. For a given value of λ and confidence coefficient $1 - \alpha$, we seek two integers b and d, $0 \leq b \leq d$, such that

$$\text{Prob}[\, b \leq x \leq d \,] = \sum_{x=b}^{d} p(x;\lambda) \geq 1 - \alpha \qquad (435)$$

where the magnitude of each term $p(x;\lambda)$ included in this sum is not smaller than each $p(x;\lambda)$ excluded. We determine b and d by starting with the modal term and forming the partial sum by adding the terms on either side of the mode in order of decreasing magnitude until (435) is satisfied. If two different values of x have equal probabilities, i.e. if $p(x';\lambda) = p(x'';\lambda)$, $x' \neq x''$, and if both terms cannot be excluded from this sum, then both terms are included. The points (b,λ) and (d,λ) are then plotted in the (x,λ) plane.

The whole region $R(\alpha)$ is determined by proceeding continuously in this manner from small to larger values of λ, starting at $\lambda = 0$, and altering b and d as required.

It is seen that tables of Poisson terms and sums must be used, and that no help is forthcoming from tables of χ^2 or the incomplete gamma function, as was possible in the central case. The method clearly does not specify any division of probability mass between the two tails, as mentioned above. Also, by including the largest terms of the Poisson series a given probability $1 - \alpha$ is attained with as few terms as possible. Since the distribution $P_x(\lambda)$ is generally skew, and markedly so for small λ, an immediate consequence of the method is to produce a region $R(\alpha)$ that is as narrow as possible. That is, of all confidence regions with a given confidence coefficient that are determined from a single observation of the Poisson variable, the region determined as above is as narrow as possible if width is measured by the length of the horizontal intercept in the x direction; [17].

The confidence limits for λ are determined as the end points of the vertical intercept of the region $R(\alpha)$ corresponding to the observation $x = k$; as was done by Ricker for the central confidence limits.

4.4.2. Crow and Gardner (1959) have calculated the confidence limits for λ by Sterne's method corresponding to a confidence coefficient of .90 and .99 and the single observation x = 0(1)10(2)20(5)30(10)50. These limits were calculated in order to compare them with another set of confidence limits obtained by a slight modification of Sterne's technique. The modified limits follow from a certain property of partial sums of the Poisson series. This property is developed by Crow and Gardner and is discussed below.

Suppose x_1 and x_2 are two possible values of the Poisson variable, $x_1 \leq x_2$. If x_1 and x_2 are fixed, then the partial sum $\sum_{x=x_1}^{x_2} p(x;\lambda)$ is an analytic function of λ, i.e. it is single valued and differentiable in (-R,R),(see the theorem of section 1.4).

To obtain the value of λ that maximises this sum, we have that

$$\begin{aligned} \frac{\partial}{\partial\lambda}\left[\sum_{x=x_1}^{x_2} p(x;\lambda) \right] &= \frac{\partial}{\partial\lambda}\left[1 - F(x_1-1;\lambda) - P(x_2+1;\lambda) \right] \\ &= p(x_1-1;\lambda) - p(x_2;\lambda) \qquad (436) \end{aligned}$$

from (107) and (111), and the value of λ we seek is the solution of the equation

$$\frac{e^{-\lambda}\lambda^{x_1-1}}{(x_1-1)!} = \frac{e^{-\lambda}\lambda^{x_2}}{(x_2)!}$$

which is $$\lambda = \lambda_{x_1,x_2} = [\, x_1(x_1+1)\ldots x_2 \,]^{\frac{1}{x_2-x_1+1}} \qquad (437)$$

It is clear that

$$p(x_1-1;\lambda) \begin{smallmatrix} > \\ = \\ < \end{smallmatrix} p(x_2;\lambda) \quad \text{according as} \quad \lambda \begin{smallmatrix} < \\ = \\ > \end{smallmatrix} \lambda_{x_1,x_2} \qquad (438)$$

If we add the partial sum $p(x_1;\lambda) + \ldots + p(x_2-1;\lambda)$ to both sides of the first inequality in (438) we have

$$\sum_{x=x_1-1}^{x_2-1} p(x;\lambda) \quad \gtreqless \quad \sum_{x=x_1}^{x_2} p(x;\lambda) \quad \text{according as } \lambda \lesseqgtr \lambda_{x_1,x_2} \qquad (439)$$

In the accompanying chart three λ_{x_1,x_2} has been plotted against x_1 for selected values of x_2 to demonstrate the fact that if x_2 is kept fixed then λ_{x_1,x_2} increases with x_1; while if x_1 is fixed λ_{x_1,x_2} increases with x_2 also.

By altering the ranges of summation in (439) by unity we obtain

$$\sum_{x=x_1}^{x_2} p(x;\lambda) \quad \gtreqless \quad \sum_{x=x_1+1}^{x_2+1} p(x;\lambda) \quad \text{according as } \lambda \lesseqgtr \lambda_{x_1+1,x_2+1} \qquad (440)$$

and in view of the above remark $\lambda_{x_1,x_2} < \lambda_{x_1+1,x_2+1}$

By repeatedly altering the ranges of summation in (439) and combining with (440) we may obtain

$$\sum_{x_1-1}^{x_2-1} p(x;\lambda) \geqslant \sum_{x_1}^{x_2} p(x;\lambda) \geqslant \sum_{x_1+1}^{x_2+1} p(x;\lambda) \geqslant \ldots\ldots \quad \text{for } \lambda \leqslant \lambda_{x_1,x_2} \qquad (441)$$

We now deduce a sequence of inequalities in the other direction by proceeding as follows. Reducing the argument by unity in (438) we have

$$p(x_2-1;\lambda) \gtreqless p(x_1-2;\lambda) \quad \text{according as } \lambda \gtreqless \lambda_{x_1-1,\,x_2-1} \qquad (442)$$

By adding to each side of the first inequality in (442) the partial sum $p(x_1-1;\lambda) + p(x_1;\lambda) + \ldots + p(x_2-2;\lambda)$ we obtain

$$\sum_{x_1-1}^{x_2-1} p(x;\lambda) \quad \gtreqless \quad \sum_{x_1-2}^{x_2-2} p(x;\lambda) \quad \text{according as } \lambda \gtreqless \lambda_{x_1-1,x_2-1} \qquad (443)$$

and in an exactly similar manner to that described above, we may obtain the sequence of inequalities

$$\sum_{x_1-1}^{x_2-1} p(x;\lambda) \geq \sum_{x_1-2}^{x_2-2} p(x;\lambda) \geq \sum_{x_1-3}^{x_2-3} p(x;\lambda) \geq \ldots\ldots \quad \text{for} \quad \lambda \geq \lambda_{x_1-1,\, x_2-1} \tag{444}$$

Comparing (441) with (444) and noting that $\lambda_{x_1-1,\, x_2-1} \dot{<} \lambda_{x_1,x_2}$ we see that both sequences of inequalities hold simultaneously for λ in the interval $[\lambda_{x_1-1,\, x_2-1},\ \lambda_{x_1,\, x_2}]$.

Thus, if the number of consecutive terms to be included in a partial sum of the Poisson series is fixed, at say r + 1 terms, then

$$\operatorname*{Max}_{x_0 = 0,1,2..} \sum_{x=x_0}^{x_0+r} p(x;\lambda) = \sum_{x=x_1}^{x_1+r} p(x;\lambda), \quad \text{when}$$

$$\lambda_{x_1,\, x_1+r} \leq \lambda \leq \lambda_{x_1+1,\, x_1+r+1} \quad \text{where } x_1 = 0,1,2.. \text{ and } r = 0,1,2..$$

and <u>these maximum sums form a non-increasing function of λ</u> .

This last statement can be made quite straightforward by referring to chart four, which is simply the ordinates of the distribution $P_0(\lambda)$ plotted for a few values of λ. If, for each λ , the r + 1 greatest ordinates are added together then the magnitude of this sum clearly cannot increase with increasing λ .

Now suppose that for a given $\lambda = \lambda_0$ say, and confidence coefficient $1 - \alpha$ the boundary points of $R(\alpha)$ are determined as the b and d of equation (435), where d = b+r. Now $\sum_{b}^{d} p(x;\lambda)$ contains the largest terms of the Poisson

CHART THREE

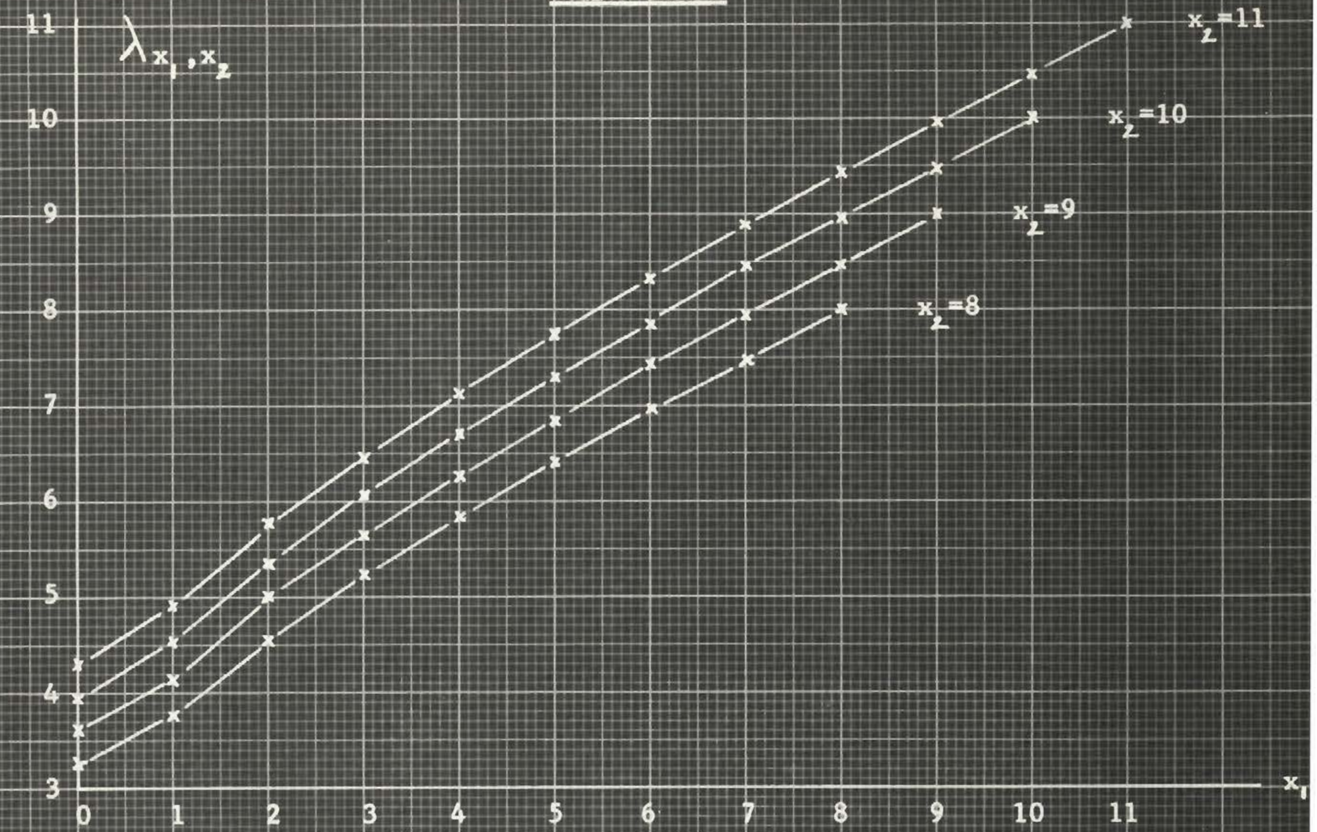

CHART FOUR

series so that no partial sum of fewer than r + 1 terms is as large as $1 - \alpha$ for the given value λ_0. Also, (because of the underlined result above) no partial sum of fewer than r + 1 terms is as large as $1 - \alpha$ for any $\lambda > \lambda_0$. This implies that the horizontal width of the region $R(\alpha)$ cannot decrease as λ increases.

Referring to (440) and writing b for x_1 and d for x_2 we have

$$\sum_{x=b}^{d} p(x;\lambda) = \sum_{x=b+1}^{d+1} p(x;\lambda) \quad \text{when } \lambda = \lambda_{b+1,d+1} \tag{445}$$

Thus when λ has increased to the value $\lambda_{b+1,d+1}$ we are able to maintain the same width of $R(\alpha)$ by substituting d+1 for b in $\sum_{b}^{d} p(x;\lambda)$. Substitution at this moment will distinguish Sterne's confidence region from Crow's.

Since Sterne includes only the largest terms in the sum $\sum_{b}^{d} p(x;\lambda)$ substitution as above must wait until λ has increased to $\lambda = \lambda_{b+1,d+1}$, when the two terms $p(b;\lambda)$ and $p(d+1;\lambda)$ are equal; [16].

However, we may perform this same substitution as soon as $\sum_{b+1}^{d+1} p(x;\lambda)$ first attains $1 - \alpha$, and this must occur at $\lambda < \lambda_{b+1,d+1}$ (recall the underlined result above). The effect of this earlier substitution is to minimise the upper confidence limit.

The confidence region obtained by Crow has the same width as Sterne's and thus has the optimum property discussed in 4.4.1. But Crow's region has the additional optimum property that of all equally narrow confidence regions of the Sterne type, it is the region with the smallest possible upper confidence limits;[17]. The confidence limits for λ, $\underline{\lambda}$ and $\bar{\lambda}$ say, are given by the end points of the vertical intercept of $R(\alpha)$. Crow has tabulated $\underline{\lambda}$ and $\bar{\lambda}$ for $1 - \alpha = .80, .90, .95, .99, .999$ and $x = 0(1)300$.

4.5 Randomised confidence intervals

4.5.1. A few fundamentals of the Neyman theory of confidence intervals were discussed in section 4.1.1. For discrete distributions the theory provides confidence intervals that contain a certain margin of safety. Equation (406) implies that it is possible to specify only the lower bound $1 - \alpha$ to our probability of being correct in making statements of the form "the interval $(\underline{\theta},\bar{\theta})$ covers the unknown parameter value." While for the continuous case, these statements can be made with the probability of being correct precisely specified.

It occurred to Stevens (1950) that the confidence limits provided by equation (406) were unnecessarily wide; and that if the discrete variable could in some way be "converted" into a continuous variable, then the limits could be narrowed until a previously specified probability is reached. Moreover, if this conversion could be achieved then further advantages would follow.

The method of interval estimation discussed in 4.4 yields confidence intervals that are indeed shorter than the central intervals. We have seen that the method does not specify any division of probability between the two tails. Consequently, the probability that the true parameter value lies below the lower limit and the probability that it lies above the upper limit are not known separately; although the sum of the two probabilities is known to be at most α.

Now if the discrete Poisson variable is converted into a continuous variable then we will be able to make one-sided statements about the true parameter value; and these statements can be made with precisely specified probability of being correct. The required conversion can be achieved by adding to the Poisson variable x an independently distributed variable u having the uniform distribution over the unit interval.

4.5.2. To show that $y = x + u$ is a continuous random variable we require the c.d.f. of y, $H(y)$ say.

If we let y', x', u' denote particular observed values of the variables, then $y' = x' + u'$ and

$$H(y') = \text{Prob}[\ y \leq y'\ ] = u'.p(x';\lambda) + \text{Prob}[\ x < x'\ ]$$
$$= u'.p(x';\lambda) + F(x'-1;\lambda)$$

Now $x' = [y']$ the largest integer contained in y'.

For convenience we will drop primes, and write

$$H(y) = \left\{ y - [y] \right\} p(\ [y]\ ;\lambda) \ + \ F(\ [y] - 1;\lambda)$$

Now $H(y)$ is clearly differentiable; but since $p(\ [y]\ ;\lambda)$ and $F(\ [y] -1;\lambda)$ take constant values for y non-integral, $\frac{d}{dy}H(y)$ will be discontinuous at all integral values of y. Thus $\frac{d}{dy}H(y)$ has a countable number of points of discontinuity, and any finite interval contains at most a finite number of discontinuity points.

The continuity of $H(y)$ itself remains to be examined. This is obvious for y non-integral, so consider $H(y)$ in the neighborhood of y_0 where y_0 is integer.

For $\triangle > 0$ we have

$$H(y_0 + \triangle) = [\ (y_0 + \triangle) - y_0\ ]\ p(y_0\ ;\lambda) \ + \ F(y_0 - 1;\lambda)$$

$$\text{and} \quad \lim_{\triangle \to 0} H(y_0 + \triangle) = F(y_0 - 1;\lambda)$$

while to approach y_0 from below consider

$$H(y_0 - \triangle) = [\ (y_0 - \triangle) - (y_0 - 1)\ ]\ p(y_0 - 1;\lambda) \ + \ F(y_0 - 2;\lambda)$$

$$\text{and} \quad \lim_{\triangle \to 0} H(y_0 - \triangle) = p(y_0 - 1;\lambda) \ + \ F(y_0 - 2;\lambda)$$
$$= F(y_0 - 1;\lambda) \text{ , also.}$$

Thus $H(y)$ is seen to be everywhere continuous; and y is a continuous random variable.

4.5.3. The device of adding a random observation of the variable u to the observed value of x is called "randomisation". The actual method of randomising adopted in practice will be an approximation to an ideal process;[2]. Thus we might choose a number of four digits from a table of random numbers and by placing a decimal point in front of them we are assuming that the decimal so obtained is a random observation of the uniform variable u. That any advantage whatsoever is to be gained by this device comes as something of a surprise. For as Pearson (1950) remarks, "we feel instinctively that having completed the experiment proper, the relevant information on which to reach a rational conclusion must be available without an appeal to any list of random numbers."

However, Stevens (1950) has shown that the device may be used to obtain confidence intervals for the parameter of a discrete distribution that are shorter than the central intervals. And, as mentioned above, statements that the true parameter value lies below, between or above certain limits can be made with a precisely specified probability of being correct. Stevens (1957) has also proposed an interpolatory procedure that may be used to obtain the shorter confidence intervals from tables of the central confidence intervals.

4.5.4. Other approaches to the problem of interval estimation for the Poisson parameter are on record. Walsh (1954) has considered functions of the confidence level α , which he then determined so as to render probability errors small over a wide range;[17].

Neyman's (1937) definition of the shortest system and the short unbiased system of confidence intervals; and Scheffe's (1942) further definition of the shortest unbiased system will not be discussed in any depth here. The subject is linked in an essential manner with the Neyman theory of hypothesis testing. Any worthwhile discussion would, in itself, be the basis of a thesis topic of considerable complexity, requiring an advanced mathematical as well as statistical background.

The Neyman-shortest intervals are shortest in the sense of being least likely to cover a false value of the parameter. Pratt (1961) has shown that this approach to interval estimation i.e. considering the probability of covering false values, is related to the approach that considers expected length of the confidence interval. He has proved that if there is an optimum procedure as regards including false values, it is also optimum as regards expected length, and vice versa. However, a knowledge of atomic measure theory is required by the reader in order to follow these results.

For a discrete distribution randomisation is required for the construction of the Neyman-shortest unbiased confidence set. This point has been discussed by Eudey (1949), who obtained the set for the binomial parameter, and Tocher (1950).

Blyth and Hutchinson (1961) have provided a table of the Neyman-shortest unbiased confidence intervals for the Poisson parameter. These authors have observed that the confidence limits obtained for $x > 250$ by using the Normal approximation (equations (421) and (422) without continuity correction) and the Neyman-shortest unbiased limits differ by less than 1% of the length of the interval.

. . .

LIST OF REFERENCES

1. ANDERSON, R.L. and BANCROFT, T.A. (1952) Statistical Theory in Research.
   McGraw-Hill Book Co. Inc. New York. Chapter 10.

2. ANSCOMBE, F.J. (1948) The validity of comparative experiments.
   Jour. Roy. Stat. Soc., Ser. A, 111, pp. 181-211.

3. BARTLETT, M.S. (1936) The square root transformation in the analysis of variance.
   Jour. Roy. Stat. Soc. Suppl. 3, pp. 68-78.

4. BARTLETT, M.S. (1947) The use of transformations.
   Biometrics, 3, pp. 39-52.

5. BLACKWELL, D. and GIRSHICK, M.A. (1954) Theory of Games and Statistical Decisions.
   John Wiley & Sons. New York. Chapter 3.

6. BLISS, C.I. (1948) Estimation of the mean and its error from incomplete Poisson distributions.
   Connecticut Agricultural Experiment Station, New Haven. Bulletin 513.

7. BLYTH, C.R. and HUTCHINSON, D.W. (1961) Table of Neyman-shortest unbiased confidence intervals for the Poisson parameter.
   Biometrika, 48, pp. 191-194.

8. BORTKIEWICZ, L.V. (1898) Das Gesetz der Kleinen Zahlen.

Leipzig.

9. BORTKIEWICZ, L.V. (1918) Homogeneitat und stabilitat in der statistik.

Scandinavisk Aktuarietidskrift.

10. BOWLEY, A.L. (1901) Elements of Statistics.

P.S. King & Son. London. page 302.

11. CACOULLOS, T. (1960) Two problems of estimation: Binomial variance and truncated Poisson mean.

Columbia University. New York.

12. COHEN, A.C. Jr. (1954) Estimation of the Poisson parameter from truncated samples and from censored samples.

Jour. Amer. Stat. Assoc. 49, pp. 158-168.

13. COHEN, A.C. Jr. (1960) Estimating the parameter in a conditional Poisson distribution.

Biometrics, 16, pp. 203-211.

14. COHEN, A.C. Jr. (1961) Estimating the Poisson parameter from samples that are truncated on the right.

Technometrics, 3, pp. 433-438.

15. CRAMÉR, H. (1946) Mathematical Methods of Statistics.

Princeton University Press. Chapters 33 and 34.

16. CROW, E.L. (1963) Personal communication, July.

17. CROW, E.L. and GARDNER, R.S. (1959) Confidence intervals for the expectation of a Poisson variable.
Biometrika, 46, pp. 441-453.

18. CURTISS, J.H. (1943) On transformations used in the analysis of variance.
Ann. Math. Stats., 14, pp. 107-122.

19. DAVID, F.N. (1951) Probability Theory for Statistical Methods.
Cambridge University Press. Chapters 6 and 15.

20. DAVID, F.N. and BARTON, D.E. (1962) Combinatorial Chance.
Hafner Publ. Co. New York. Chapter 3.

21. DAVID, F.N. and JOHNSON, N.L. (1950) The probability integral transformation when the variable is discontinuous.
Biometrika, 37, pp. 42-49.

22. DAVID, F.N. and JOHNSON, N.L. (1952) The truncated Poisson.
Biometrics, 8, pp. 275-285.

23. EUDEY, M.W. (1949) On the treatment of discontinuous random variables.
Statistical Lab., University of California, Berkley. Technical Report #13.

24. FELLER, W. (1957) An Introduction to Probability Theory and its Applications.
John Wiley & Sons. New York. Vol. I, 2nd edition. Chapter 4 and p. 248.

25.	FINNEY, D.J. and VARLEY, G.C.	(1955)	An example of the truncated Poisson distribution. Biometrics, 11, pp. 387-394.
26.	FISHER, R.A.	(1934)	Statistical Methods for Research Workers. Oliver & Boyd Ltd. Edinburgh. 5th edition. Table 3.
27.	FRASER, D.A.S.	(1958)	Statistics: An Introduction John Wiley & Sons. New York. Chapter 9.
28.	FREUND, J.E.	(1962)	Mathematical Statistics. Prentice-Hall, Inc. p. 159.
29.	GARWOOD, F.	(1936)	Fiducial limits for the Poisson distribution. Biometrika, 28, pp. 437-442.
30.	GENERAL ELECTRIC COMPANY, DEFENSE SYSTEMS DEPARTMENT	(1962)	Tables of the Individual and Cumulative Terms of Poisson Distribution. D.Van Nostrand Co., Inc.
31.	GUTTMAN, I.	(1958)	A note on a series solution of a problem in estimation. Biometrika, 45, pp. 565-567.
32.	HALD, A.	(1952)	Statistical Theory with Engineering Applications. John Wiley & Sons. New York. page 144 and chapter 22.
33.	HALMOS, P.R. and SAVAGE, L.J.	(1949)	Application of the Radon-Nikodym theorem to the theory of sufficient statistics. Ann. Math. Stats., 20, p. 225

34. HARTLEY, H.O. (1958) Maximum likelihood estimation from incomplete data.

Biometrics, 14, pp. 174-194.

35. HOEL, P.G. (1954) Introduction to Mathematical Statistics.

John Wiley & Sons. New York. 2nd edition, page 86.

36. HOGG, R.V. and CRAIG, A.T. (1959) Introduction to Mathematical Statistics.

The MacMillan Co. New York. Chapter 5.

37. IRWIN, J.O. (1937) The frequency distribution of the difference between two independent variates following the same Poisson distribution.

Jour. Roy. Stat. Soc. 100, Part 3, pp. 415-416

38. IRWIN, J.O. (1959) On the estimation of the mean of a Poisson distribution from a sample with the zero class missing.

Biometrics, 15, pp. 324-326.

39. JEFFERY, R.L. (1956) Calculus.

University of Toronto Press. page 182.

40. JOHNSON, N.L. (1959) On an extension of the connexion between Poisson and χ^2 distributions.

Biometrika, 46, pp. 352-363.

41. JORDAN, C. (1950) Calculus of Finite Differences.

Chelsea, New York. page 169.

42. KENDALL, M.G. (1951) The Advanced Theory of Statistics.

Hafner Publ. Co. New York.
Vol. II, 3rd edition.
Chapters 18 and 19.

43. KEYNES, J.M. (1921) A Treatise on Probability.

MacMillan & Co. Ltd. London.
page 403.

44. KITAGAWA, T. (1952) Tables of the Poisson Distribution.

Biafukan, Tokyo.

45. LANCASTER, H.O. (1949) The combination of probabilities arising from data in discrete distributions.

Biometrika, 36, pp. 370-382.

46. LIDSTONE, G.J. (1942) Notes on the Poisson frequency distribution.

Jour. Institute of Actuaries 71, pp. 284-291.

47. McKENDRICK, A.G. (1914) Studies on the theory of continuous probabilities, with special reference to its bearing on natural phenomena of a progressive nature.

Proc. London Math. Soc., 13, pp. 401-416.

48. MIKSA, F.L. Unpublished table of Stirling numbers of the second kind.

49. MOLINA, E.C. (1942) Poisson's Exponential Binomial Limit.

D.Van Nostrand Co., Inc.

50. MOOD, A.M. (1963) Introduction to the Theory of Statistics.

McGraw-Hill Book Co., Inc. 2nd edition. Chapters 8 and 11.

51. MOORE, P.G. (1952) The estimation of the Poisson parameter from a truncated distribution.

Biometrika, 39, pp. 247-251.

52. MOORE, P.G. (1954) A note on truncated Poisson distributions.

Biometrics, 10, pp. 402-406.

53. MORAN, P.A.P. (1952) A characteristic property of the Poisson distribution.

Proc. Camb. Phil. Soc. 48, Part 1, pp. 206-207.

54. MORANT, G. (1921) On random occurrences in space and time, when followed by a closed interval.

Biometrika, 13, pp. 309-337.

55. MORTARA (1912) Sulle variazioni di frequenza di alcuni fenomeni demografica rari.

Annali di Statistica, Serie 5, 4, pp. 5-81.

56. MURAKAMI, M. ASAI, A. and KAWAMURA, M. (1954) The estimation of the Poisson parameter from a truncated distribution and a censored sample.

Jour. Coll. Arts and Sciences. Chiba University. Vol. I, pp. 141-148.

57. NEYMAN, J. (1935) On the problem of confidence intervals.

Ann. Math. Stats. 6, pp. 111-116.

58. NEYMAN, J. (1937) Outline of a theory of statistical estimation based on the classical theory of probability.

Phil. Trans. Roy. Soc. London, Ser. A, 236, pp. 333-380.

59. NEYMAN, J. (1939) On a new class of "contagious" distributions, applicable in entomology and bacteriology.

Ann. Math. Stats. 8, pp. 35-57.

60. NOACK, A. (1950) A class of random variables with discrete distributions.

Ann. Math. Stats. 21, pp. 127-132.

61. PATIL, G.P. (1959) Contributions to estimation in a class of discrete distributions.

Doctoral Thesis at the University of Michigan.

62. PATIL, G.P. (1961) Asymptotic bias and variance of ratio estimates in generalized power series distributions and certain applications.

Sankhya, Ser. A, 23, pp. 269-280.

63. PATIL, G.P. (1962,a) Estimation by two-moments method for generalized power series distribution and certain applications.

Sankhya, Ser. B, 24, pp. 201-214.

64. PATIL, G.P. (1962,b) Minimum variance unbiased estimation for the generalized power series distribution and certain problems of additive number theory.

Paper read at the International Congress of Mathematicians, Stockholm.

65. PATIL, G.P. (1962,c) Maximum likelihood estimation for generalized power series distributions and its application to a truncated binomial distribution.

Biometrika, 49, pp. 227-237.

66. PEARSON, K. (1905) Skew variation, a rejoinder.

Biometrika, 4, page 208.

67. PEARSON, K. (1922) Tables of the Incomplete Gamma Function.

Cambridge University Press for the Biometrika Trustees.

68. PEARSON, E.S. (1950) On questions raised by the combination of tests based on discontinuous distributions.

Biometrika, 37, pp. 383-398.

69. PEARSON, E.S. and HARTLEY, H.O. (1950) Tables of the χ^2 integral and of the cumulative Poisson distribution.

Biometrika, 37, pp. 313-325.

70. PEARSON, E.S. and HARTLEY, H.O. (1954) Biometrika Tables for Statisticians.

Cambridge University Press.

71. PLACKETT, R.L. (1953) The truncated Poisson distribution.

Biometrics, 9, pp. 485-488.

72. POISSON, S.D. (1837) Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités.

Bachelier, Imprimeur-Libraire, Paris.
pp. 40, 41, 54, 189, 205-207.

73.	PRATT, J.W.	(1961)	Length of confidence intervals. Jour. Amer. Stat. Assoc. 56, pp. 549-567.
74.	PRZYBOROWSKI, J. and WILENSKI, H.	(1935)	Statistical principles for routine work in testing clover seed for dodder. Biometrika, 27, pp. 273-292.
75.	RAIKOV, D.A.	(1937)	On the decomposition of the Poisson law. Dokl. Akad. Nauk. USSR 14, pp. 9-12.
76.	RAMANUJAN, S.	(1927)	Collected papers of Srinivasa Ramanujan, edited by G.H. Hardy, P.V. Seshu Aiyar and B.M. Wilson. Cambridge University Press.
77.	RAO, C.R.	(1945)	Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37, pp. 81-91.
78.	RAO, C.R.	(1952)	Advanced Statistical Methods in Biometric Research. John Wiley & Sons. New York. Chapter 2.
79.	RICKER, W.E.	(1937)	The concept of confidence or fiducial limits applied to the Poisson frequency distribution. Jour. Amer. Stat. Assoc. 32, pp. 349-356.
80.	RIDER, P.R.	(1953)	Truncated Poisson distributions. Jour. Amer. Stat. Assoc. 48, pp. 826-830.

81.	ROY, J. and MITRA, S.K.	(1957)	Unbiased minimum variance estimation in a class of discrete distributions. Sankhya, 18, pp. 371-378.
82.	RUDIN, W.	(1953)	Principles of Mathematical Analysis. McGraw-Hill Book Co., Inc. New York. Chapter 8.
83.	SATTERTHWAITE, F.E.	(1957)	Binomial and Poisson confidence limits. Industrial Quality Control 13, #11, pp. 56-59.
84.	SCHEFFÉ, H.	(1942)	On the theory of testing composite hypotheses with one constraint. Ann. Math. Stats. 13, pp. 280-293.
85.	SEXTON, C.R., SEXTON, C.A. and SEXTON, J.A.	(1959)	Table errata #270. Mathematical Tables and Other Aids to Computation. 13, pp. 141-142.
86.	SOPER, H.E.	(1914)	Tables of Poisson's exponential binomial limit. Biometrika, 10, pp. 25-35.
87.	STERNE, T.E.	(1954)	Some remarks on confidence or fiducial limits. Biometrika, 41, pp. 275-278.
88.	STEVENS, W.L.	(1950)	Fiducial limits of the parameter of a discontinuous distribution. Biometrika, 37, pp. 117-129.

89. STEVENS, W.L. (1957) Shorter intervals for the parameter of the binomial and Poisson distributions.

Biometrika, 44, pp. 436-440.

90. STUDENT (1907) On the error of counting with a haemacytometer.

Biometrika, 5, pp. 351-360.

91. STUDENT (1919) An explanation of deviations from Poisson's law in practice.

Biometrika, 12, pp. 211-215.

92. SZEGÖ, G. (1928) Über einige von S. Ramanujan gestellte aufgaben.

Jour. London Math. Soc. 3, pp. 225-232.

93. TATE, R.F. and GOEN, R.L. (1958) Minimum variance unbiased estimation for the truncated Poisson distribution.

Ann. Math. Stats. 29, pp. 755-765.

94. TIPPETT, L.H.C. (1932) A modified method of counting particles.

Proc. Royal Soc., Ser. A 137, pp. 434-446.

95. TOCHER, K.D. (1950) Extension of the Neyman-Pearson theory of tests to discontinuous variates.

Biometrika, 37, pp. 130-144.

96. TUCKER, H.G. (1962) An Introduction to Probability and Mathematical Statistics.

Academic Press. New York. Chapter 8.

97. TUKEY, J. (1949) Sufficiency, truncation and selection.

Ann. Math. Stats. 20, pp. 309-310.

98. RUTHERFORD, E. and GEIGER, H. (1910) The probability variations in the distribution of α-particles; with a note by H. Bateman.

Philosophical Magazine. Ser. 6, 20, pp. 698-707.

99. WALSH, J.E. (1954) Analytic tests and confidence intervals for the mean value, probabilities, and percentage points of a Poisson distribution.

Sankhya, 14, pp. 25-38.

100. WATSON, G.N. (1929) Theorems stated by Ramanujan (v); approximations connected with e^x.

Proc. London Math. Soc. Ser. 2, 29, pp. 293-308.

101. WEATHERBURN, C.E. (1949) Mathematical Statistics.

Cambridge University Press. page 63.

102. WHITAKER, LUCY (1914) On the Poisson law of small numbers.

Biometrika, 10, pp. 36-71.

103. WHITWORTH, W.A. (1886) Choice and Chance.

Camb., Deighton, Bell & Co., 4th edition. Proposition 51.

104. WILKS, S.S. (1938) Shortest average confidence intervals from large samples.

Ann. Math. Stats. 9, pp. 166-175.

105. WILKS, S.S. (1962) Mathematical Statistics.

John Wiley & Sons. New York. Chapters 1, 5 and 6; problems 6.18, 6.19, 7.12 and 12.27.

106. WILSON, E.B. and HILFERTY, M.M. (1931) The distribution of chi-square.

Proc. Nat. Acad. Sci. Washington. 17, pp. 684-688.

107. YULE, G.U. (1910) On the distribution of deaths with age when the causes of death act cumulatively, and similar frequency distributions.

Jour. Roy. Stat. Soc. 73, pp. 26-35.

. . .