ESTIMATING THE PARAMETER IN A CONDITIONAL POISSON DISTRIBUTION*

A. CLIFFORD COHEN, JR.
The University of Georgia
Athens, Georgia, U.S.A.

1. INTRODUCTION

The problem of estimating the Poisson parameter when zero values of the random variable may not be observed has recently been the subject of considerable attention. Samples from distributions of the number of persons per residence suffering from a contagious disease and of the number of accidents per worker in a factory during specified intervals of time are of this type. Such samples have been considered by David and Johnson [3], Hartley [5], Moore [8], Plackett [9], Rider [10], Tate and Goen [11], and, while this paper was being processed for publication, by Irwin [6]. They constitute a special case of the various types of restricted Poisson samples studied earlier by this author [2]. In the author's previous paper, maximum likelihood estimating equations were derived for truncated and censored samples in both singly and doubly restricted cases. However, tables necessary for the rapid easy solution of these equations were not provided at that time.

David and Johnson, Rider, and Moore seemed primarily interested in estimates that are easy to calculate. When necessary, they appeared willing to make slight sacrifices in efficiency for the sake of easier calculations. Plackett, and Tate and Goen were further concerned with unbiased estimators. Maximum likelihood estimators were dismissed as being unsuited for ordinary practical use because of bias and the burdensome calculations involved.

Admittedly, maximum likelihood estimates are troublesome to calculate without proper tables since it is necessary to solve a somewhat complicated non-linear equation. These estimates, however, are consistent and asymptotically efficient. Except in small samples, bias appears unlikely to be a major source of trouble. Therefore when samples are large, if the labor involved in calculating maximum likelihood estimates can be sufficiently reduced, there appears to be little justification for employing any other estimator. Furthermore, one might quite appropriately question the advisability of even attempting

*Sponsored by the Office of Ordnance Research, U. S. Army.

to estimate the Poisson parameter from very small samples since sampling errors inherent in these estimates are of such magnitude as to limit the usefulness of results obtained regardless of the estimator employed.

The reason for reexamining the problem under consideration here is to provide tables and charts necessary to render calculation of maximum likelihood estimates feasible and to emphasize the requirement for large samples. Abbreviated tables with entries to three decimals were given by David and Johnson [3] and by Rider [10], but they are hardly adequate for general use.

Irwin [6] gave an explicit solution for the maximum likelihood estimating equation in the form of a Lagrange series, but convergence is slow, particularly for small values of $\bar{x}$. Because of the number of terms of the exapansion which therefore must be evaluated, calculations using his results are still troublesome. The tables and charts presented here greatly reduce the labor of computing maximum likelihood estimates in practical applications.

Hartley [5] gave a general iterative procedure for obtaining maximum likelihood estimates of parameters of any type population. However, calculations using his method are rather laborious in comparison with linear interpolation in the tables provided here, a procedure which is adequate in the case under consideration. Unless one is fortunate enough to begin with a close first approximation, calculations using Hartley's method are likely to be tedious and time consuming.

2. MAXIMUM LIKELIHOOD ESTIMATION

The conditional Poisson probability function under consideration here may be written as

$$p(x) = e^{-\lambda}\lambda^{x}/x!\,(1 - e^{-\lambda}), \qquad x = 1, 2, \cdots ; \lambda > 0. \tag{1}$$

The likelihood function for a sample consisting of n observations of random variable x, having probability function (1), may be written as

$$P(x_1\,, x_2\,, \cdots x_n\,; \lambda) = (1 - e^{-\lambda})^{-n} \prod_{i=1}^{n} (e^{-\lambda}\lambda^{x_i}/x_i!). \tag{2}$$

Taking logarithms of (2) and differentiating, we obtain

$$\partial L/\partial\lambda = -n/(1 - e^{-\lambda}) + \sum_{i=1}^{n} x_i/\lambda, \tag{3}$$

where L has been written for $1nP$. On setting (3) equal to zero, we obtain the estimating equation

$$\lambda/(1 - e^{-\lambda}) = \bar{x}, \tag{4}$$

where $\bar{x}$ is the sample mean ($\bar{x} = \sum_{i=1}^{n} x_i/n$). The required estimate, which we designate $\hat{\lambda}$ as distinguished from λ, the parameter being estimated, must then be a positive real root of (4). Incidently, equation (4) above is a special case of equation (10) of reference [2].

In order to simplify the solution of (4) for given values of the sample mean $\bar{x}$ Table 1 has been prepared with the aid of Molina's Tables of

TABLE: 1 $\bar{x} = \lambda/(1 - e^{-\lambda})$

$\bar{x}$	λ	$\bar{x}$	λ	$\bar{x}$	λ	$\bar{x}$	λ	$\bar{x}$	λ	$\bar{x}$	λ
		1.155	0.2955	1.46	0.8115	2.30	1.9836	4.30	4.2379	7.6	7.5962
1.0005	0.0010	1.160	.3046	1.47	.8273	2.35	2.0464	4.35	4.2904	7.7	7.6965
1.0010	.0020	1.165	.3137	1.48	.8430	2.40	2.1086	4.40	4.3428	7.8	7.7968
1.0015	.0030	1.170	.3227	1.49	.8586	2.45	2.1703	4.45	4.3951	7.9	7.8971
1.0020	.0040	1.175	.3317	1.50	.8742	2.50	2.2316	4.50	4.4473	8.0	7.9973
1.0025	0.0050	1.180	0.3407	1.51	0.8897	2.55	2.2924	4.55	4.4994	8.1	8.0976
1.0030	.0060	1.185	.3497	1.52	.9052	2.60	2.3527	4.60	4.5515	8.2	8.1977
1.0035	.0070	1.190	.3586	1.53	.9207	2.65	2.4126	4.65	4.6034	8.3	8.2979
1.0040	.0080	1.195	.3675	1.54	.9361	2.70	2.4721	4.70	4.6553	8.4	8.3981
1.0045	.0090	1.200	.3764	1.55	.9514	2.75	2.5312	4.75	4.7071	8.5	8.4983
1.005	0.0100	1.205	0.3853	1.56	0.9667	2.80	2.5899	4.80	4.7588	8.6	8.5984
1.010	.0199	1.210	.3942	1.57	.9819	2.85	2.6483	4.85	4.8105	8.7	8.6986
1.015	.0299	1.215	.4030	1.58	.9970	2.90	2.7063	4.90	4.8622	8.8	8.7987
1.020	.0397	1.220	.4118	1.59	1.0121	2.95	2.7640	4.95	4.9137	8.9	8.8988
1.025	.0496	1.225	.4206	1.60	1.0272	3.00	2.8214	5.00	4.9652	9.0	8.9989
1.030	0.0594	1.230	0.4294	1.61	1.0422	3.05	2.8786	5.1	5.0679	9.1	9.0990
1.035	.0692	1.235	.4381	1.62	1.0571	3.10	2.9354	5.2	5.1704	9.2	9.1991
1.040	.0790	1.240	.4468	1.63	1.0720	3.15	2.9919	5.3	5.2728	9.3	9.2992
1.045	.0887	1.245	.4555	1.64	1.0869	3.20	3.0482	5.4	5.3750	9.4	9.3992
1.050	.0984	1.250	.4642	1.65	1.1017	3.25	3.1042	5.5	5.4770	9.5	9.4993
1.055	0.1081	1.26	0.4815	1.66	1.1165	3.30	3.1600	5.6	5.5789	9.6	9.5993
1.060	.1177	1.27	.4987	1.67	1.1312	3.35	3.2156	5.7	5.6806	9.7	9.6994
1.065	.1273	1.28	.5158	1.68	1.1458	3.40	3.2709	5.8	5.7821	9.8	9.7995
1.070	.1369	1.29	.5329	1.69	1.1604	3.45	3.3259	5.9	5.8836	9.9	9.8995
1.075	.1464	1.30	.5499	1.70	1.1750	3.50	3.3808	6.0	5.9849	10.0	9.9995
1.080	0.1559	1.31	0.5668	1.71	1.1895	3.55	3.4356	6.1	6.0871	10.1	10.0996
1.085	.1654	1.32	.5836	1.72	1.2040	3.60	3.4902	6.2	6.1873	10.2	10.1996
1.090	.1749	1.33	.6003	1.73	1.2185	3.65	3.5446	6.3	6.2883	10.3	10.2997
1.095	.1843	1.34	.6170	1.74	1.2329	3.70	3.5988	6.4	6.3893	10.4	10.3997
1.100	.1937	1.35	.6335	1.75	1.2472	3.75	3.6528	6.5	6.4902	10.5	10.4997
1.105	0.2031	1.36	0.6500	1.80	1.3184	3.80	3.7067	6.6	6.5910	10.6	10.5997
1.110	.2125	1.37	.6665	1.85	1.3885	3.85	3.7604	6.7	6.6917	10.7	10.6998
1.115	.2218	1.38	.6829	1.90	1.4578	3.90	3.8140	6.8	6.7924	10.8	10.7998
1.120	.2311	1.39	.6992	1.95	1.5261	3.95	3.8674	6.9	6.8930	10.9	10.8998
1.125	.2404	1.40	.7154	2.00	1.5936	4.00	3.9207	7.0	6.9936	11.0	10.9998
1.130	0.2496	1.41	0.7316	2.05	1.6603	4.05	3.9739	7.1	7.0942	11.2	11.1998
1.125	.2588	1.42	.7477	2.10	1.7263	4.10	4.0269	7.2	7.1946	11.3	11.2999
1.140	.2680	1.43	.7637	2.15	1.7916	4.15	4.0798	7.3	7.2951	11.5	11.4999
1.145	.2772	1.44	.7797	2.20	1.8562	4.20	4.1326	7.4	7.3955	12.0	11.9999
1.150	.2864	1.45	.7956	2.25	1.9202	4.25	4.1853	7.5	7.4958	12.5	12.5000

the Poisson Function [4] and the W.P.A. Tables of the Exponential Function [12]. It is necessary only that we enter our table with the sample value $\bar{x}$ and read $\hat{\lambda}$ directly. Linear interpolation will ordinarily yield an accuracy of four (at least three) significant digits in this value.

For use when a quick solution of the estimating equation is desired

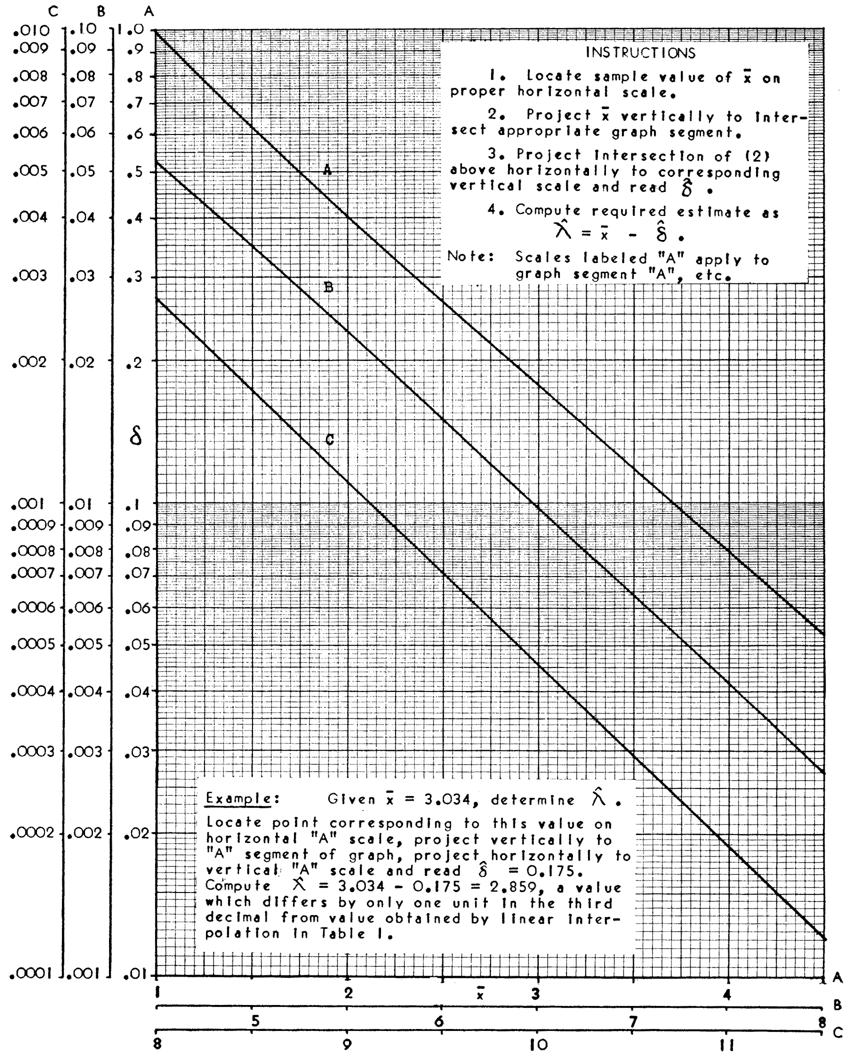

FIGURE 1.

Folded Scale δ vs $\bar{x}$ Chart and when a slight sacrifice in accuracy is permissible, a folded scale chart of δ as a function of $\bar{x}$ is given in Figure 1, where $\delta = \bar{x} - \lambda$. Thus with $\bar{x}$ given, $\hat{\delta}$ can be read from this chart and the required estimate follows as $\hat{\lambda} = \bar{x} - \hat{\delta}$. By plotting δ rather than λ, it has been possible to achieve a higher degree of accuracy in a chart of fixed size.

3. VARIANCE OF THE ESTIMATE

The asymptotic variance of $\hat{\lambda}$ may be expressed as

$$V(\hat{\lambda}) = -[E(\partial^2 L/\partial\lambda^2)]^{-1}. \tag{5}$$

The second partial derivative follows from (3) as

$$\partial^2 L/\partial\lambda^2 = -n[\{\bar{x}/\lambda^2\} - \{e^{-\lambda}/(1 - e^{-\lambda})^2\}]. \tag{6}$$

Therefore

$$V(\hat{\lambda}) \sim \psi(\lambda)[\lambda/n], \tag{7}$$

where

$$\psi(\lambda) = (1 - e^{-\lambda})^2/[1 - (\lambda + 1)e^{-\lambda}]. \tag{8}$$

We note that $\psi(\lambda)$ is continuous and monotonic decreasing, that $\lim_{\hat{\lambda}\to 0} \psi(\lambda) = 2$, and $\lim_{\hat{\lambda}\to\infty} \psi(\lambda) = 1$. Therefore, regardless of the value of λ, the asymptotic variance satisfies the inequality

$$\lambda/n \leq V(\hat{\lambda}) \leq 2\lambda/n. \tag{9}$$

TABLE 2
THE VARIANCE FUNCTION
$\psi(\lambda) = (1 - e^{-\lambda})^2/[1 - (\lambda + 1)e^{-\lambda}]$

λ	$\psi(\lambda)$	λ	$\psi(\lambda)$	λ	$\psi(\lambda)$
.001	2.000000	1.0	1.512159	7	1.005512
.1	1.935503	1.5	1.364906	8	1.002018
.2	1.875156	2.0	1.258674	9	1.000988
.3	1.818676	2.5	1.182216	10	1.000409
.4	1.765808	3.0	1.127426	14.5	1.000007
.5	1.716315	3.5	1.088421		
.6	1.669964	4.0	1.060855		
.7	1.626561	4.5	1.041538		
.8	1.585911	5	1.028135		
.9	1.547833	6	1.012619		

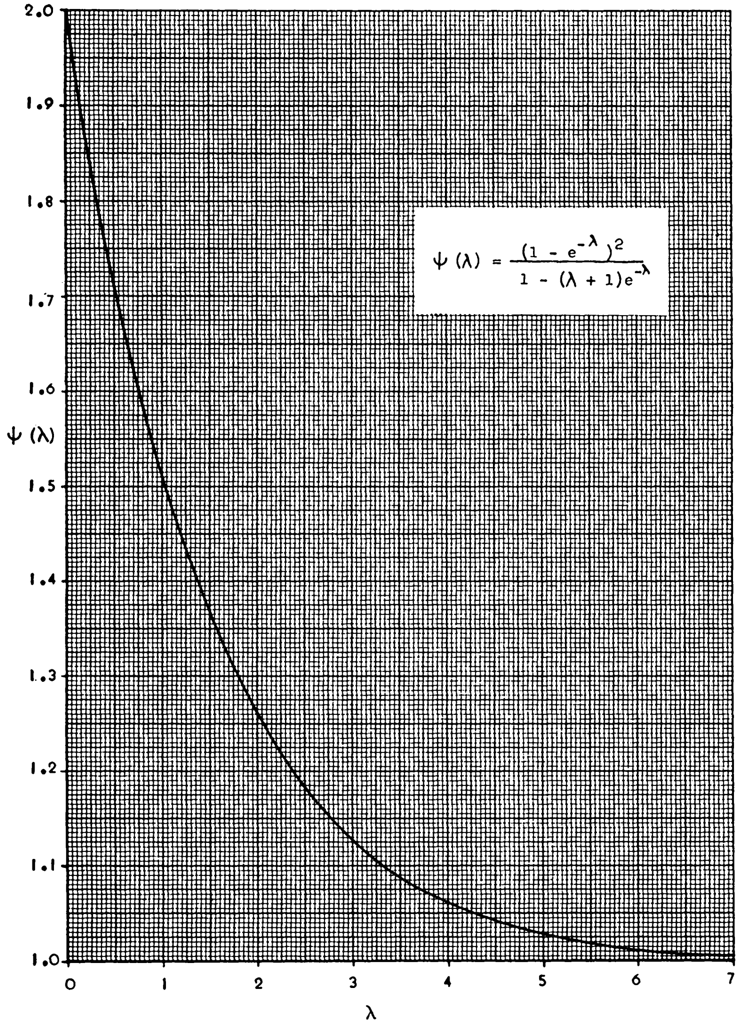

FIGURE 2.
GRAPH OF VARIANCE FUNCTION

To facilitate calculation of variances of estimates, Table 2, which is an abbreviated table of $\psi(\lambda)$, has been included. In addition, a graph of this function is given as Figure 2.

4. ILLUSTRATIVE EXAMPLES

To illustrate the practical application of these results, Bortkiewicz's [1] classic example on deaths from the kick of a horse in the Prussian

Army has been selected. These deaths were collected from records of a certain group of ten Prussian Army Corps over the twenty year period from 1875–1894. A total of 122 deaths were recorded in the 200 annual reports included in the study, and the mean number of deaths per army corps per year is $122/200 = 0.610$. On the basis of the full (unrestricted) sample, this is the estimate of λ. Following is a tabulation of the complete data for this example.

No. deaths per Army Corps per year	No. observations
x	f
0	109
1	65
2	22
3	3
4	1
5	0

For the purposes of this illustration, zero observations are eliminated and there remains a sample consisting of $n = 91$ observations with

$$\bar{x} = \sum xf/n = 122/91 = 1.3407,$$

which we consider as being from a conditional Poisson distribution with probability function (1). Perhaps the sample thus formed might more appropriately be considered simply as a truncated Poisson sample from which the zero class is missing. Regardless of the point of view adopted, estimating equation (4) is applicable.

Entering Table 1 with $\bar{x} = 1.3407$, we interpolate linearly as summarized below and round off to three decimals to obtain $\hat{\lambda} = 0.618$, which is to be compared with the complete sample estimate of 0.610.

$\bar{x}$	λ
1.3500	0.6335
1.3407	*0.6182*
1.3400	0.6170

The chart of Figure 1 can be employed to read $\hat{\delta} = 0.72$, from which it follows that $\hat{\lambda} = 1.34 - 0.72 = 0.62$, a value that is correct to the two decimals given.

We employ (7) with λ replaced by its estimate 0.618 in calculating the variance of $\hat{\lambda}$. From the chart of Figure 2, we read $\psi(0.618) = 1.66$ and accordingly

$$V(\hat{\lambda}) \doteq 1.66[0.618/91] = 0.0113,$$

$$\sigma_{\hat{\lambda}} = \sqrt{V(\hat{\lambda})} \doteq 0.106.$$

For comparison, we compute the standard deviation of $\hat{\lambda}$ based on the full unrestricted sample of 200 observations, and thus obtain

$$\sigma_{\hat{\lambda}} \doteq \sqrt{\hat{\lambda}/n} = \sqrt{0.610/200} = 0.055,$$

a value which represents a reduction of almost one-half from the standard deviation of the estimate based only on the 91 non-zero observations. The comparison afforded here serves to emphasize the necessity for large samples if reliable estimates of the Poisson parameter are to be obtained when the zero class is missing.

For an additional illustration, we consider the distribution of eggs laid in the unopened flower heads of the black knapweed by the Knapweed gall-fly in two different years, 1935 and 1936, studied by Finney and Varley [4]. The number of flower heads in which no eggs were laid is unavailable, and the 1935 data consisted of 148 observations for which the number of eggs is one or more with $\bar{x} = 3.020$. Linear interpolation in Table 1 immediately yields $\hat{\lambda} = 2.8443$ as compared with 2.845 given by Finney and Varley. More accurate calculations using entries from Tables of the Exponential Function [12] subsequently verified the interpolated value given here to be correct. The chart of Figure 1 can be employed to read $\hat{\delta} = 0.18$, from which it follows that $\hat{\lambda} = 3.02 - 0.18 = 2.84$, a result correct to two decimals.

The 1936 data consisted of 88 flower heads with $\bar{x} = 3.034$. Linear interpolation in Table 1 gives $\hat{\lambda} = 2.8603$ which, when rounded off to three decimals, agrees exactly with the corresponding estimate calculated by Finney and Varley. Again the chart of Figure 1 can be employed to read $\hat{\delta} = 0.175$, from which it follows that $\hat{\lambda} = 3.034 - 0.175 = 2.859$, a result correct to within one unit in the third decimal.

REFERENCES

[1] Bortkiewicz, L. von. [1898]. *Das Gesetz der Kleinen Zahlen*, Teubner, Leipzig.
[2] Cohen, A. C., Jr. [1954]. Estimation of the Poisson parameter from truncated samples and from censored samples, *J. Am. Stat. Assn. 49*, 158–68.
[3] David, F. N., and Johnson, N. L. [1952]. The truncated Poisson. *Biometrics 8*, 275–85
[4] Finney, D. J. and Varley, G. C. [1955]. An example of the truncated Poisson distribution, *Biometrics II*, 387–94.

[5] Hartley, H. O. [1958]. Maximum likelihood estimation from incomplete data. *Biometrics 14*, 174–94.
[6] Irwin, J. O. [1959]. On the estimation of the mean of a Poisson distribution from a sample with the zero class missing. *Biometrics 15*, 324–26.
[7] Molina, E. C. [1942]. *Poisson's Exponential Binomial Limit*, D. Van Nostrand Co., Inc.
[8] Moore, P. G. [1952]. The estimation of the Poisson parameter from a truncated distribution. *Biometrics 39*, 247–51.
[9] Plackett, R. L. [1953]. The truncated Poisson distribution. *Biometrics 9*, 485–88.
[10] Rider, Paul R. [1953]. Truncated Poisson distributions. *J. Am. Stat. Assn. 48*, 826–30.
[11] Tate, R. F., and Goen, R. L. [1958]. Minimum variance unbaised estimation for the truncated Poisson distribution. *Ann. Math. Stat. 29*, 755–65.
[12] W. P. A. [1939]. *Tables of the Exponential Function*, W. P. A. Project No. 765-97-3-10, U.S. National Bureau of Standards.