ESTIMATION OF THE MEAN AND ITS ERROR FROM INCOMPLETE POISSON DISTRIBUTIONS

C. I. Bliss

CONNECTICUT AGRICULTURAL
EXPERIMENT STATION
NEW HAVEN

in 513—January, 1948

ESTIMATION OF THE MEAN AND ITS ERROR FROM INCOMPLETE POISSON DISTRIBUTIONS

C. I. Bliss[1,2]

The biologist often needs to count organisms or events which are distributed at random in space or in time. This bulletin describes a method for estimating the average number of individuals in each counting unit when his count is not recorded completely. It is applicable to experiments in which the individuals occur at random as defined by the Poisson distribution [5]. Small units are selected for counting so that some of them contain no organism or event. The average number in all units can then be estimated with the present method even though those with four or more individuals are grouped. With the tables in the present bulletin one can estimate rapidly and easily the mean and its standard error from such incomplete series. Two types of biological data are encountered in practice to which the Poisson distribution may be applicable.

In the first type randomness is expected from the nature of the data. In counts with the haemocytometer, for example, the number of cells in each square is distributed at random when the experimental technique is adequate [5]. From a series of such counts the experimenter can estimate the concentration of cells in the original suspension. Another example is the distribution of noxious weed seeds per ounce of field seed. In the inspection of clover seed, the number of dodder seeds in 100 gram samples from different parts of the sack or from different sacks in a lot has been found to be random [9]. In cases such as these the variability among replicated random counts should follow the Poisson distribution if a single population is involved.

The distribution of most plants or animals in nature, on the other hand, is modified by non-random factors. Sometimes the latter may contribute less to the total variability than the part attributable to random or Poisson variation. This was the case, for example, in the occurrence of Japanese beetle larvae in relatively light infestations [3]. In studies on the distribution of leafhoppers in fields of sugar beets, Bowen [4] found that the mean number of leafhoppers per plant could be estimated with the Poisson series from the percentage of uninfested plants. Even in so favorable a case, however, the possible occurrence of non-random factors would need to be checked continuously in order to insure the validity of the estimate. In

[1] Station biometrician.

[2] The author is especially indebted to Miss Marian C Jackman, Technician in the Department of Pharmacology, Yale University. Miss Jackman computed Tables 1, 2 and 5 which, as the reader will soon discover, are the most useful parts of the present bulletin. The author is equally indebted to Dr. Raimon L. Beard, entomologist at this station, for the experimental data in Table 3.

ecological studies on the occurrence of grasses in nature, Blackman [2] reported several cases where counts of the number of plants per quadrat square could be described adequately by the Poisson series. Thus, even though non-random factors usually cause significant departures from a random distribution of plants and animals in nature, the Poisson series determines the minimum variability to be expected by chance alone.

When a Poisson series can be assumed, there may be advantages in truncating the distribution arbitrarily and computing the mean from an incomplete count. Units containing few individuals, such as 0, 1, 2 and 3, can be identified at a glance and the number of each type counted rapidly. The number of units with four or more individuals can be counted equally rapidly when grouped into a single category. If a small enough unit is selected that the mean number per unit does not exceed 3 to 5, the precision lost in pooling the larger occurrences can easily be balanced by counting more units. The method of computation is based upon a paper by Tippett [10], which should be consulted for a mathematical derivation of the underlying equations and for nomographs for estimating the mean.

Estimation of the Mean

A Poisson distribution is defined by a single parameter, its mean (m). Since its variance (σ^2) is equal to the mean, the expected frequencies of the units containing $x = 0, 1, 2, 3, \cdots$ events per unit are given by the successive terms in the series

$$Ne^{-m}\left\{1, m, \frac{m^2}{2!}, \frac{m^3}{3!}, \cdots\right\}, \tag{1}$$

where $e = 2.7183$ or the base of the natural logarithms and N is the total number of units. Since the parameter m is seldom known *a priori*, it is replaced by its estimate $\hat{m}$ computed from the observations. When the series is complete,

$$\hat{m} = \bar{x} = S(fx)/N \tag{2}$$

or the arithmetic mean. When the larger units are pooled into a single class, the total number of individuals in all units $S(fx)$ is unknown and the mean cannot be computed directly. It can be estimated, however, by maximum likelihood as described by Tippett [10].

A preliminary approximation to $\hat{m}$ is obtained from a proportionate cumulative frequency based upon the actual observations. For this purpose the frequencies f for $x = 0, 1, 2, \cdots$ may be added until their sum $S(f)$ is equal to about half of the total or until it includes all classes that were not pooled in counting. This number is expressed as a proportion of the total (N) and referred to the appropriate column of Table 1 to obtain the two or three values of $\hat{m}$ in the first column which bracket it. Table 1 gives for each value of the mean $\hat{m}$ the proportionate expected frequency for $x = 0$ and that accumulated from $x = 0$ to $x = 1$, 2, and 3. It has

been prepared from a table of the Poisson distribution such as Pearson's Table 51 [8] or Molina's Table 1 [7].

TABLE 1. PROPORTIONATE EXPECTED FREQUENCIES ϕ_{ot} FOR THE POISSON DISTRIBUTION CUMULATED FROM $x = 0$ TO DESIGNATED VALUES OF x.

Mean $\hat{m}$	Cumulated values of x				Mean $\hat{m}$	Cumulated values of x			
	0	0 and 1	0 to 2	0 to 3		0	0 and 1	0 to 2	0 to 3
.1	.90484	.99532	.99984	1.00000	3.1	.04505	.18470	.40116	.62484
.2	.81873	.98248	.99885	.99994	3.2	.04076	.17120	.37990	.60252
.3	.74082	.96306	.99640	.99973	3.3	.03688	.15860	.35943	.58034
.4	.67032	.93845	.99207	.99922	3.4	.03337	.14684	.33974	.55836
.5	.60653	.90980	.98561	.99825	3.5	.03020	.13589	.32085	.53663
.6	.54881	.87810	.97688	.99664	3.6	.02732	.12569	.30275	.51522
.7	.49658	.84420	.96586	.99425	3.7	.02472	.11620	.28543	.49415
.8	.44933	.80879	.95258	.99092	3.8	.02237	.10738	.26890	.47348
.9	.40657	.77248	.93714	.98654	3.9	.02024	.09918	.25312	.45325
1.0	.36788	.73576	.91970	.98101	4.0	.01832	.09158	.23810	.43347
1.1	.33287	.69903	.90042	.97426	4.1	.01657	.08452	.22381	.41418
1.2	.30119	.66263	.87949	.96623	4.2	.01500	.07798	.21024	.39540
1.3	.27253	.62682	.85711	.95690	4.3	.01357	.07191	.19736	.37715
1.4	.24660	.59183	.83350	.94628	4.4	.01228	.06630	.18514	.35945
1.5	.22313	.55782	.80885	.93436	4.5	.01111	.06110	.17358	.34230
1.6	.20190	.52493	.78336	.92119	4.6	.01005	.05629	.16264	.32571
1.7	.18268	.49325	.75722	.90681	4.7	.00910	.05184	.15230	.30968
1.8	.16530	.46284	.73062	.89129	4.8	.00823	.04773	.14254	.29423
1.9	.14957	.43375	.70372	.87470	4.9	.00745	.04394	.13333	.27934
2.0	.13534	.40601	.67668	.85712	5.0	.00674	.04043	.12465	.26503
2.1	.12246	.37962	.64963	.83864	5.1	.00610	.03719	.11648	.25127
2.2	.11080	.35457	.62271	.81935	5.2	.00552	.03420	.10879	.23807
2.3	.10026	.33085	.59604	.79935	5.3	.00499	.03145	.10155	.22541
2.4	.09072	.30844	.56971	.77872	5.4	.00452	.02891	.09476	.21329
2.5	.08208	.28730	.54381	.75758	5.5	.00409	.02656	.08838	.20170
2.6	.07427	.26738	.51843	.73600	5.6	.00370	.02441	.08239	.19062
2.7	.06721	.24866	.49362	.71409	5.7	.00335	.02242	.07677	.18005
2.8	.06081	.23108	.46945	.69194	5.8	.00303	.02059	.07151	.16996
2.9	.05502	.21459	.44596	.66962	5.9	.00274	.01890	.06658	.16035
3.0	.04979	.19915	.42319	.64723	6.0	.00248	.01735	.06197	.15120

The total number of individuals *expected* in the classes which are listed separately is determined for each provisional $\hat{m}$ by means of the expression

$$\hat{m}(f_{ot} - \eta f_c), \tag{3}$$

where f_{ot} is the observed total frequency from $x = 0$ to $x = t$, the largest class which is recorded separately, and f_c is the observed composite frequency for all values of x larger than t, so that $f_{ot} + f_c = N$. Corresponding

to each $\hat{m}$ is an expected ratio $\eta = \phi_t/\phi_c$, which is the ratio of the expected frequency for $x = t$ to the cumulated expected frequency corresponding to f_c. These η ratios are listed in Table 2 for each $\hat{m}$ from 0.1 to 6.0 and for x_t from 0 to 3.

TABLE 2. RATIO η OF EXPECTED FREQUENCY IN LARGEST CLASS (x_t) THAT IS COUNTED (ϕ_t) TO THE CUMULATED EXPECTED FREQUENCY IN ALL LARGER CLASSES (ϕ_c).

Mean $\hat{m}$	Ratio $\eta = \phi_t/\phi_c$ where x_t is				Mean $\hat{m}$	Ratio $\eta = \phi_t/\phi_c$ where x_t is			
	0	1	2	3		0	1	2	3
.1	9.50833	19.3383	29.1871	37.7500	3.1	.047174	.171291	.361469	.596218
.2	4.51666	9.34463	14.2639	19.1579	3.2	.042494	.157383	.336564	.560068
.3	2.85829	6.01699	9.26028	12.5338	3.3	.038295	.144656	.313514	.526404
.4	2.03324	4.35612	6.76583	9.21392	3.4	.034525	.132999	.292155	.495008
.5	1.54150	3.36199	5.26939	7.21233	3.5	.031137	.122312	.272338	.465689
.6	1.21637	2.70126	4.27368	5.88356	3.6	.028092	.112506	.253936	.438276
.7	.986432	2.23106	3.56344	4.93447	3.7	.025351	.103504	.236833	.412615
.8	.815966	1.87996	3.03197	4.22280	3.8	.022883	.095235	.220922	.388570
.9	.685118	1.60828	2.61961	3.67026	3.9	.020660	.087635	.206112	.366019
1.0	.581975	1.39221	2.29063	3.22904	4.0	.018658	.080649	.192316	.344848
1.1	.498960	1.21659	2.02228	2.86854	4.1	.016852	.074221	.179458	.324961
1.2	.431012	1.07132	1.79947	2.56875	4.2	.015224	.068307	.167469	.306262
1.3	.374630	.949391	1.61167	2.31563	4.3	.013756	.062866	.156284	.288673
1.4	.327310	.845820	1.45142	2.09915	4.4	.012430	.057856	.145847	.272117
1.5	.287216	.756928	1.31319	1.91205	4.5	.011234	.053243	.136104	.256526
1.6	.252971	.679972	1.19288	1.74880	4.6	.010154	.048996	.127004	.241835
1.7	.223516	.612845	1.08733	1.60520	4.7	.009178	.045085	.118505	.227987
1.8	.198033	.553906	.994078	1.47800	4.8	.008298	.041483	.110567	.214930
1.9	.175874	.501862	.911202	1.36460	4.9	.007503	.038165	.103149	.202613
2.0	.156517	.455679	.837151	1.26295	5.0	.006784	.035109	.096218	.190992
2.1	.139544	.414515	.770662	1.17139	5.1	.006134	.032294	.089741	.180024
2.2	.124610	.377682	.710718	1.08852	5.2	.005548	.029702	.083688	.169672
2.3	.111430	.344611	.656462	1.01323	5.3	.005017	.027314	.078031	.159899
2.4	.099769	.314829	.607189	.944581	5.4	.004537	.025116	.072745	.150669
2.5	.089425	.287935	.562304	.881773	5.5	.004104	.023090	.067804	.141955
2.6	.080233	.263591	.521305	.824143	5.6	.003712	.021226	.063189	.133725
2.7	.072048	.241508	.483760	.771115	5.7	.003357	.019509	.058875	.125952
2.8	.064747	.221437	.449301	.722203	5.8	.003037	.017929	.054845	.118612
2.9	.058227	.203164	.417613	.676984	5.9	.002746	.016474	.051081	.111679
3.0	.052396	.186503	.388416	.635097	6.0	.002485	.015136	.047566	.105131

We next compute $S_{ot}(fx)$, the total number of individuals *observed* in the classes which were counted separately. The corresponding expected values obtained by solving expression (3) at each selected $\hat{m}$ should bracket that computed from the observations. If not, expression (3) is solved with

additional values of $\hat{m}$ until two are obtained which enclose the observed $S_{ot}(fx)$. The mean $\hat{m}$ to which the latter corresponds is then interpolated linearly from the adjacent expected values. Within the limits imposed by linear interpolation the result is a maximum likelihood estimate of the mean of the incomplete distribution.

The computation may be illustrated by haemocytometer counts, kindly supplied by Dr. R. L. Beard [1], of blood from an infected Japanese beetle larva. The infection was milky disease and the object of the count was to determine the concentration of spores per mg. of blood. One drop weighing 22.5 mg. was dispersed in 50 cc. of water and the concentration determined by four series of counts, each of 64 squares. Each square represented a volume of 25×10^{-8} cc. of suspension or 11.25×10^{-8} mg. of blood. Squares containing four or more spores were listed under a single heading, leading to the frequency distributions in Table 3.

TABLE 3. FREQUENCY DISTRIBUTIONS OF SPORE COUNTS FROM THE BLOOD OF A JAPANESE BEETLE LARVA INFECTED WITH MILKY DISEASE. DATA OF R. L. BEARD [1]

No. of spores per square x	Frequency (f) of squares in series			
	I	II	III	IV
0	3	4	9	10
1	11	13	15	14
2	19	16	18	12
3	14	12	14	14
$\geqslant 4$	17	19	8	14
Total (N)	64	64	64	64
$S(x)$ from complete count	169	156	135	144

The calculation may be illustrated with the counts from series I. In the range from $x = 0$ to $x = 3$, the cumulated observed frequency was $f_{03} = 3 + 11 + 19 + 14 = 47$ and the total number of spores recorded was $S_{03}(fx) = 11 + 2 \times 19 + 3 \times 14 = 91$. For a trial estimate of $\hat{m}$, the frequencies were cumulated from $x = 0$ to $x = 2$ to obtain $f_{02}/N = 33/64 = 0.516$. When referred to the fourth column of Table 1, 0.516 lay between the expected proportions for means of $\hat{m} = 2.6$ and $\hat{m} = 2.7$. The expected total number of spores from $x = 0$ to $x = 3$ was then computed for each $\hat{m}$ with the values of η from the fifth column of Table 2 by expression (3) as $2.6(47 - 17 \times 0.8241) = 85.77$ and $2.7(47 - 17 \times 0.7711) = 91.51$ respectively. The number of spores observed, $S_{03}(fx) = 91$, corresponded to a mean $\hat{m}$ between 2.6 and 2.7. Its value was estimated more exactly by linear interpolation as $\hat{m} = 2.6 + 0.1\left(\frac{91 - 85.77}{91.51 - 85.77}\right) = 2.691$ spores per square. This represented a density of $2.691/11.25 \times 10^{-8} = 23.9 \times$

10^6 spores per mg. of blood. The above calculation is summarized in the following Work Form.

WORK FORM

For calculating the mean number of spores per square ($\hat{m}$) and its standard error ($s_{\hat{m}}$) from the data for series I in Table 3.

Haemocytometer counts of the number of spores of milky disease in the blood of a Japanese beetle larva

No. of spores per square x	No. of squares f		fx
0	3		0
1	11	47	11
2	19	$= f_{03}$	38
3	14		42
$\geqslant 4$	17	$= f_c$	
Total.....	64 $(= N)$		91

First approximation to $\hat{m}$

$f_{02} = 3 + 11 + 19 = 33$

$f_{02}/N = 0.516$

From Table 1, find for $x = 0$ to 2

$\hat{m}$	ϕ_{02}	f_{02}/N
2.6	.5184	
		.516
2.7	.4936	

For range from $x = 0$ *to* $x = 3$*:*

Compute expected total number of spores for $\hat{m} = 2.6$ and 2.7 with η from Table 2:

$\hat{m}(f_{03} - \eta f_c) = 2.6(47 - .8241 \times 17) = 85.77$ for $\hat{m} = 2.6$

and $2.7(47 - .7711 \times 17) = 91.51$ " $\hat{m} = 2.7$

Observed total number of spores, $S_{03}(fx) = 91$.

Interpolate $\hat{m}$ *corresponding to* 91 *as*

$$\hat{m} = 2.6 + 0.1\left\{\frac{91 - 85.77}{91.51 - 85.77}\right\} = 2.691 \text{ spores per square.}$$

Since 11.25×10^{-8} mg. of blood cover each square, each mg. of blood contains $2.691/11.25 \times 10^{-8} = 23.9 \times 10^6$ spores.

Standard error of $\hat{m}$. Interpolate $\sigma_{\hat{m}}^2$ from Table 5 for $\hat{m} = 2.691$ from values for $\hat{m} = 2.6$ and 2.7.

$\hat{m}$	Tabular $\sigma_{\hat{m}}^2$	
2.6	$2.8966 \times .09 =$	.261
2.7	$3.0457 \times .91 =$	2.772
2.691		$3.033 = \sigma_{\hat{m}}^2$

Compute $s_{\hat{m}} = \sqrt{\sigma_{\hat{m}}^2/N} = \sqrt{3.033/64} = 0.218$ spores per square.

Similar estimates have been computed from the data of the other three series and listed in Table 4. For comparative purposes the total number of spores in each series was recorded at the time of counting (Table 3). The means ($\bar{x}$) from the complete counts have been computed by equation (2) and listed beside the means ($\hat{m}$) estimated from the grouped distributions. The two estimates agreed within the sampling error in all series.

TABLE 4. STATISTICS COMPUTED FROM THE OBSERVATIONS IN TABLE 3.

Series no.	Complete count $\bar{x} \pm s_{\bar{x}}$	Partial count $\hat{m} \pm s_{\hat{m}}$	χ^2 for agreement with Poisson ($n = 3$)
I	2.641 ± .203	2.691 ± .218	1.275
II	2.438 ± .195	2.683 ± .217	0.528
III	2.109 ± .182	2.020 ± .182	1.000
IV	2.250 ± .188	2.260 ± .195	3.547

Standard Error of the Mean

Since the variance of a Poisson distribution is equal to its mean, the standard error of the mean of a complete count may be computed as

$$s_{\hat{m}} = \sqrt{m/N}, \tag{4}$$

where the expected mean (m) is replaced necessarily by its estimate $\hat{m} = \bar{x}$. This equation does not apply, however, when the upper end of the distribution is grouped, so that we know the number of units (f_c) but not the number of individuals it contains. Tippett [10] has shown that the variance ($\sigma_{\hat{m}}^2$) applying to the mean ($\hat{m}$) estimated from incomplete data is a function of the expected proportionate frequencies (ϕ). On the basis of a single unit, his equation may be written as

$$\sigma_{\hat{m}}^2 = \hat{m}/\{\phi_{ot} - \phi_{t-1}\hat{m} + \phi_t[(1 + \eta)\hat{m} - 1]\}, \tag{5}$$

where for a given $\hat{m}$, ϕ_{ot} is the cumulated proportionate frequency in Table 1, η is the ratio in Table 2 and ϕ_t and ϕ_{t-1} are the proportionate frequencies for the largest and next-to-largest values of x that are counted completely. Solving this equation for the values of $\hat{m}$ in Tables 1 and 2 led to the variances in Table 5.

The standard error of the mean $\hat{m}$ estimated from a given series of N units is computed as

$$s_{\hat{m}} = \sqrt{\sigma_{\hat{m}}^2/N}, \tag{6}$$

where $\sigma_{\hat{m}}^2$ is interpolated directly from Table 5 at the required $\hat{m}$.

Since for a complete count $\sigma_{\hat{m}}^2 = \hat{m}$, one can compute how many units would be needed to estimate the mean with equal precision from an in-

TABLE 5. VARIANCE OF THE MEAN OF AN INCOMPLETE POISSON DISTRIBUTION AS COMPUTED BY EQUATION 4.

$\hat{m}$	$\sigma_{\hat{m}}^2$ at t =				$\hat{m}$	$\sigma_{\hat{m}}^2$ at t =			
	0	1	2	3		0	1	2	3
.1	.10517	.10016	.10005	.10000	3.1	21.198	7.5156	4.6591	3.7039
.2	.22140	.20125	.20006	.20000	3.2	23.533	8.1338	4.9504	3.8863
.3	.34986	.30413	.30030	.30002	3.3	26.113	8.8016	5.2593	4.0768
.4	.49182	.40956	.40089	.40007	3.4	28.964	9.5236	5.5874	4.2760
.5	.64872	.51830	.50210	.50020	3.5	32.116	10.3040	5.9359	4.4846
.6	.82212	.63105	.60417	.60047	3.6	35.598	11.1486	6.3064	4.7030
.7	1.01376	.74851	.70742	.70096	3.7	39.446	12.0622	6.7006	4.9322
.8	1.22554	.87141	.81217	.80180	3.8	43.701	13.0512	7.1200	5.1725
.9	1.45960	1.00042	.91878	.90309	3.9	48.402	14.1227	7.5668	5.4249
1.0	1.71829	1.13630	1.02758	1.00497	4.0	53.597	15.2824	8.0428	5.6902
1.1	2.0042	1.27982	1.13898	1.10762	4.1	59.34	16.540	8.5502	5.9692
1.2	2.3201	1.43175	1.25336	1.21122	4.2	65.68	17.903	9.0912	6.2629
1.3	2.6693	1.59297	1.37114	1.31596	4.3	72.70	19.380	9.6688	6.5721
1.4	3.0552	1.76434	1.49270	1.42204	4.4	80.45	20.982	10.2854	6.8977
1.5	3.4817	1.94683	1.61852	1.52974	4.5	89.02	22.721	10.9438	7.2415
1.6	3.9530	2.1414	1.7490	1.63925	4.6	98.48	24.608	11.648	7.6038
1.7	4.4740	2.3492	1.8847	1.75085	4.7	108.95	26.656	12.400	7.9865
1.8	5.0496	2.5715	2.0260	1.86477	4.8	120.51	28.880	13.205	8.3905
1.9	5.6859	2.8094	2.1735	1.98133	4.9	133.28	31.295	14.066	8.8174
2.0	6.3890	3.0642	2.3278	2.10080	5.0	147.41	33.918	14.987	9.2686
2.1	7.1662	3.3376	2.4893	2.2235	5.1	163.0	36.771	15.974	9.7460
2.2	8.0250	3.6311	2.6587	2.3497	5.2	180.0	39.870	17.031	10.2510
2.3	8.9742	3.9465	2.8367	2.4797	5.3	199.3	43.237	18.164	10.7858
2.4	10.0232	4.2854	3.0239	2.6140	5.4	220.4	46.899	19.378	11.3523
2.5	11.1825	4.6501	3.2210	2.7528	5.5	243.7	50.887	20.680	11.9522
2.6	12.464	5.0428	3.4288	2.8966	5.6	269.4	55.216	22.076	12.589
2.7	13.880	5.4656	3.6482	3.0457	5.7	297.9	59.932	23.576	13.263
2.8	15.445	5.9214	3.8799	3.2005	5.8	329.2	65.065	25.186	13.979
2.9	17.174	6.4129	4.1250	3.3616	5.9	364.1	70.650	26.914	14.739
3.0	19.085	6.9432	4.3843	3.5292	6.0	402.4	76.734	28.769	15.546

complete distribution and from a complete count. The relative number of units is measured by the ratio $\sigma_{\hat{m}}^2/\hat{m}$, which has been plotted in Fig. 1 against the expectation $\hat{m}$ per cell for four different systems of counting.

Figure 1. Dependence of relative number of units needed for a given accuracy upon the expected frequency per unit ($\hat{m}$) with different methods of counting. The different curves show the largest number per unit counted individually (x_t) for values of x_t = 0, 1, 2 and 3, all larger numbers being pooled in each case.

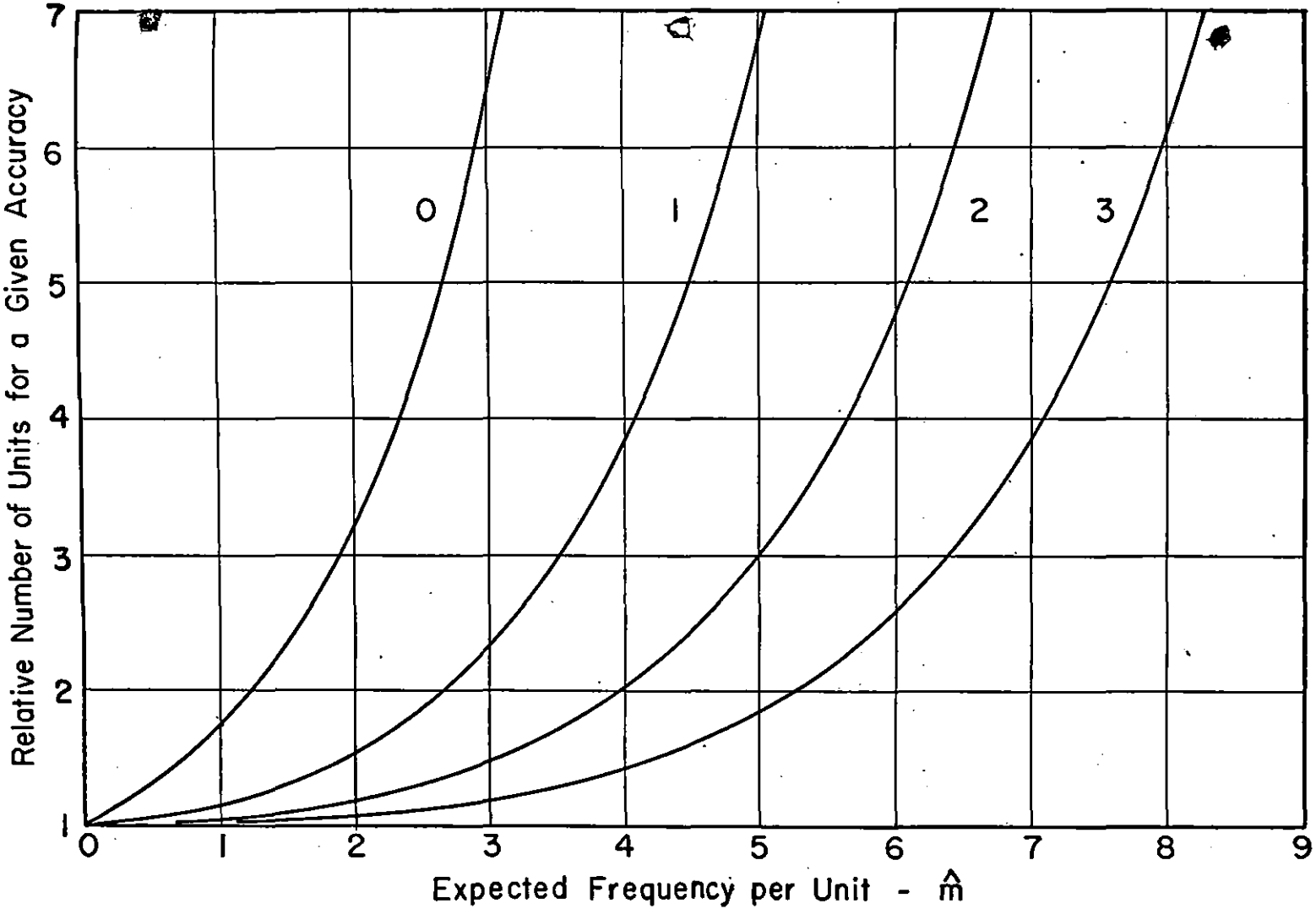
Relative Number of Units for a Given Accuracy
Expected Frequency per Unit - $\hat{m}$
0
1
2
3
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7
8
9

Figure 1

It is obvious from the figure that the pooling of all units in which $x \geqslant 1$, unavoidable under some experimental conditions [6], leads rapidly to a marked loss in efficiency as $\hat{m}$ increases. Pooling units for which $x \geqslant 4$ gives a relatively efficient estimate when the expected mean frequency per unit ($\hat{m}$) is less than 4.

The standard errors of the two means for each series in Table 3 have been computed by equations (4) and (6) respectively, as may be illustrated with series I. The mean of the grouped series was $\hat{m} = 2.691$. By linear interpolation in the column for $t = 3$ of Table 5, its variance was determined as $\sigma_{\hat{m}}^2 = 0.09 \times 2.8966 + 0.91 \times 3.0457 = 3.033$, from which $s_{\hat{m}} = \sqrt{3.033/64} = 0.218$ by equation (6). The standard error of the mean from the complete count of the same series by equation (4) was $s_{\bar{x}} = \sqrt{2.641/64} = 0.203$. The standard errors for the remaining three series were obtained similarly (Table 4). In every case the mean determined by the short-cut agreed well with that based upon the complete count.

To obtain the same precision from a partial as from a complete count of 64 squares, the number of squares in the first series should be increased by the factor $3.033/2.691 = 1.127$ or from 64 to 72 squares. An experimental test would be needed to determine whether it would be easier to examine eight additional squares or to record completely the number of spores in every square. One would also have to allow for the additional computation required with the partial count.

χ^2 Test for Agreement with the Poisson Series

The statistics computed from a partially-recorded distribution are valid only if the distribution is Poisson. This is inferred initially from the conditions giving rise to the data. The number of individuals or events in any given unit should be independent of that in any other unit and all units should be equally exposed to the chance occurrence of one or more events. In using a haemocytometer, for example, the randomness of the distribution might depend upon the care with which the slide was prepared for counting. The number of spores in one square should not modify the number in another square and all squares should be equally exposed to the chance occurrence of spores.

Even where a Poisson distribution would be expected from the available evidence, a given sample may depart significantly from expectation. In other cases there may be no information other than the observed distribution. The agreement of an observed frequency distribution with that expected by the Poisson series can be tested objectively by χ^2 [5]. An expected frequency ($m_i' = N\phi_i$) is computed to match each observed frequency (f_i) and χ^2 is determined from the paired frequencies as

$$\chi^2 = S\left\{\frac{(f_i - m_i')^2}{m_i'}\right\}. \qquad (7)$$

For a given $\hat{m}$ the expected frequencies (m'_i) at each x_i can be computed directly by equation (1) to obtain the number of units containing x_i = 0, 1, 2 and 3 individuals or events. The expected frequency for pooled values such as beyond $x_i = 3$, is given by subtraction from N. Alternatively and more easily, the expected frequencies can be interpolated from the proportionate frequencies (ϕ_i) in a table of the Poisson distribution [7, 8].

To judge the goodness of fit the χ^2 computed with equation (7) is referred to a table of the χ^2 distribution [5]. The degrees of freedom (n) in χ^2 are equal to the number of paired frequency groups diminished by two, representing $\hat{m}$ and N. If the probability P of obtaining as large a χ^2 as that observed exceeds 0.05, the data are consistent with the Poisson series. Since the uncounted units may contain an excessive number of individuals or events, χ^2 provides only a partial check. For this reason it is desirable to count a few representative series in full and to test their agreement with the Poisson series.

TABLE 6. ESTIMATION OF EXPECTED FREQUENCIES (m'_i) BY INTERPOLATION AND THEIR COMPARISON WITH THE OBSERVED FREQUENCIES (f_i) BY χ^2 FOR SERIES I IN TABLE 3.

Spores per square x	Proportionate frequency (ϕ) for		Expected frequency m'_i for $\hat{m} = 2.691$	Observed frequency f_i	$\frac{(f_i - m'_i)^2}{m'_i}$
	$\hat{m} = 2.6$	$\hat{m} = 2.7$			
0	.0743	.0672	4.34	3	.414
1	.1931	.1815	11.68	11	.040
2	.2510	.2450	15.71	19	.689
3	.2176	.2205	14.10	14	.001
⩾4	.2640	.2859	18.17	17	.075
Factor.....	5.8	58.2			
Total.....			64.00	64	1.219 = χ^2

The first series in Table 3, with an estimated mean of $\hat{m} = 2.691$ has been tested for goodness of fit. The tabular proportionate frequencies for $\hat{m}$ = 2.6 and $\hat{m}$ = 2.7 have been copied in the second and third columns of Table 6. To obtain the expected frequencies by linear interpolation, each value of m'_i for $\hat{m} = 2.6$ was multiplied by $(1 - 0.91)64 = 5.8$ and added to the products of the corresponding value for $\hat{m} = 2.7$ multiplied by $0.91 \times 64 = 58.2$. These factors have been entered at the foot of each column. The expected frequency for $x = 0$, for example, was computed as $m'_i = 5.8 \times .0743 + 58.2 \times .0672 = 4.34$ and the others were obtained similarly.

The expected frequencies m'_i were then compared with the observed values f_i by equation (7), computing from each pair its contribution to χ^2. The component for the first entry was $(3 - 4.34)^2/4.34 = 0.414$, which has been listed with the other contributions in the last column of Table 6.

The sum of the five entries gave $\chi^2 = 1.219$ with $5 - 2 = 3$ degrees of freedom. By interpolation in a table for χ^2, $P = 0.74$, indicating good agreement with the Poisson. The χ^2's for the other series (Table 4) gave similar results. A χ^2 test [5] comparing the total numbers of spores in each of the four series showed that they differed no more than would be expected by chance ($\chi^2 = 4.331, n = 3$).

Summary

When organisms or events occur at random in space or time the number of individuals in each unit follows the Poisson distribution. It may be necessary or convenient to record in full only the units containing few observations such as 0, 1, 2 and 3, combining the rest into a single category. This paper presents tables to facilitate the estimation of the population mean and its standard error from such incomplete counts. Agreement of the observed frequencies with those expected by the Poisson distribution can be tested readily by χ^2. The calculation of these statistics is illustrated by haemocytometer counts for measuring the density of the spores of milky disease in the blood of an infected Japanese beetle larva.

References

[1] Beard, R. L. Personal communication.

[2] Blackman, G. E. A study by statistical methods of the distribution of species in grassland associations. Ann. Bot. **49**: 749–777. 1935.

[3] Bliss, C. I. Statistical problems in estimating populations of Japanese beetle larvae. J. Econ. Ent. **34**: 221–232. 1941.

[4] Bowen, M. F. A method of estimating beet leafhopper populations from the proportion of uninfested plants. U. S. Dept. Agr., Bur. Ent. and Pl. Quar. **ET-225**: 1–6. 1945.

[5] Fisher, R. A. Statistical methods for research workers. 10th Ed. Oliver and Boyd, Ltd. 1946.

[6] ———. The design of experiments. 4th Ed. Oliver and Boyd, Ltd. 1947.

[7] Molina, E. C. Poisson's exponential binomial limit. D. Van Nostrand Co., Inc. 1942.

[8] Pearson, Karl. Tables for statisticians and biometricians. Part I. 2nd Ed. Biometric Laboratory, University College, London. 1924.

[9] Przyborowski, J. and H. Wilenski. Statistical principles of routine work in testing clover seed for dodder. Biometrika **27**: 273–292. 1935.

[10] Tippett, L. H. C. A modified method of counting particles. Proc. Roy. Soc. Series A. **137**: 434–446. 1932.