ON THE POISSON LAW OF SMALL NUMBERS.

By LUCY WHITAKER, B.Sc.

PART I. THEORY AND APPLICATION TO CELL-FREQUENCIES.

(1) *Introductory.*

Let p denote the probability of the happening of a certain event A, and $q = 1 - p$, the probability of its failure in one trial. Then it is well known that the distribution of the frequencies of occurrence n, $n-1$, $n-2$ ... times in a series N of n trials is given by the terms of the point binomial

$$N(p+q)^n \quad \text{.....(i).}$$

The fitting of point-binomials plotted on an elementary base c to observed frequency distributions has been discussed by Pearson*, and he has indicated that, if c be unknown, the problem can be solved in terms of the three moment coefficients μ_2, μ_3, μ_4 required to find c, p and n. In actual practice but few cases of frequency can be found which are describable in terms of a point-binomial, and of these few a considerable section have n negative, p greater than unity and q negative; thus defying at present interpretation, however well they may serve as an analytical expression of the frequency.

The hypothesis made in deducing the binomial $(p+q)^n$ as a description of frequency is clearly that each trial shall be absolutely independent of those which precede it. In this respect it may be said that binomial frequencies belong to the teetotum class of chances, and not to those of card-drawings, when each drawing is unreplaced. In the latter case the "contributory cause groups are not independent," and our series corresponds to the hypergeometrical rather than to the binomial type of progression†.

Using the customary notation $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$, the binomial is determined from:

$$\left.\begin{aligned} n &= 2/\{3 - \beta_2 + \beta_1\}, \quad c = \sigma\sqrt{6 - 2\beta_2 + 3\beta_1} \\ pq &= \tfrac{1}{2}(3 - \beta_2 + \beta_1)/(6 - 2\beta_2 + 3\beta_1) \end{aligned}\right\} \quad \text{.....(ii).}$$

* "Skew Variation in Homogeneous Material," *Phil. Trans.* Vol. 186, A, p. 347, 1895.

† *Phil. Trans.* Vol. 186, A, p. 381, 1895.

In order that n should be positive, it is needful that

$$3 - \beta_2 + \beta_1, \quad = \tfrac{1}{2}(6 - 2\beta_2 + 2\beta_1),$$

should be positive. If this is satisfied clearly c will be real because β_1 is always positive. Further then

$$pq = p(1-p) = \frac{1}{4} \times \frac{6 - 2\beta_2 + 2\beta_1}{6 - 2\beta_2 + 3\beta_1}$$

is always less than a quarter and p and q will therefore be real. If the reader will turn to Rhind's diagram, *Biometrika*, Vol. VII. p. 131, he will see that the line $3 - \beta_2 + \beta_1 = 0$ cuts off all curves of Types III, IV, V and VI, and includes a portion only of Type I, with a part of its U and J varieties. The binomial description of frequency, therefore, is not—considering our experience of frequency distributions—likely to be of very universal application.

(2) *Further Limitations.*

Now let us still further limit our binomial by supposing:

(i) that the unit of grouping of the observed frequencies corresponds to the actual binomial base unit c and (ii) that the first of the observed frequencies corresponds to the term Np^n of the binomial*.

In this case the mean m of the observed frequency measured from the first term of the frequency will be equal to the nq of the binomial and the standard deviation of the observed distribution will be equal to $\sqrt{npq}$. We have thus:

$$p = \sigma^2/m, \quad q = 1 - \sigma^2/m, \quad n = m^2/(m - \sigma^2) \ldots\ldots\ldots\ldots\ldots\ldots\text{(iii)}$$

and n and q will both be negative, if m be less than σ^2. The condition for a positive binomial is therefore that σ be less than $\sqrt{m}$.

(3) *Probable errors of the constants of a Binomial Frequency.*

It is desirable to find the probable errors of p and n as determined by these formulae. We have:

$$\mu_1' = nq, \qquad \mu_2 = npq,$$
$$\delta\mu_1' = q\delta n + n\delta q, \quad \delta\mu_2 = pq\delta n + nq\delta p + np\delta q,$$

assuming deviations may be represented by differentials.

Hence, since $dp = -dq$:

$$\delta\mu_2 - (p-q)\,\delta\mu_1' = q^2\delta n \quad \text{and} \quad p\delta\mu_1' - \delta\mu_2 = nq\delta q.$$

Square each of these results, sum for all samples and divide by the number of samples, and we have:

$$\sigma^2_{\mu_2} + (p-q)^2\sigma^2_{\mu_1'} - 2(p-q)\,\sigma_{\mu_2}\sigma_{\mu_1'}r_{\mu_2\mu_1'} = q^4\sigma_n^2$$
$$\sigma^2_{\mu_2} + p^2\sigma^2_{\mu_1'} - 2p\sigma_{\mu_2}\sigma_{\mu_1'}r_{\mu_2\mu_1'} = n^2q^2\sigma_q^2.$$

* The exact nature of these limitations must be fully appreciated. The best fitting binomial to a given frequency distribution will usually be far from one in which the first term of the binomial corresponds to the first observed frequency. The modes of the binomial and the observed frequency will closely correspond, but the "tails" of the binomial may be quite insignificant and correspond to no observed frequencies.

Now σ_{μ_2} is the standard deviation of variations in μ_2 and therefore

$$\sigma^2_{\mu_2} = (\mu_4 - \mu_2^2)/N.$$

Similarly $\sigma_{\mu_1'}$ is the standard deviation of variations in the mean and therefore $\sigma^2_{\mu_1'} = \mu_2/N$. Lastly the product $\sigma_{\mu_2}\sigma_{\mu_1'}r_{\mu_2\mu_1'}$ measures the correlation between deviations in μ_2 and μ_1' and is known to be μ_3/N*.

Thus we have:

$$q^4\sigma_n^2 = \frac{1}{N}\{\mu_4 - \mu_2^2 + (p-q)^2\mu_2 - 2(p-q)\mu_3\},$$

$$n^2q^2\sigma_q^2 = \frac{1}{N}\{\mu_4 - \mu_2^2 + p^2\mu_2 - 2p\mu_3\}.$$

But† $$\left.\begin{aligned}\mu_4 &= npq\{1 + 3(n-2)pq\},\\ \mu_3 &= npq(p-q), \quad \mu_2 = npq\end{aligned}\right\} \quad \text{.....(iv).}$$

Whence after some purely algebraical reductions we deduce:

$$\sigma_n = \frac{n}{\sqrt{N}}\frac{p}{q}\sqrt{2\left(1-\frac{1}{n}\right)} = \frac{\sigma^2}{\sqrt{N}q^2}\sqrt{2\left(1-\frac{1}{n}\right)} \quad \text{.....(v),}$$

$$\sigma_p = \sigma_q = \frac{p}{\sqrt{N}}\sqrt{2 + \frac{1-3p}{np}} \quad \text{.....(vi).}$$

Formulae (v) and (vi) are very important; they enable us to obtain the probable errors for n and p when a binomial limited in the present manner is fitted to a frequency distribution‡.

We see at once, that as n grows large and q grows small

$$\sigma_p = \sigma_q \text{ approaches the limit } \sqrt{2/N},$$

or the probable error, ·67449 $\sqrt{2/N}$, of p and q is finite. But σ^2 being finite σ_n becomes infinitely great, or the probable error of n indefinitely large. Thus when the n of the binomial is very large, q being very small, the probable error of its determination is so great that its actual value is not capable of being found accurately. Again, suppose N embraced 200 observations, the probable error of q would be of the order ·07; if N corresponded to only eighteen observations, then the probable error of q would be of the order ·22. It is clearly wholly impossible

* *Biometrika*, Vol. II. "On the Probable errors of Frequency Constants," see p. 275 (iv), p. 276 (vii), and p. 279 (xii).

† *Phil. Trans.* Vol. 186, A, p. 347, 1895.

‡ There is no difficulty in obtaining the probable errors of n and p from the more general values in (ii). In this case

$$\sigma_n = \tfrac{1}{2}n^2\sqrt{\sigma^2_{\beta_1} + \sigma^2_{\beta_2} - 2\sigma_{\beta_1}\sigma_{\beta_2}r_{\beta_1\beta_2}},$$

$$\sigma_p = \sigma_q = \frac{pq}{\beta_1}\sqrt{\sigma^2_{\beta_1} + \tfrac{1}{4}n^2\beta_1^2\sigma^2_{\beta_2} - n\beta_1\sigma_{\beta_1}\sigma_{\beta_2}r_{\beta_1\beta_2}}.$$

The values of σ_{β_1}, σ_{β_2} and $r_{\beta_1\beta_2}$ for different values of β_1 and β_2 have been tabled by Rhind, *Biometrika*, Vol. VII. pp. 136—141.

from series of observations even of the order 200, much less of order 18, to assert that q is or is not really a "small quantity." Thus the observed value of q corresponding to a population of extremely small q might easily show $q = \cdot 15$ to $\cdot 50$!.

(4) *Poisson—Law of Small Numbers.*

A last limitation of the point-binomial is made by supposing the mean $m = nq$ to remain finite, but q to be indefinitely small. We write:

$$N(p+q)^n = N(1-q+q)^n = N(1-q)^{\frac{m}{q}}\left(1+\frac{q}{1-q}\right)^{\frac{m}{q}}$$
$$= N(1-q)^{\frac{m}{q}}(1+q)^{\frac{m}{q}} \text{ nearly}$$
$$= Ne^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\ldots\right).$$

Here the successive terms give the frequency of occurrence of 0, 1, 2, 3... successes on the basis of each success not being prejudiced by what has previously occurred. This is the Law of Small Numbers. It was first published by Poisson in 1837*. It was adopted later by Bortkewitsch, who published a small treatise expanding by illustrations Poisson's work†. The same series was deduced later by "Student" in ignorance of both Poisson and Bortkewitsch's papers, when dealing with the counts made with a haemacytometer‡.

The mean is at m from the first group, the other moments as "Student" has shewn § are:

$$\mu_2 = m, \quad \mu_3 = m, \quad \mu_4 = 3m^2 + m.$$

Hence $$\beta_1 = 1/m, \quad \beta_2 - 3 = 1/m.$$

When the mean value is large, β_1, β_2 and the higher β's approach the values given by the Gaussian curve.

Clearly the Poisson-Exponential formula contains only the single constant $m = \mu_1'$ and its probable error is therefore $\cdot 67449\sigma/\sqrt{N} = \cdot 67449\sqrt{\frac{m}{N}}$. This will, if N be reasonably large and m not too big, be a small or at any rate a finite quantity (i.e. not like σ_n for q very small). Hence it might be supposed, although erroneously, that the Poisson-Exponential formula was capable of great accuracy in addition to its great simplicity. But this is to neglect the fundamental assumptions on which it is based, namely:

(i) that the data actually correspond to a binomial,

(ii) that in that binomial q is small and n large.

Clearly (i) shows us that, if we can find the binomial, it will actually be closer to the observed frequency than Poisson's merely approximate formula.

* *Recherches sur la Probabilité des Jugements.* Paris, 1837, pp. 205 *et seq.*

† *Das Gesetz der kleinen Zahlen*, Leipzig, 1898.

‡ "On the Error of Counting with a Haemacytometer," *Biometrika*, Vol. v, pp. 351—5, 1907.

§ They may be deduced at once from (iv).

Secondly (ii) can only be justified as an assumption by actually ascertaining the form of the binomial from the data and testing whether n is large and q small and *positive*. It appears absurd to base our formula on an approximation to a binomial of a particular kind when, on testing in the actual problem, such a binomial does not describe the results. As a merely empirical formula, the Poisson-Exponential of course can be tested by the usual processes for measuring goodness of fit, but no such test nor any discussion of the probable errors of their results have been provided by Bortkewitsch himself nor by Mortara, who has followed recently his lines in a work to be considered later. As a matter of fact in the cases dealt with by Bortkewitsch, by Mortara and by "Student," n will be found almost as frequently small and negative as large and positive, and q takes a great variety of values large and negative and large and positive, as well as small and positive. Thus the initial assumptions made from which the "law of small numbers" is deduced are by no means justified on the material to which it has so far been applied.

(5) *Application of the Law of Small Numbers to determine the Probable Errors of Small Frequencies.* Given a distribution of frequency for a population $\bar{N}$ let $\bar{n}_{st}$ be the frequency in the cell of the sth row and tth column of a contingency table (or if we drop t, $\bar{n}_s$ would stand for the frequency of any class). Then if we take a random sample of N individuals from this population, the chance that an individual is taken out of the $\bar{n}_{st}$ cell is $\bar{n}_{st}/\bar{N}$, and that it is not is $1-\frac{\bar{n}_{st}}{\bar{N}}$. Therefore if the original population be so large that the withdrawal of an individual does not affect the next draw, the frequency of individuals in M random samples of N will be given by the terms of the binomial:

$$M\left\{\left(1-\frac{\bar{n}_{st}}{\bar{N}}\right)+\frac{\bar{n}_{st}}{\bar{N}}\right\}^N.$$

Now, if $\bar{n}_{st}/\bar{N}$ be very small, and N large this will approximate to the Poisson series:

$$Me^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\ldots\right),$$

where $m=\frac{\bar{n}_{st}}{\bar{N}}\times N$. But $\bar{n}_{st}/\bar{N}$ will approximately be the mean proportion of the whole in the st cell of the sample itself $=n_{st}/N$, or $m=n_{st}$. Thus if in any cell of a contingency table, or in any sub-class of a frequency whatsoever, we have a frequency n_{st} small as compared to the population N, then in sampling, this small frequency will have a distribution approximating to the Poisson Law, and tending as n_{st} becomes larger to approach the Gaussian distribution*. It would appear,

* Such approach is usually *assumed* when we speak of

$$\cdot 67449\sqrt{n_{st}\left(1-\frac{n_{st}}{N}\right)}$$

as the probable error of the frequency n_{st}. But such a "probable error" has really no meaning if n_{st} be very small and the exponential law be applied.

therefore, that the Poisson Law of Small Numbers should be applied in order to deal with the errors of random sampling in any *small* frequency, and an appeal should not be made—as is usually the case—to Sheppard's Tables on the assumption that the frequency is Gaussian.

The following Table I illustrates the results obtained (*a*) from the Binomial, (*b*) from the Poisson-Exponential and (*c*) from the normal curve on the two hypotheses that (i) the frequency is 10 in the 1000 and (*b*) is 30 in the 1000. But here a word must be said as to which Gaussian is to be compared with the Binomial or the Poisson-Exponential. The usual method of fitting a Gaussian is to give it the same mean and standard-deviation as the material to which we are fitting it. For example, we should compare the Poisson exponential with a Gaussian at mean m and with standard-deviation $\sqrt{m}$, or the point binomial with mean nq

TABLE I.

Comparison of Binomial, Poisson-Exponential and Gaussian for cell-frequency variations in samples for case of 10 *and* 30 *in a total population of* 1000

Percentage Frequency

Number in Cell	*10* in *1000*				*30* in *1000*		
	Binomial	Poisson-Exponential	Gaussian		Binomial	Poisson-Exponential	Gaussian
0	·00004	·00005	·00132	*19*	·00848	·00894	·01100
1	·00044	·00045	·00327	*20*	·01287	·01341	·01553
2	·00020	·00227	·00735	*21*	·01857	·01916	·02118
3	·00739	·00757	·01491	*22*	·02556	·02613	·02792
4	·01861	·01892	·02736	*23*	·03362	·03408	·03544
5	·03745	·03783	·04539	*24*	·04233	·04260	·04373
6	·06274	·06306	·06806	*25*	·05110	·05112	·05198
7	·08999	·09080	·09224	*26*	·05927	·05898	·05970
8	·11282	·11260	·11300	*27*	·06613	·06553	·06625
9	·12561	·12511	·12514	*28*	·07107	·07021	·07104
				29	·07367	·07263	·07360
10	·12574	·12511	·12526	*30*	·07375	·07263	·07367
11	·11431	·11374	·11334				
12	·09516	·09478	·09271	*31*	·07137	·07029	·07126
13	·07305	·07291	·06854	*32*	·06684	·06590	·06659
14	·05202	·05208	·04580	*33*	·06064	·05991	·06013
15	·03454	·03472	·02767	*34*	·05334	·05286	·05246
16	·02148	·02170	·01511	*35*	·04553	·04531	·04423
17	·01256	·01276	·00746	*36*	·03775	·03776	·03602
18	·00693	·00709	·00333	*37*	·03042	·03061	·02835
19	·00362	·00373	·00134	*38*	·02384	·02417	·02156
20	·00179	·00187	·00049	*39*	·01819	·01859	·01584
21	·00085	·00089	·00016	*40*	·01351	·01394	·01125
22	·00038	·00040	·00005	*41*	·00979	·01020	·00771
23	·00016	·00018	·00001	*42*	·00691	·00729	·00511

and standard-deviation $\sqrt{npq}$. These will, however, not be identical standard deviations as p is not truly unity. In ordinary practice, in testing for example the 30 in 1000 frequency, we should put the centre of our Gaussian at our 30 group, and use a standard deviation $= \sqrt{30\,(1-30/1000)} = \sqrt{30 \times \cdot 97} = 5{\cdot}39444$ to enter the table of the probability integral. This is, of course, the Gaussian we obtain by the method of least squares, but to assume that it is "the best" is to argue in a circle, because we then take least squares as a test of what is best*. It is not the Gaussian which is directly reached by proceeding either to a limit of the Binomial or to the Exponential, for example, by applying Stirling's Theorem. It will be seen by examining Table II that the Gaussian curve develops out of the exponential by a mode at the point midway between the two equal terms, rather than by a mode at the mean, which coincides with the centre of the second of them. If we apply Stirling's Theorem to the term†

$$N \frac{\lfloor n}{\lfloor n-r \lfloor r} p^{n-r} q^r$$

of the binomial $N(p+q)^n$ it becomes

$$u_r = \frac{N}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{1}{2}\{r-nq+\frac{1}{2}(p-q)\}^2/(npq)},$$

i.e. the *ordinate* of a Gaussian curve of Standard Deviation $\sqrt{npq}$ and mean at $nq - \frac{1}{2}(p-q)$. These give for the Poisson-Exponential the Gaussian with standard-deviation $\sqrt{m}$ and mean $m - \frac{1}{2}$. The above type of curve which gives frequencies by coordinates and not by areas has been termed by Sheppard a 'spurious curve of frequency'; at the same time it is the method by which Laplace and Poisson first reached the normal curve, and the real point at issue is whether we shall get better approximations to the discontinuous frequencies of the binomials by using Gaussian ordinates than by using the areas of a Gaussian curve. At the same time it has been shewn‡ that if a Gaussian curve gives a series of frequencies by its areas, then if its standard-deviation be σ^2, a spurious Gaussian frequency curve with standard deviation given by $\sigma_0^2 = \sigma^2 + \frac{1}{12}h^2$, h being the sub-range, will closely give the frequencies by its ordinates. It seems probable therefore that the Gaussian curve with mean at $nq - \frac{1}{2}(p-q)$ and standard deviation $\sqrt{npq - \frac{1}{12}}$ will more closely represent the binomial for cell frequency variation by its areas,

* There is a further flaw in this treatment—the Gaussian is continuous, the Binomial and the Poisson-Exponential are not. If t_r be the rth term of either of the latter series, we ought really to make

$$S_0^{\infty}\left[\left\{t_r - \int_{r-\bar{m}}^{r+1-\bar{m}} \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x}{\sigma^2}} dx\right\}^2 \Big/ t_r\right] = u,$$

a minimum by the conditions $du/d\bar{m} = du/d\sigma = 0$. No complete solution of this problem has hitherto been determined.

† The final form for u_r may be obtained by neglecting the terms in $\frac{1}{n^2}$ in the formula given by Pearson, *Phil. Trans.* Vol. 186, A, p. 347, footnote.

‡ *Biometrika*, Vol. III. p. 311.

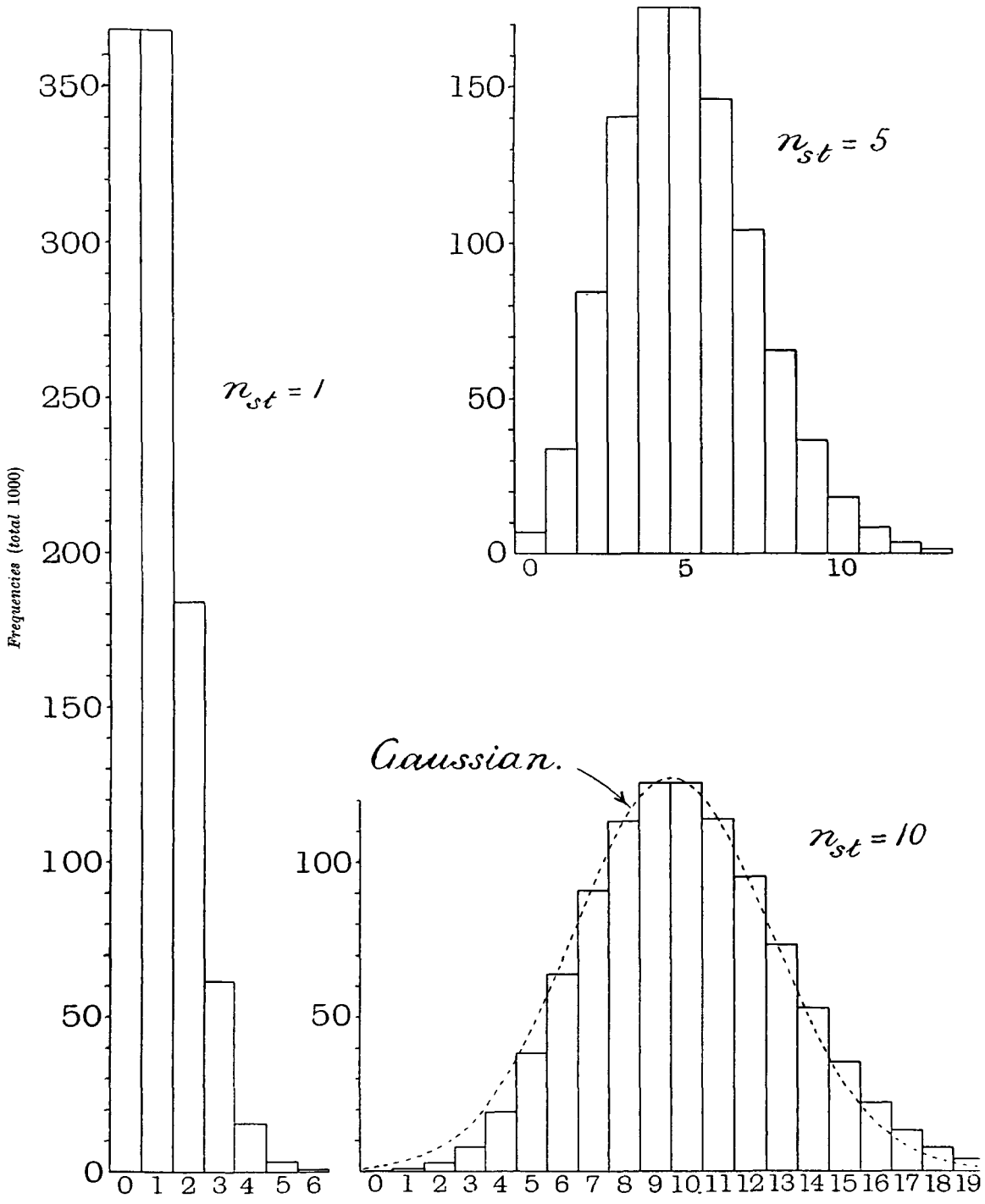
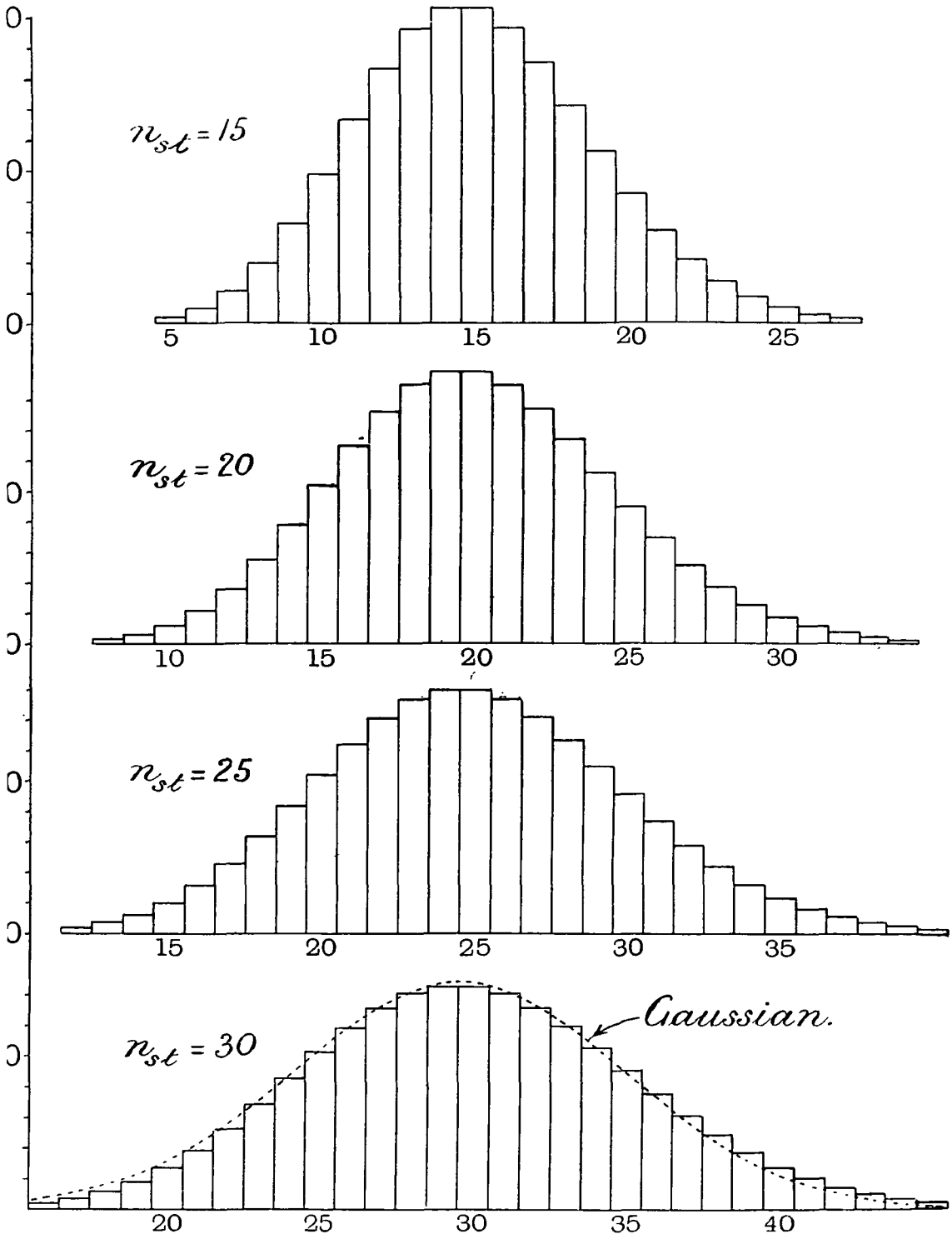


Plate VII



than if we apply the ordinary process of mean nq, standard deviation $\sqrt{npq}$, and Sheppard's table for areas to the frequencies. It will be noted that this amounts to using Sheppard's correction on the crude second-moment and slightly shifting the central ordinate towards the side of greater frequency. This is the Gaussian curve used in Table I.

The object of the present section of our work is to indicate how far it is legitimate to use the Poisson-Exponential up to cell frequencies of the order 30 in a population of about 1000* and how far we then reach a state of affairs, which for practical purposes may be described by ordinary tables of the Gaussian. It will be seen from Table I that the Poisson-Exponential even for $n_{st} = 10$ and 30 is not extremely divergent from the Binomial.

In Plate VII the transition of the exponential histograms of frequency towards the Gaussian form is indicated for cell-frequency = 1, 5, 10, 15, 20, 25 and 30; in the cases of 10 and 30 the corresponding Gaussian curves are drawn.

It will be seen that with due caution the Poisson-Exponential may be reasonably used up to frequencies of about 30 in the 1000, and that after that it would be fairly satisfactory to use the areas of the Gaussian curve as provided in the usual tables.

(6) In order to table the results of the Poisson-Exponential for easy use, it seemed desirable to turn them into percentages of excess and defect. For example take the distribution for a frequency 5. It is:

		Per cent. of Cases in which:			
0	·006,737,945	a defect of 5		occurs:	0·674
1	·033,689,725	„	4 or more	„ :	4·043
2	·084,224,310	„	3 or more	„ :	12·465
3	·140,373,850	„	2 or more	„ :	26·503
4	·175,467,310	„	1 or more	„ :	44·049
5	·175,467,310	the true value		„ :	17·547
6	·146,222,755	an excess of 1	or more	„ :	38·404
7	·104,444,825	„	2 or more	„ :	23·782
8	·065,278,015	„	3 or more	„ :	13·337
9	·036,265,564	„	4 or more	„ :	6·809
10	·018,132,782	„	5 or more	„ :	3·183
11	·008,242,173	„	6 or more	„ :	1·370
12	·003,434,238	„	7 or more	„ :	0·545
13	·001,320,860	„	8 or more	„ :	0·202

* Of course in the Poisson-Exponential itself the total frequency plays no part; it is only useful in testing the validity of the approximation.

Thus we see that if the true value of the frequency be 5 for the average sample, it will only lie outside the range 1 to 10 in ·674 + 1·370 = 2·044 cases per cent., or the odds are 49 to 1 that the value found will be from 1 to 10.

On the other hand it will lie outside the range 2 to 8 in 4·043+6·809=10·852 °/₀ of cases, or once in about 9 trials the frequency will lie outside this range. Or, again, once in about every four trials (25·8 °/₀) the result will fall outside the range 3 to 7.

On the other hand if we write $\sigma = \sqrt{5(1 - \cdot 005)} = 2 \cdot 23047$, we have $-4 \cdot 5$ and $+5 \cdot 5$ as the deviations from a mean 5 of all beyond 0·5 and above 10·5, giving $x/\sigma = -2 \cdot 0175$ and $+2 \cdot 4658$ respectively. These cut off tail areas of ·02181 and ·00684, respectively. Thus in 2·865—not 2·044—per cent. of cases we should assert that the frequency would lie outside the range 1 to 10, or the odds that it would lie inside this range are now only about 34 to 1, not 49 to 1. Calculated from the Gaussian the frequencies outside ranges 2 to 8 and 3 to 7 correspond to 10·1 °/₀ and 26·2 °/₀ of the trials instead of 10·9 °/₀ and 25·8 °/₀. If we take for the standard-deviation of our Gaussian $\sqrt{npq - \frac{1}{12}} = 2 \cdot 21171$, we find that the odds in the first case are still only 35 to 1, but the percentages in the other two cases are 11·3 and 25·8.

It will be clear that near the centre of the curve—especially when we equalise the excess and defect of the Gaussian by taking equal ranges on both sides—it does not give bad percentages of frequency, but that it does not lend itself to the accurate determination of the range for reasonable working odds such as 50 to 1.

It will be noted that the total area in excess and defect of 2 and more = 23·782 + 26·503 = 50·285, or corresponds very nearly to the "probable error." Actually the Gaussians with standard deviations of 2·23047 and 2·21171 give probable errors of 1·504 and 1·492 respectively, so that the Gaussian with 1·5 as the probable error is very nearly accurate.

Table II gives the Poisson-Exponential; it will enable the reader to appreciate the range of probable variation in small frequencies. Thus we realise that in 37 °/₀ of cases in which the true frequency is 1, the cell will be found empty; in 13·5 per cent. of cases it will be empty when the actual frequency is 2, and in 5 °/₀ of cases when the frequency is 3 and in 1·8 °/₀ when the frequency is 4. These results indicate how rash it is to assume that a sample 4-fold table with one zero quadrant signifies perfect dependence or association in the attributes of the material sampled. The second line below gives the percentages of cases that 0 would appear in a cell when the actual number to be expected is that in the first line calculated from Table II on the usual theory of *a priori* probabilities:

Actual ...	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9* & over
Percentage ...	63·21	23·25	8·55	3·15	1·16	0·43	0·16	0·06	0·02	0·01

TABLE II.

Table of Poisson-Exponential for Cell Frequencies 1 *to* 30.

Cell Frequencies

	x	1	2	3	4	5	6	7	8	9	10
Per cent. occurrence of values differing by x or more in defect from Actual.	22										
	21										
	20										
	19										
	18										
	17										
	16										
	15										
	14										
	13										
	12										
	11										
	10										·005
	9									·012	·050
	8								·034	·123	·277
	7							·091	·302	·623	1·033
	6						·248	·730	1·375	2·123	2·925
	5					·674	1·735	2·964	4·238	5·496	6·708
	4				1·832	4·043	6·197	8·177	9·963	11·569	13·014
	3			4·979	9·158	12·465	15·120	17·299	19·124	20·678	22·022
	2		13·534	19·915	23·810	26·503	28·506	30·071	31·337	32·390	33·282
	1	36·788	40·601	42·319	43·347	44·049	44·568	44·971	45·296	45·565	45·793
	Actual	36·788	27·067	22·404	19·537	17·547	16·062	14·900	13·959	13·176	12·511
Per cent. occurrence of values differing by x or more in excess from Actual.	1	26·424	32·332	35·277	37·116	38·404	39·370	40·129	40·745	41·259	41·696
	2	8·030	14·288	18·474	21·487	23·782	25·602	27·091	28·338	29·401	30·323
	3	1·899	5·265	8·392	11·067	13·337	15·276	16·950	18·411	19·699	20·845
	4	·366	1·656	3·351	5·113	6·809	8·392	9·852	11·192	12·422	13·554
	5	·059	·453	1·191	2·136	3·183	4·262	5·335	6·380	7·385	8·346
	6	·008	·110	·380	·813	1·370	2·009	2·700	3·418	4·146	4·875
	7	·001	·024	·110	·284	·545	·883	1·281	1·726	2·203	2·705
	8	·000	·005	·029	·092	·202	·363	·572	·823	1·110	1·428
	9	—	·001	·007	·027	·070	·140	·241	·372	·532	·719
	10	—	·000	·002	·008	·023	·051	·096	·159	·242	·346
	11	—	—	·000	·002	·007	·018	·036	·065	·105	·160
	12	—	—	—	·001	·002	·006	·013	·025	·044	·071
	13	—	—	—	·000	·001	·002	·005	·009	·017	·030
	14	—	—	—	—	·000	·001	·002	·003	·007	·013
	15	—	—	—	—	—	·000	·001	·001	·002	·006
	16	—	—	—	—	—	—	·000	·000	·001	·002
	17	—	—	—	—	—	—	—	—	·000	·001
	18	—	—	—	—	—	—	—	—	—	·001
	19	—	—	—	—	—	—	—	—	—	·000
	20	—	—	—	—	—	—	—	—	—	—
	21	—	—	—	—	—	—	—	—	—	—
	22	—	—	—	—	—	—	—	—	—	—
	23	—	—	—	—	—	—	—	—	—	—
	24	—	—	—	—	—	—	—	—	—	—
	25	—	—	—	—	—	—	—	—	—	—
	26	—	—	—	—	—	—	—	—	—	—
	27	—	—	—	—	—	—	—	—	—	—
	28	—	—	—	—	—	—	—	—	—	—

TABLE II—(*continued*).

Cell Frequencies

	x	11	12	13	14	15	16	17	18	19	20
Per cent. occurrence of values differing by *x* or more in defect from Actual.	22										
	21										
	20										—
	19									—	—
	18								—	—	—
	17							—	—	·000	·000
	16						—	·000	·000	·001	·002
	15					·000	·000	·001	·002	·004	·007
	14				·000	·001	·002	·004	·008	·015	·026
	13			·000	·001	·004	·009	·018	·032	·052	·078
	12		·001	·003	·009	·021	·040	·067	·104	·151	·209
	11	·002	·008	·022	·047	·086	·138	·206	·289	·387	·500
	10	·020	·052	·105	·181	·279	·401	·543	·706	·886	1·081
	9	·121	·229	·374	·553	·763	1·000	1·260	1·538	1·832	2·139
	8	·492	·760	1·073	1·423	1·800	2·199	2·612	3·037	3·467	3·901
	7	1·510	2·034	2·589	3·162	3·745	4·330	4·912	5·489	6·056	6·613
	6	3·752	4·582	5·403	6·206	6·985	7·740	8·467	9·167	9·840	10·486
	5	7·861	8·950	9·976	10·940	11·846	12·699	13·502	14·260	14·975	15·651
	4	14·319	15·503	16·581	17·568	18·475	19·312	20·087	20·808	21·479	22·107
	3	23·198	24·239	25·168	26·004	26·761	27·451	28·084	28·665	29·203	29·703
	2	34·051	34·723	35·317	35·846	36·322	36·753	37·146	37·505	37·836	38·142
	1	45·989	46·150	46·311	46·445	46·565	46·674	46·774	46·865	46·948	47·026
	Actual	11·938	11·437	10·994	10·599	10·244	9·922	9·629	9·360	9·112	8·884
Per cent. occurrence of values differing by *x* or more in excess from Actual.	1	42·073	42·404	42·695	42·956	43·191	43·404	43·597	43·776	43·939	44·091
	2	31·130	31·846	32·486	33·064	33·588	34·066	34·503	34·909	35·283	35·630
	3	21·871	22·798	23·639	24·408	25·114	25·765	26·367	26·928	27·451	27·939
	4	14·596	15·559	16·450	17·280	18·053	18·776	19·451	20·088	20·686	21·251
	5	9·261	10·129	10·953	11·736	12·478	13·184	13·852	14·491	15·099	15·677
	6	5·593	6·297	6·983	7·650	8·297	8·923	9·526	10·111	10·675	11·219
	7	3·219	3·742	4·266	4·791	5·311	5·825	6 329	6·826	7·313	7·789
	8	1·769	2·128	2·501	2·884	3·275	3·669	4·064	4·461	4·856	5·248
	9	·929	1·160	1·407	1·671	1·947	2·232	2·523	2·824	3·127	3·433
	10	·467	·607	·762	·933	1·117	1·312	1·516	1·732	1·954	2·182
	11	·225	·305	·396	·502	·619	·746	·882	1·030	1·185	1·348
	12	·104	·148	·201	·261	·331	·411	·497	·595	·699	·809
	13	·047	·069	·097	·131	·172	·219	·272	·333	·400	·473
	14	·020	·031	·046	·063	·086	·114	·144	·182	·223	·269
	15	·008	·014	·021	·030	·042	·057	·074	·096	·121	·149
	16	·003	·006	·009	·013	·020	·028	·036	·050	·064	·081
	17	·001	·002	·004	·006	·009	·014	·017	·025	·033	·042
	18	·001	·000	·002	·002	·004	·006	·008	·012	·017	·022
	19	·000	·000	·001	·001	·002	·003	·003	·006	·008	·011
	20	—	—	·001	·000	·001	·002	·002	·003	·004	·005
	21	—	—	·000	·000	·000	·001	·001	·002	·002	·003
	22	—	—	—	—	—	·000	·000	·001	·001	·001
	23	—	—	—	—	—	—	—	·000	·000	·001
	24	—	—	—	—	—	—	—	—	—	·000
	25	—	—	—	—	—	—	—	—	—	—
	26	—	—	—	—	—	—	—	—	—	—
	27	—	—	—	—	—	—	—	—	—	—
	28	—	—	—	—	—	—	—	—	—	—

TABLE II—(*continued*).

Cell Frequencies

	x	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*
Per cent. occurrence of values differing by *x* or more in defect from Actual.	*22*		—	—	—	—	—	—	—	—	·000
	21	—	—	—	—	—	—	·000	·000	·000	·001
	20	—	—	—	—	·000	·000	·001	·001	·001	·002
	19	—	—	·000	·000	·001	·001	·002	·003	·004	·006
	18	·000	·000	·001	·001	·002	·004	·006	·009	·012	·017
	17	·001	·002	·003	·005	·008	·011	·016	·023	·031	·041
	16	·003	·006	·010	·015	·022	·031	·043	·056	·073	·092
	15	·012	·020	·030	·043	·059	·078	·102	·129	·160	·195
	14	·039	·058	·081	·109	·142	·180	·224	·273	·328	·387
	13	·111	·150	·198	·252	·314	·384	·460	·543	·632	·727
	12	·277	·355	·443	·540	·647	·762	·884	1·014	1·151	1·293
	11	·625	·763	·912	1·072	1·240	1·417	1·601	1·791	1·987	2·187
	10	1·290	1·512	1·743	1·983	2·229	2·482	2·739	3·000	3·263	3·528
	9	2·455	2·778	3·107	3·440	3·775	4·111	4·446	4·781	5·114	5·444
	8	4·336	4·769	5·200	5·626	6·048	6·463	6·872	7·274	7·669	8·057
	7	7·157	7·689	8·208	8·713	9·204	9·682	10·147	10·599	11·038	11·465
	6	11·107	11·704	12·277	12·827	13·358	13·867	14·357	14·830	15·285	15·724
	5	16·292	16·900	17·477	18·025	18·549	19·048	19·525	19·981	20·417	20·836
	4	22·696	23·250	23·771	24·263	24·730	25·172	25·591	25·990	26·371	26·734
	3	30·168	30·603	31·010	31·391	31·753	32·094	32·416	32·721	33·011	33·287
	2	38·426	38·691	38·938	39·168	39·387	39·593	39·786	39·970	40 143	40·308
	1	47·097	47·164	47·226	47·283	47·340	47·392	47·440	47·486	47·530	47·572
	Actual	8·671	8·473	8·288	8·115	7·952	7·799	7·654	7·517	7·387	7·264
Per cent. occurrence of values differing by *x* or more in excess from Actual.	*1*	44·232	44·363	44·485	44·603	44·708	44·810	44·906	44·997	45·083	45·165
	2	35·955	36·258	36·542	36·812	37·062	37·299	37·525	37·739	37·942	38·135
	3	28·397	28·828	29·235	29·620	29·982	30·326	30·653	30·965	31·262	31·546
	4	21·785	22·290	22·770	23·227	23·660	24·074	24·469	24·847	25·208	25·555
	5	16·230	16·758	17·264	17·748	18·211	18·655	19·083	19·493	19·888	20·269
	6	11·744	12·251	12·740	13·213	13·669	14·110	14·538	14·951	15·351	15·738
	7	8·254	8·709	9·153	9·585	10·007	10·418	10·819	11·210	11·591	11·962
	8	5·637	6·022	6·402	6·777	7·146	7·509	7·866	8·218	8·562	8·901
	9	3·742	4·052	4·362	4·670	4·978	5·284	5·588	5·890	6·188	6·484
	10	2·415	2·654	2·895	3·138	3·385	3·632	3·880	4·129	4·377	4·625
	11	1·517	1·692	1·873	2·057	2·246	2·438	2·633	2·831	3·030	3·230
	12	·927	1·051	1·181	1·315	1·456	1·599	1·747	1·899	2·053	2·210
	13	·552	·637	·727	·821	·921	1·025	1·134	1·247	1·362	1·481
	14	·320	·376	·437	·500	·570	·643	·720	·801	·885	·973
	15	·181	·217	·256	·298	·345	·394	·448	·504	·564	·626
	16	·100	·122	·147	·173	·204	·237	·272	·311	·352	·395
	17	·054	·067	·082	·098	·118	·139	·162	·188	·215	·245
	18	·028	·036	·045	·055	·067	·080	·095	·111	·129	·149
	19	·015	·019	·024	·030	·037	·045	·054	·065	·076	·089
	20	·007	·010	·013	·016	·020	·025	·031	·037	·044	·052
	21	·004	·005	·007	·009	·011	·014	·017	·021	·025	·030
	22	·002	·002	·004	·005	·006	·007	·009	·011	·014	·017
	23	·001	·001	·002	·003	·003	·004	·005	·006	·007	·010
	24	·001	·001	·001	·002	·002	·002	·003	·003	·004	·006
	25	·000	·000	·001	·001	·001	·001	·001	·002	·002	·003
	26	—	—	·000	·000	·000	·001	·001	·001	·001	·002
	27	—	—	—	—	—	·000	·000	·000	·000	·001
	28	—	—	—	—	—	—	—	—	—	·000

PART II. CRITICISMS OF PREVIOUS APPLICATIONS OF POISSON'S LAW OF SMALL NUMBERS.

(7) We now turn to the illustrations which various authors have given of the Law of Small Numbers.

"Student's" Cases. We take first the series given by "Student" in his memoir on counting with a Haemacytometer*. They are of special importance because the series at first appear of fairly adequate size, namely consisting of 400 individuals, and further we should anticipate that the Law of Small Numbers would hold in his cases. He obtains better fits with the binomial than with the exponential but, as he remarks, he has one more constant at his disposal. On the other hand, if the exponential be a true approximation, the binomial ought to come out with a large n and a small but positive q. "Student" finds for his four series:

I. $400 \times (1{\cdot}1893 - {\cdot}1893)^{-3{\cdot}6054}$.

II. $400 \times ({\cdot}97051 + {\cdot}02949)^{46{\cdot}2084}$.

III. $400 \times (1{\cdot}0889 - {\cdot}0889)^{-20{\cdot}2473}$.

IV. $400 \times ({\cdot}9525 + {\cdot}0475)^{98{\cdot}5263}$.

II. and IV. may, perhaps, be held fairly to satisfy the conditions, although it is not certain if 46 is to be considered a large n or ·05 a very small q.

I. and III. fail to satisfy the conditions at all, unless the probable errors of q and n are such that q might really be a small positive quantity and n really large and positive. The following are the values for the four series of n and q and their probable errors:

I. $q = -{\cdot}1893 \pm {\cdot}0647$, $n = -\ 3{\cdot}6054 \pm 1{\cdot}2209$.

II. $q = +{\cdot}0295 \pm {\cdot}0457$, $n = \ 46{\cdot}2084 \pm 71{\cdot}7373$.

III. $q = -{\cdot}0889 \pm {\cdot}0534$, $n = -20{\cdot}2473 \pm 12{\cdot}1165$.

IV. $q = +{\cdot}0475 \pm {\cdot}0452$, $n = \ 98{\cdot}5263 \pm 93{\cdot}7494$.

Now while these results are very satisfactory for II. and IV., they are not wholly conclusive for I. and III. We can approach the matter from another standpoint; the probable error of q for $p = 1$ is

$$\cdot 67449 \frac{1}{\sqrt{N}} \sqrt{2} = \cdot 67449 \times \cdot 0707$$

in "Student's" cases. Thus the deviation of q from q a very small quantity is for I. 2·68 times the S. D., and for III. 1·26 times the S. D. Since q may be either positive or negative, we may reasonably apply the probability tables and the odds against deviations occurring as great as these are in one trial about 250 to 1 and 9 to 1 respectively. Hence in four trials we should still have large odds against their combined appearance.

* *Biometrika*, Vol. v. p. 356.

We have said that the results for II. and IV. are fairly satisfactory, *i.e.* we mean that they are consistent with q being small and positive and n being large; but of course they are also consistent with q being negative and n being small and negative.

It will be obvious from these results for "Student's" data that it is extremely difficult to test the legitimacy of the hypothesis on which the "Law of Small Numbers" is based. In none of the cases dealt with by Bortkewitsch, much less in those dealt with by Mortara, are the populations (N) anything like as extensive as those considered by "Student." But populations of even 400 give, as we see, too large values of the probable errors of q and n for us to be certain of our conclusions.

(8) *Bortkewitsch's Cases.* Taking Bortkewitsch next, he deals with the following cases:

I. Suicides of Children in Prussia for 25 years: (*a*) Boys, (*b*) Girls, 25 cases.

II. Suicides of Women in eight German States for 14 years: 112 cases or 8 subseries of 14.

III. Accidental Deaths in 11 Trade Societies in 9 years: 99 cases, or 11 subseries of 9.

IV. Deaths from the Kick of a Horse in 14 Prussian Army Corps for 20 years: 280, or, as Bortkewitsch, 200 cases.

It will be noted at once that Bortkewitsch's populations (N) are far too small for any effective determination of the legitimacy of his application of Poisson's formula to his data.

We take his cases in order:

I. (*a*) *Suicides of Boys.*

TABLE III.

Number of Suicides ...	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7* and over
Number of Years ...	4	8	5	3	4	0	1	0

The binomial is:

$$25\,[1{\cdot}2033 - {\cdot}2033]^{-9{\cdot}6425}.$$

Mean 1·9600 and $\mu_2 = 3{\cdot}2584$.

We have $q = -{\cdot}2033 \pm {\cdot}2421$, $n = -9{\cdot}6425 \pm 10{\cdot}9416$.

If q were really zero its probable error would be $\pm {\cdot}1908$. Clearly 25 cases are wholly inadequate to test the legitimacy of applying the Poisson-Exponential to the frequency*. But to what extent is the reader made conscious by Bortkewitsch that his cases fail entirely to demonstrate the legitimacy of applying his hypotheses?

* The χ^2 for the binomial is 2·379 and for the exponential 2·836, showing a somewhat better result for the binomial.

I. (*b*) *Suicides of Girls.*

TABLE IV.

Number of Suicides ...	0	1	2	3
Number of Years ...	15	9	1	0

The binomial is:

$$25\,[\cdot7418 + \cdot2582]^{1\cdot7041}.$$

$$\text{Mean} = \cdot4400 \text{ and } \mu_2 = \cdot3264.$$

We find $q = \cdot2582 \pm \cdot1012$, $n = 1\cdot7041 \pm \cdot7850$.

As in the case of the boys' suicides, if q were practically zero its probable error would be $\pm$ ·1908, and there is nothing in this result again to justify us in asserting that q is indefinitely small and n indefinitely large.

Actually we have:

TABLE V.

Number of Suicides per Year.

	0	1	2	3
Actual	15	9	1	0
Bortkewitsch ...	16·1	7·1	1·8	—
Binomial (*a*) ...	15·0	8·9	1·1	—
Binomial (*b*) ...	15·2	8·7	1·1	—

(*a*) is the binomial considered above, (*b*) is the binomial obtained by taking n a whole number = 2, and q = mean/2 = ·22, *i.e.* $25\,(\cdot78 + \cdot22)^2$.

It is clear that either (*a*) or (*b*) gives better results than the Poisson-Exponential. Applying the test of goodness to fit, we have

$\chi^2 = \cdot007$ for the binomial (*a*),

$\chi^2 = \cdot610$ for Bortkewitsch's solution.

Both give $P > \cdot60$ but the first is much better than the second.

If both boys and girls are taken together, we find the binomial

$$25\,(\cdot9333 + \cdot0667)^{36}.$$

This is the nearest approach to a small q and big n we have so far found—*i.e.* the nearest approach so far to an exponential, but it is reached by a process, *i.e.* that of adding together two series of entirely different means and variabilities in a manner which cannot be justified, for Bortkewitsch's hypothesis depends essentially on the *homogeneity* of his material. Even here the fit of the point binomial is slightly better than that of the exponential.

II. *Suicides of Women in Eight German States.* Bortkewitsch gives the following table:

TABLE VI.

State	Number of Suicides of Women per Year											Totals
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
(*a*) Schaumburg-Lippe ...	4	4	2	4	—	—	—	—	—	—	—	14
(*b*) Waldeck	1	4	3	4	1	1	—	—	—	—	—	14
(*c*) Lübeck	1	3	2	4	3	1	—	—	—	—	—	14
(*d*) Reuss ä. L.	1	3	3	3	2	1	1	—	—	—	—	14
(*e*) Lippe	2	3	1	2	3	1	2	—	—	—	—	14
(*f*) Schwarzburg-Rudolstadt ...	—	1	—	2	—	5	3	2	1	—	—	14
(*g*) Mecklenburg-Strelitz ...	—	1	2	1	4	—	1	—	2	2	1	14
(*h*) Schwarzburg-Sonderhausen	—	—	4	—	2	2	1	—	—	3	2	14
Totals	9	19	17	20	15	11	8	2	3	5	3	112

The resulting binomials are:

$$(a)\quad 14(\cdot 9714 + \cdot 0286)^{50\cdot 0024},$$
$$(b)\quad 14(\cdot 8571 + \cdot 1429)^{15\cdot 4996},$$
$$(c)\quad 14(\cdot 5819 + \cdot 4181)^{6\cdot 1503},$$
$$(d)\quad 14(1\cdot 0058 - \cdot 0058)^{-456\cdot 2944},$$
$$(e)\quad 14(1\cdot 3929 - \cdot 3929)^{-7\cdot 2727},$$
$$(f)\quad 14(\cdot 6071 + \cdot 3929)^{13\cdot 0909},$$
$$(g)\quad 14(1\cdot 5792 - \cdot 5792)^{-9\cdot 1267},$$
$$(h)\quad 14(1\cdot 6609 - \cdot 6609)^{-8\cdot 5376}.$$

Thus it will be seen that of the eight binomials only four have a positive q, and of these only *one* can be said to have a very small q, and even in this case the n is not indefinitely large. Of the four negative binomials three have quite substantial q's, and the fourth with its small negative q corresponds most closely to the Poisson-Exponential. The probable error of q for $q=0$ is $\pm \cdot 2549$. The number, 14, of cases taken is therefore wholly inadequate to test whether the Poisson-Exponential may be applied to these data. The mean value of q is negative and $= -\cdot 0820 \pm \cdot 0901$, and the standard deviation of $q = \cdot 3928 \pm \cdot 0637$, which are within the limits of random sampling of $q=0$ with a standard deviation of $\cdot 3779$. We shall return to a different manner of considering the point later. At present we wish only to indicate that the hypothesis is that q is a very small positive quantity and that data which give q a standard deviation of $\cdot 3928$, or in the next example of $\cdot 4714$ are really inadequate to test such a hypothesis; for in the resulting binomials q may easily lie anywhere between $+\cdot 8$ and $-\cdot 8$, and it is not possible to demonstrate that its real value is practically an exceeding small positive quantity.

III. *Accidental Deaths in* 11 *Trade Societies.* Bortkewitsch provides data from which the following table is deduced:

TABLE VII.

Index Number of Society	Accidental Deaths															Totals
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	
13	—	—	—	—	1	1	1	1	3	1	—	—	—	—	1	9
14	—	2	3	2	1	1	—	—	—	—	—	—	—	—	—	9
12	2	1	3	—	1	1	—	1	—	—	—	—	—	—	—	9
20	—	—	1	3	2	2	—	—	—	—	1	—	—	—	—	9
23	—	—	—	1	2	1	2	—	1	1	—	1	—	—	—	9
27	—	4	3	1	1	—	—	—	—	—	—	—	—	—	—	9
29	—	—	—	2	3	—	—	1	2	—	—	—	1	—	—	9
41	—	1	—	1	1	2	1	2	1	—	—	—	—	—	—	9
40	2	1	2	1	—	1	1	1	—	—	—	—	—	—	—	9
42	1	—	—	1	1	4	1	—	1	—	—	—	—	—	—	9
55	—	—	2	1	1	3	1	1	—	—	—	—	—	—	—	9
Totals ...	5	9	14	13	14	16	7	7	8	2	1	1	1	—	1	99

The resulting binomials are:

$$(13)\quad 9(\cdot4914+\cdot5086)^{15\cdot5108},$$
$$(14)\quad 9(\cdot6184+\cdot3816)^{6\cdot6962},$$
$$(12)\quad 9(1\cdot9227-\cdot9227)^{-2\cdot7696},$$
$$(20)\quad 9(1\cdot1282-\cdot1282)^{-33\cdot8000},$$
$$(23)\quad 9(\cdot9921+\cdot0079)^{784\cdot0502},$$
$$(27)\quad 9(\cdot5229+\cdot4771)^{3\cdot9589},$$
$$(29)\quad 9(1\cdot4130-\cdot4130)^{-14\cdot2589},$$
$$(41)\quad 9(\cdot8454+\cdot1546)^{23\cdot0626},$$
$$(40)\quad 9(2\cdot0342-1\cdot0342)^{-2\cdot7934},$$
$$(42)\quad 9(\cdot9322+\cdot0678)^{67\cdot2397},$$
$$(55)\quad 9(\cdot6154+\cdot3846)^{11\cdot2667}.$$

Of these eleven binomials seven have a positive q; only one of these (23) actually corresponds to a really small q and large n, although a second, (42), approximates to this condition. In the five other cases the q's are quite substantial; in (13) the q is larger than p. Of the four negative q's none can be said to be so small and the n so large as to suggest that they really correspond to the Poisson-Exponential. The probable error of q for $q = 0$ is, however, $\pm$ ·3180, and thus for such small series, no test whatever can be really reached of the legitimacy of applying the Poisson-Exponential to such data. We may note, indeed, that seven of the eleven values of q exceed the probable error and two of these are more than three times the probable error. We should only expect *two* negative values of q as great or greater than ·9227 in 80 trials, whereas two have occurred in 9 trials,

so that the odds are considerably against such an experience. The mean value of q is $-\cdot0469 \pm \cdot0959$ and the standard deviation of q is $\cdot5127 \pm \cdot0678$, both results compatible with q indefinitely small and a standard deviation $= \cdot4714$. The main problem, however, of the legitimacy of applying the Poisson-Exponential to such series cannot be answered by data involving only total frequencies of 9 to 14 cases in the individual series.

Bortkewitsch examines the matter from another standpoint. He clubs the results given for each application of the Poisson-Exponential together and examines the observed totals against the sums of the calculated totals. Thus calculating the 11 Poisson-Exponential series* and adding them together Bortkewitsch finds for observed and calculated deaths:

TABLE VIII.

Accidental Deaths in 11 *Trade-Societies.*

Number of Deaths	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13* & over	Totals
Observed Frequencies	5	9	14	13	14	16	7	7	8	2	1	1	1	1	99
Sums of 11 Exponentials	3·7	9·6	13·9	15·2	14·3	12·3	9·8	7·3	5·8	3·3	2·0	1·2	0·7	0·6	99
Single Binomial ...	3·8	9·5	13·9	15·6	14·8	12·4	9·6	6·9	4·8	3·1	2·0	1·2	0·7	0·7	99

If we attempt to fit a *single* binomial to the observed line of totals, we obtain:

$$m = 4\cdot3636, \quad \sigma^2 = 7\cdot5849$$

leading to the negative binomial:

$$99\,(1\cdot7382 - \cdot7382)^{-5\cdot9111}.$$

Here: $q = -\cdot7382 \pm \cdot1829$†, $n = -5\cdot9111 \pm \cdot1391$,

or the constants are significantly substantial with regard to their probable errors. The resulting frequencies are given in the last line of the table above. The reader

* The values of the means and standard deviations for the eleven societies are:

	m	σ		m	σ		m	σ
13	7·889	1·969	*23*	6·222	2·485	*40*	2·889	2·424
14	2·556	1·343	*27*	1·889	0·994	*42*	4·556	2·061
12	2·556	2·217	*29*	5·889	2·885	*55*	4·333	1·633
20	4·333	2·211	*41*	5·111	2·079			

All these means are less than 10, which is the limit reached by Bortkewitsch's Tables for the Poisson-Exponential. Bortkewitsch says he has taken the societies for which "the statistics indicated the smallest numbers of such accidents." This is not very clear. It is certain that a society with a mean number of accidents $=100$, if it consisted of 200,000 members, would be more suitable for application of the exponential, than one with a mean of 8 if it only contained 10,000 members. Both Bortkewitsch and Mortara confine their results to means less than 10, and seem to indicate that "smallness" has been determined by the absolute frequencies, but clearly it is relative frequency with which we have to deal. The use of such a term as *Das Gesetz der kleinen Zahlen* for the Poisson-Exponential seems open to serious objection, if it be associated with "m" an absolutely small number, and not with smallness of "q."

† For $q=0$, the probable error would be $\pm\cdot0959$ and accordingly q is very divergent from the Poisson-Exponential value of zero.

will be surprised to see how closely the single negative binomial determined by *two* constants gives the same result as the sum of the eleven Poisson-Exponentials determined by *eleven* constants, no one of which is really of any significance for its own exponential*. If we apply the condition for "goodness of fit," $\chi^2 = 5{\cdot}83$ for the single binomial and $\chi^2 = 5{\cdot}88$ for the sum of the eleven Poisson exponentials, leading to $P = {\cdot}950$ and $P = {\cdot}951$ respectively, or the fit with a single negative binomial is slightly better than that with eleven exponentials. The two constants are significant, the eleven constants have no real significance for their individual series, as is demonstrated by the fact that the binomials for these series do not approximate to the Poisson-Exponential type.

We may now consider the previous case of suicides of women from the same standpoint†. The following are the data as given by Bortkewitsch:

TABLE IX.

Suicides of Women in Eight German States.

Number of Suicides	0	1	2	3	4	5	6	7	8	9	10 & over	Totals
Observed Frequencies	9	19	17	20	15	11	8	2	3	5	3	112
Sum of 8 Exponentials	8·0	16·9	20·3	18·7	15·1	11·4	8·3	5·6	3·6	2·1	2·0	112
Single Binomial ...	12·6	18·4	18·8	16·4	13·2	9·9	7·2	5·1	3·5	2·4	4·5	112

For the single binomial we have:

$$m = 3{\cdot}4732, \quad \sigma^2 = 8{\cdot}2312,$$

leading to:

$$112\,(2{\cdot}3699 - 1{\cdot}3699)^{-2{\cdot}5354},$$

where

$$q = -1{\cdot}3699 \pm {\cdot}1490, \quad n = -2{\cdot}5354 \pm {\cdot}3076.$$

If q were very small its probable error would be $\pm {\cdot}0901$. The values of q and n are quite significant, q is large and negative and n is small and negative. The resulting frequencies are given in the last line of the table as "Single Binomial." Turning now to the test of "goodness of fit," we have for the sum of the 8 exponentials $\chi^2 = 7{\cdot}957$, and for the single binomial $\chi^2 = 7{\cdot}740$, leading to $P = {\cdot}633$

* If the reader will turn to the first footnote on p. 53 he will note that for *nine* cases, the standard deviations of the means ($\sigma/\sqrt{9}$) are roughly about ·7 or errors of ± 1 to $\pm 1{\cdot}5$ may easily occur in the means. Hence with the possible exception of (13) and (27) the m's have not significant differences, and are not typical of the individual societies.

† The values of the means and standard deviations are:

	m	σ		m	σ
Schaumburg-Lippe	1·429	1·178	Lippe	2·857	1·995
Waldeck ...	2·214	1·378	Schwarzburg-Rudolstadt ...	5·143	1·767
Lübeck	2·571	1·223	Mecklenburg-Strelitz ...	5·286	2·889
Reuss ä. L. ...	2·643	1·631	Schwarzburg-Sonderhausen	5·642	3·061

The standard deviation of the mean is here $\sigma/\sqrt{14}$, or, say, ·5. Thus errors of 1 might easily occur in the values of m. There are probably significant differences between the first five and the last three states, but not between the first five among themselves or the last three among themselves. Thus the Poisson-Exponentials, if correct in theory, are not significant for the individual states.

and ·654 respectively. Thus again the single binomial with only two constants give a fit slightly better, than the sum of eight exponentials with eight constants.

Bortkewitsch looking at the observed frequencies and the sum of 8 or 11 exponentials—without using any satisfactory test for "goodness of fit"—assumes that the coincidence is so good as to justify his hypothesis. But a better fit can be obtained with two instead of 8 or 11 constants by simply using a negative binomial. We must note here that Bortkewitsch is using the final coincidence merely as justification of the Poisson-Exponential; the total frequency is not describable in terms of the 8 or 11 constants as it is in terms of the two, for these eight constants are not really significant for his individual eleven trade societies or for the suicides in the individual eight states. If he wants to describe the total, he has no constants by which he can do it. If, on the other hand, he wishes to describe what has occurred in the individual societies or states, we have seen that their binomials differ very widely from Poisson-Exponentials. If, lastly, no stress be laid on the individual cases as having too large probable errors, but only on the general coincidence with total frequencies, then the same coincidence would justify us in using a single binomial with two constants only*. It appears to us that to properly test the Poisson-Exponential, we need not 9 or 14 instances in the individual case, but several hundred instances,—more, indeed, than "Student" has taken—and that no proof of the "Law of Small Numbers" can be obtained on data such as those of Bortkewitsch or Mortara.

IV. *Deaths from the Kick of a Horse in Prussian Army Corps, omitting four Corps with Bortkewitsch.*

Here the results are:

TABLE X.

Number of Deaths ...	*0*	*1*	*2*	*3*	*4*	Totals
Number of Corps ...	109	65	22	3	1	200

Whence:

$$m = \cdot 61, \quad \mu_2 = \cdot 6079$$

and the binomial is:

$$200\,(\cdot 996{,}557 + \cdot 003{,}443)^{177\cdot 17107}.$$

This is the first of Bortkewitsch's illustrations for which his hypothesis that q is small and n large is really justified by his data. For:

$$q = \cdot 0034 \pm \cdot 0670,$$
$$n = 177\cdot 1711 \pm 3449\cdot 103.$$

The probable error of q for q really zero is $\pm \cdot 0674$.

* Of course immensely better general total fits are obtained by using the sums of the actual 8 or 11 binomials than by the Poisson-Exponential sum or the single binomial, but the results in that case involve 16 or 22 non-significant constants.

The actual results as given by the binomial and the Poisson-Exponential are:

TABLE XI.

Number of Deaths ...	*0*	*1*	*2*	*3*	*4* and over
Observed	109	65	22	3	1
Binomial	108·6	66·4	20·2	4·1	0·7
Exponential	108·7	66·3	20·2	4·1	0·7

Actually if we work to two decimal places in the frequencies we have $\chi^2 = \cdot 61$ for both binomial and exponential, or the goodness of fit is practically identical.

In this case it seemed worth discussing the binomial fit more at length. Taking the moment coefficients about the mean we have:

(i) Mean $= nq = \cdot 6100.$

(ii) $\mu_2 = npq = \cdot 6079.$

(iii) $\mu_3 = npq\,(p - q) = \cdot 590{,}562.$

(iv) $\mu_4 = npq\,(1 + 3npq - 6pq) = 1 \cdot 643{,}373.$

We have already discussed the binomial from (i) and (ii), giving χ^2 for goodness of fit $= \cdot 6096$. Using (ii) and (iii) we have for the binomial

$$200\,(\cdot 985{,}739 + \cdot 014{,}261)^{43 \cdot 24351},$$

giving $$\chi^2 = \cdot 665.$$

Using (iii) and (iv) we have:

$$200\,(\cdot 979{,}524 + \cdot 020{,}057)^{30 \cdot 3090},$$

giving $$\chi^2 = \cdot 707.$$

Putting: $$\beta_2 = \mu_4/\mu_2^2 \quad \text{and} \quad \beta_1 = \mu_3^2/\mu_2^3,$$

we have: $$\beta_2 - 3 = (1 - 6pq)/npq, \quad \beta_1 = (1 - 4pq)/npq,$$

and working from β_1 and β_2 we find:

$$200\,(\cdot 969{,}150 + \cdot 030{,}850)^{18 \cdot 9645},$$

and in this case $$\chi^2 = 1 \cdot 1286.$$

This of course does not give a bad fit, but it is clear that working from the *lowest moment coefficients*, as we might anticipate, gives the best results.

But if q be the chance of death from the kick of a horse, and n the number of men in an army corps, then the binomial should be

$$200\,(p + q)^n.$$

Now it is obvious that none of the binomials give, by their value of n any approach to the real number of men in an army corps. If we start with the

number of men n in an army corps as 50,000*, we have $nq = \cdot 61$ and $q = \cdot 000,0122$, thus reaching the binomial

$$200\,(\cdot 999,9878 + \cdot 000,0122)^{50,000},$$

giving as compared against Bortkewitsch:

	Binomial	*Bortkewitsch*
0	108·6876	108·6703
1	66·3002	66·2889
2	20·2213	20·2181
3	4·1115	4·1110
4 and over	·7035	·7034
and $\chi^2 =$	·608,298	·608,318

or, the slight advantage to the binomial exists but is of no significance.

Now it seems to us that in this case the use of the exponential is justified for the total frequencies, but as far as describing those frequencies is concerned, it gives no better result than the binomial. But as in the other five of Bortkewitsch's cases the Exponential is not justified by the individual series themselves†.

It is perfectly true that the exponential has a definite theory behind it, and is interpretable in terms of that theory, i.e. we must suppose the probability of an occurrence very small and the chance of its repetition absolutely identical. But is the second of these conditions ever likely to be demonstrable *a priori*, or must

* This supposes that every man in the army corps is equally liable to death from the kick of a horse; of course a very arbitrary assumption.

† To illustrate the idleness of the application of the Poisson-Exponential even to these data for the Prussian Army Corps, we give here the binomials for the whole of the 14 corps.

Index Number of Corps	*Binomial*
G	$20\,(\cdot 95 + \cdot 05)^{16\cdot 0000}$
I	$20\,(1\cdot 325 - \cdot 325)^{-2\cdot 4615}$
II	$20\,(1\cdot 5667 - \cdot 5667)^{-1\cdot 0588}$
III	$20\,(\cdot 9 + \cdot 1)^{6\cdot 0000}$
IV	$20\,(\cdot 6 + \cdot 4)^{1\cdot 0000}$
V	$20\,(\cdot 6318 + \cdot 3682)^{1\cdot 4038}$
VI	$20\,(1\cdot 0912 - \cdot 0912)^{-9\cdot 3202}$
VII	$20\,(\cdot 9 + \cdot 1)^{6\cdot 0000}$
VIII	$20\,(\cdot 65 + \cdot 35)^{1\cdot 0000}$
IX	$20\,(\cdot 8115 + \cdot 1885)^{3\cdot 4483}$
X	$20\,(1\cdot 05 - \cdot 05)^{-15\cdot 0000}$
XI	$20\,(1\cdot 11 - \cdot 11)^{-11\cdot 3636}$
XIV	$20\,(1\cdot 05 - \cdot 05)^{-24\cdot 0000}$
XV	$20\,(1\cdot 1 - \cdot 1)^{-4\cdot 0000}$

One seeks in vain through these binomials for any approach to q very small and positive and n very large and positive. In no case does n approach the number of men in an army corps, say 50,000, or q equal the chance of a death from the kick of a horse, say, ·0000122! It seems impossible by clubbing such equations together to give any satisfactory proof that the Poisson-Exponential really does apply to individual cases. In the 20 years involved, there were doubtless great changes in both the training and the personnel of each army corps, and the results obtained may be just as much due to such causes as to the errors of small samples.

not we *a posteriori* demonstrate it from the data themselves? Child suicide may be influenced by example, by environmental conditions in different districts, possibly even by meteorological conditions in different years. Again, even in different army corps the conditions may be far from uniform, the spirit of the corps, the teaching with regard to the handling of horses, the experience of past life according to whether the corps is raised in town or rural districts may all tell. Even Bortkewitsch before he gets his best fit removes four corps or 80 observations from his data. We do not criticise this removal, but even unremoved he says the fit of theory with experience leaves "wie man sieht, nichts zu wünschen übrig" (p. 25). But the binomial is before removal:

$$280\,(1{\cdot}085{,}714 - {\cdot}085{,}714)^{-8{\cdot}166{,}694}$$

in which q is not very small and is negative, and n is not very large and is not positive. It is true that the probable error of q for q insignificant is in this case $\pm$ ·0570, but this only shows that the data were insufficient in quantity to determine whether the exponential could be applied or not.

(9) *Mortara's Cases.*

Mortara* in an interesting paper has realised the possibility of repetitions not being independent and has discussed a constant Q', by which he proposes to test such influence. This quantity Q' should be unity, if the Bortkewitschian hypothesis can be applied. He then takes 16 or 17 districts with records of 10 years, and calculates the mean number of deaths from some special cause per year, say, for each district for those years. If this mean number exceeds 10, he casts out that district, presumably on the ground either (i) that such a number is no longer small, or (ii) that it differentiates the district from those with lower numbers. Thus Bologna with 10·9 deaths by murder is excluded and Bergamo with 8·4 is included, although $Q' = 1$ for both. Bologna with 7·1 deaths from smallpox is included, but Pavia with 12·3 is excluded although the Q' of the former is 2·5 and that of the latter 1·7. What method should be employed in dealing with the frequency of the excluded districts which may amount to 50 °/₀ of all districts is not discussed. Having thus reduced his available districts, Mortara proceeds to apply the exponential to each *individual* district; he adds up the results for each district and compares his totals with the observed totals. It will thus be observed that he fits his exponential to *ten* observations, and then adds together five or more districts to get his totals. We can equally well apply this process by fitting a binomial to each 10 observations and then adding up such results. But it is quite clear that on the basis of *ten* observations, it is, owing to the large probable errors, wholly impossible to assert, whether a binomial of the kind required by the Bortkewitsch-Mortara hypothesis,—i.e. one of very small positive q and very large positive n—really is justified. We can illustrate this at once from Mortara's Tables (see his pp. 42 and 45) for deaths from Chronic Alcoholism. The

* "Sulle variazioni di frequenza di alcuni fenomeni demografici rari," *Annali di Statistica*. Serie v. Vol. IV. pp. 5—81. Roma, 1912.

observed numbers, and those deduced from the binomials are given in the accompanying table. At the foot are the observed totals, Mortara's exponential totals and the binomial totals.

TABLE XII. *Deaths from Chronic Alcoholism.*

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14* & over	
Calabria	1	3	4	—	2	—	—	—	—	—	—	—	—	—	—	Observed
	1·49	2·84	2·70	1·71	·81	·31	·10	·03	·01	—	—	—	—	—	—	Mortara
	1·18	2·85	3·06	1·91	·76	·20	·03	—	—	—	—	—	—	—	—	Binomial
Foggia	1	2	4	—	2	1	—	—	—	—	—	—	—	—	—	O.
	1·00	2·30	2·65	2·03	1·17	·54	·21	·07	·02	·01	—	—	—	—	—	M.
	·96	2·29	2·70	2·08	1·18	·53	·19	·06	·01	—	—	—	—	—	—	B.
Siracusa	2	1	3	—	2	2	—	—	—	—	—	—	—	—	—	O.
	·82	2·05	2·56	2·14	1·34	·67	·28	·10	·03	·01	—	—	—	—	—	M.
	1·12	2·16	2·33	1·85	1·21	·69	·35	·17	·07	·03	·01	—	—	—	—	B.
Potenza	2	—	2	2	1	1	1	1	—	—	—	—	—	—	—	O.
	·41	1·30	2·09	2·23	1·78	1·14	·61	·28	·11	·04	·01	—	—	—	—	M.
	·78	1·61	1·95	1·80	1·41	·98	·63	·38	·21	·12	·06	·03	·01	·01	—	B.
Catanzaro	1	1	3	1	—	—	1	1	1	—	—	1	—	—	—	O.
	·15	·63	1·32	1·85	1·95	1·63	1·14	·69	·36	·17	·07	·03	·01	—	—	M.
	·88	1·35	1·46	1·36	1·17	·95	·75	·57	·43	·31	·23	·16	·12	·08	·17	B.
Salerno	1	1	1	—	2	—	1	1	2	—	—	1	—	—	—	O.
	·06	·31	·79	1·35	1·72	1·75	1·49	1·09	·69	·39	·20	·09	·04	·02	·01	M.
	·40	·86	1·18	1·31	1·27	1·14	·95	·77	·59	·45	·33	·24	·17	·12	·22	B.
Cosenza	2	—	1	—	1	—	3	1	—	1	—	—	1	—	—	O.
	·06	·29	·75	1·29	1·68	1·75	1·51	1·12	·73	·42	·22	·10	·05	·02	·01	M.
	·43	·88	1·17	1·27	1·23	1·10	·93	·75	·59	·44	·33	·24	·17	·13	·33	B.
Bologna	—	—	3	1	1	1	1	—	—	—	1	—	—	—	2	O.
	·01	·06	·21	·49	·88	1·24	1·47	1·49	1·32	1·04	·74	·48	·28	·15	·14	M.
	·40	·76	·97	1·06	1·05	·98	·87	·76	·64	·53	·43	·35	·28	·21	·71	B.
Totals	10	8	21	4	11	5	7	4	3	1	1	2	1	—	2	O.
	4 00	9·78	13·07	13·09	11·33	9·03	6·81	4·87	3·27	2·08	1·24	·70	·38	·19	·16	M.
	6·15	12·75	14·82	12·64	9·28	6·57	4·70	3·46	2·54	1·88	1·38	1·02	·75	·55	1·43	B.

The following are the binomials for the 8 districts out of 16 which Mortara has selected.

Reggio Calabria	$10\,(\ \cdot7842 + \cdot2158)^{+8\cdot8049}$
Foggia	$10\,(\ \cdot9609 + \cdot0391)^{58\cdot7709}$
Siracusa	$10\,(1\cdot3000 - \cdot3000)^{-8\cdot3223}$
Potenza	$10\,(1\cdot5500 - \cdot5500)^{-5\cdot8182}$
Catanzaro	$10\,(2\cdot7524 - 1\cdot7524)^{-2\cdot3907}$
Salerno	$10\,(2\cdot3510 - 1\cdot3510)^{-3\cdot7750}$
Cosenza	$10\,(2\cdot5308 - 1\cdot5308)^{-3\cdot2970}$
Bologna	$10\,(3\cdot3161 - 2\cdot3161)^{-2\cdot6769}$

Examining these we see that there are only *two* in which q and n are positive and only *one* in q is small and positive and n moderately large. The probable error of q for 10 observations on the assumption that n is very large and q very small is $\pm$ ·3016 and is quite inconsistent with the last four districts being samples from exponentially distributed frequencies. The other four districts may or may not belong to such frequencies—the data are wholly inadequate to determine whether they do or not. Reggio Calabria and Foggia have the lowest Q's, i.e. 0·9 and 1·0. But that six districts out of an *already selected* eight give *negative* q and a seventh a relative large q and small n suggests the inapplicability of the hypothesis adopted. If we seek for "goodness of fit" of the totals, we find:

	Binomial	*Exponential*
$\chi^2 =$	25·12	47·92
$P =$	·0336	·0000

Thus the odds against the binomial system are 28 to 1, but the odds against the exponential are enormous. It does not seem possible to justify the treatment of such data by the use of the Poisson-Exponential.

Let us turn to a second of Mortara's illustrations, that of deaths from small-pox. He rejects first six out of the 17 districts, the remaining ten are given in Table XIII. The districts give the following binomials:

Venezia	$10\,(\ \cdot9500 + \ \cdot0500)^{16}$
Bologna	$10\,(\ \cdot9889 + \ \cdot0111)^{81}$
Treviso	$10\,(\ 2\cdot2000 - \ 1\cdot2000)^{-\cdot8333}$
Pavia	$10\,(\ 1\cdot8000 - \ \cdot8000)^{-1\cdot5000}$
Cagliari	$10\,(\ 4\cdot5190 - \ 3\cdot5190)^{-\cdot5968}$
Padova	$10\,(\ 3\cdot6833 - \ 2\cdot6833)^{-\cdot8944}$
Verona	$10\,(\ 5\cdot6000 - \ 4\cdot6000)^{-\cdot5217}$
Brescia	$10\,(\ 9\cdot9727 - \ 8\cdot9727)^{-\cdot3678}$
Bergamo	$10\,(\ 2\cdot3821 - \ 1\cdot3821)^{-2\cdot8219}$
Catanzaro	$10\,(15\cdot6128 - 14\cdot6128)^{-\cdot2669}$
Vicenza	$10\,(\ 3\cdot4854 - \ 2\cdot4854)^{-1\cdot6497}$

Out of the eleven cases only two give q small and positive; not *a single one* gives for q anything like the chance of a death from small-pox in the district, nor for n anything like the population of the district. There is an increasing divergence from the positive binomial as Mortara's Q' increases in value. We see that in nine cases, however, a negative binomial not the exponential is required to describe the frequencies. The probable error of q, for insignificant q is as before $\pm$ ·3016, and therefore it is improbable that q is zero in at least 9 out of these 11 districts.

Examining the totals we find

	Binomial	*Exponential*
$\chi^2 =$	9·64	570·79
$P =$	·67	·000,000

TABLE XIII.

Deaths from Small-pox (1900—1909).

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12* or more	
Venezia	4	5	—	1	—	—	—	—	—	—	—	—	—	Observed
	4·49	3·60	1·44	·38	·08	·01	—	—	—	—	—	—	—	Mortara
	4·40	3·71	1·46	·36	·06	·01	—	—	—	—	—	—	—	Binomial
Bologna	4	4	1	1	—	—	—	—	—	—	—	—	—	O.
	4·07	3·66	1·65	·49	·11	·02	—	—	—	—	—	—	—	M.
	4·04	3·68	1·65	·49	·11	·02	·01	—	—	—	—	—	—	B.
Treviso	5	3	1	—	—	1	—	—	—	—	—	—	—	O.
	3·68	3·68	1·84	·61	·15	·03	·01	—	—	—	—	—	—	M.
	5·18	2·36	1·18	·61	·32	·17	·09	·05	·03	·01	—	—	—	B.
Pavia	4	3	2	—	—	1	—	—	—	—	—	—	—	O.
	3·01	3·62	2·17	·87	·26	·06	·01	—	—	—	—	—	—	M.
	4·14	2·76	1·53	·79	·40	·19	·09	·05	·02	·01	·01	—	—	B.
Cagliari	5	1	1	1	—	1	—	—	—	—	1	—	—	O.
	1·23	2·57	2·70	1·89	·99	·42	·15	·04	·01	—	—	—	—	M.
	4·07	1·89	1·17	·79	·55	·39	·28	·21	·15	·11	·08	·06	·25	B.
Padova	3	3	—	2	—	1	—	—	—	—	1	—	—	O.
	·91	2·18	2·61	2·09	1·25	·60	·24	·08	·03	·01	—	—	—	M.
	3·12	2·03	1·40	·98	·70	·50	·36	·26	·19	·13	·10	·07	·16	B.
Verona	4	3	—	1	—	—	1	—	—	—	—	—	1	O.
	·91	2·18	2·61	2·09	1·25	·60	·24	·08	·03	·01	—	—	—	M.
	4·07	1·74	1·09	·75	·54	·40	·31	·23	·18	·14	·11	·09	·35	B.
Brescia	2	3	2	2	—	—	—	—	—	—	—	—	1*	O.
	·37	1·22	2·01	2·21	1·82	1·20	·66	·31	·13	·05	·02	—	—	M.
	4·29	1·42	·87	·62	·47	·37	·30	·24	·20	·17	·14	·12	·79	B.
Bergamo	2	—	2	2	—	1	—	1	1	1	—	—	—	O.
	·20	·79	1·54	2·00	1·95	1·52	·99	·55	·27	·12	·04	·02	·01	M.
	·86	1·41	1·57	1·46	1·23	·98	·74	·54	·38	·27	·18	·12	·24	B.
Catanzaro	3	3	1	1	1	—	—	—	—	—	—	—	1*	O.
	·20	·79	1·54	2·00	1·95	1·52	·99	·55	·27	·12	·04	·02	·01	M.
	4·80	1·20	·71	·50	·38	·31	·25	·21	·18	·16	·14	·12	1·04	B.
Vicenza	3	—	1	1	1	1	—	1	1	—	—	—	1	O.
	·17	·68	1·39	1·91	1·95	1·60	1·09	·64	·33	·15	·06	·02	·01	M.
	1·28	1·50	1·42	1·23	1·02	·82	·65	·51	·39	·30	·23	·17	·48	B.
Totals	39	28	11	12	2	6	1	2	2	1	2	—	4	O.
	19·24	24·97	21·50	16·54	11·76	7·58	4·38	2·25	1·07	·46	·16	·06	·03	M.
	40·25	23·70	14·05	8·58	5·78	4·16	3·08	2·30	1·72	1·20	·99	·75	3·31	B.

* 1 at '12 or more' in cases of Brescia and Catanzaro was found to signify 1 at 20 in the case of Brescia, and 1 at 27 in case of Catanzaro, if the means were to agree with those given by Mortara.

In other words the binomials give a reasonable total fit, the exponentials a practically impossible one.

But there is another question to be asked in such series as those of Mortara: What justification is there in cutting off at 10 cases, say of murder? A province may have a million inhabitants and, perhaps, 40 murders occur in a year*. Hence the binomial is for ten year returns

$$10 \times \left(\frac{24{,}999}{25{,}000} + \frac{1}{25{,}000}\right)^{1{,}000{,}000}$$

but this is as close as anything can be desired to the exponential series. It may be reasonable to apply a separate series to districts giving 4·2 and 36·6 murders per annum respectively, but it is difficult to see why the latter district should be altogether excluded from treatment. If the theory of the binomial be applicable *at all*, then it applies practically as well to districts with 40 murders as to districts with 4; for, we need no indefinitely small q to get a closely exponential series. If we take the case of deaths by murder, Mortara has retained only 6 out of 16 provinces, yet his criterion Q' (see his Table, p. 51) is not more divergent from unity for the rejected provinces than for those retained; the binomials are indeed

Reggio Treviso	$10(\cdot7000+\cdot3000)^{9\cdot3333}$
Venezia	$10(\cdot5619+\cdot4381)^{9\cdot5869}$
Vicenza	$10(\cdot9571+\cdot0429)^{114\cdot2191}$
Padova	$10(\cdot4774+\cdot5226)^{11\cdot3638}$
Pavia	$10(1\cdot8162-\cdot8162)^{-9\cdot0064}$
Bergamo	$10(\cdot8857+\cdot1143)^{73\cdot4908}$

only one of which gives q small and positive and n large.

The mean Q' for the retained provinces is ·967 with a range from ·7 to 1·4 and for the rejected 1·03 with a range from ·8 to 1·4. Even if—which is not the case—the probability of an individual being murdered were too great for the exponential, it ought to follow the binomial, but this, as a rule, it does not do, unless we give some wholly new interpretations to q and n; the actual values render the theory of the binomial as stated inapplicable.

(10) *Mortara's Criterion.*

As a matter of fact the only test of whether an exponential will legitimately fit a given series or not is to determine the binomial $(p+q)^n$ and ascertain whether p is slightly less than unity. But:

$$p = npq/nq$$
$$= \frac{(\text{Standard Deviation})^2}{\text{Mean}}.$$

* We assume that each individual is equally likely to be murdered. But if there be a graduated probability for murder throughout the community, what right have we to apply Poisson's series at all? The essential basis of the application—equal chance of each individual—is wanting.

Now if m_s be the number of deaths, say, occurring in any year and there be l years under consideration, then:

$$(\text{Standard Deviation})^2 = \frac{S_1^2 (m_s - nq)^2}{l},$$

or, if we use the form preferred by Bortkewitsch*

$$= \frac{S_1^2 (m_s - nq)^2}{l-1}.$$

Hence:
$$p = \frac{S_1^2 (m_s - nq)}{(l-1)\, nq}.$$

This in other notation is Mortara's Q'^2, the only criterion he actually uses provided by his equation (17 *ter*), p. 18. Thus his Q', which he says must not differ much from 1, is only $\sqrt{p}$, and it would be better to use p—which has a direct physical meaning—than Mortara's $Q' = \sqrt{p}$. Clearly Mortara's somewhat elaborate process of deducing Q', does not amount to more than saying: Fit a point binomial and test if p is slightly less than unity. We contend that it is best straight off to fit the binomial.

It is true that Mortara does not reach his Q'^2, our p, by the simple process of asking whether the binomial is one with a positive probability less than unity. He endeavours to obtain it by considering whether there is "lumpiness" in the observations. But it seems to us clearer and briefer to ask: Are the contributory cause-groups independent as in teetotum spinning? If so, the data will fit a true binomial and p will of necessity be a positive quantity less than unity. If they are not of this character then p must of necessity be greater than unity. It is of interest to see how Mortara's test of dependence of contributory cause groups leads to a criterion, but he actually only gets his Q'^2, i.e. our binomial p after a series of hypotheses which much limit, and that in no very obvious manner,

* The use of $\sqrt{l}$ or $\sqrt{l-1}$ in the value of the standard deviation when l is small has been several times discussed. It may be dealt with as follows: The probable errors of a mean as deduced by the two processes are

$$E = \cdot 67449 \,.\, \sigma/\sqrt{l},$$

and
$$E' = \cdot 67449 \,.\, \sigma/\sqrt{l-1},$$

now
$$E' = \cdot 67449\, \sigma/\sqrt{l}\left(1 + \frac{1}{2l} + \ldots\right)$$

$$= \cdot 67449\, \frac{1}{\sqrt{l}}\left(\sigma + \frac{1}{\sqrt{2l}}\,\frac{\sigma}{\sqrt{2l}} + \ldots\right).$$

Now the probable error of σ is $\cdot 67449\, \frac{\sigma}{\sqrt{2l}}$, and $\frac{1}{\sqrt{2l}}$ is less and often much less than $\cdot 67449$. Hence if we only know σ from the observations themselves, and this is the usual case, we have:

$$E' = \cdot 67449\, \frac{1}{\sqrt{l}}\, \sigma',$$

where σ' differ from σ by a quantity usually far less than the probable error of σ. In other words the refinement of using E' for E is idle having regard to the accuracy of our observations; and the form used by Bortkewitsch and Mortara with $\sqrt{l-1}$ for $\sqrt{l}$ is of no importance.

the nature of those contributory causes groups. Of course if their dependence were of the nature of successive draws from a pack, then the result would be a hypergeometrical series and Q'^2 would have no physical meaning for the series at all.

(11) We will deal with one further illustration out of many considered by Mortara which are of like character. In the case of Marriages of Uncle and Niece (see Table XIV, p. 65), where the distribution of Q's is the most favourable for his theory, the binomials are

Reggio Marche	$10(\cdot7000 + \cdot3000)^{1\cdot0}$
Umbria	$10(\cdot9000 + \cdot1000)^{5\cdot0}$
Basilicata	$10(1\cdot4000 - \cdot4000)^{-1\cdot5}$
Sardegna	$10(\cdot44545 + \cdot55455)^{1\cdot9836}$
Emilia	$10(\cdot9818 + \cdot0182)^{120\cdot1000}$
Abruzzi	$10(\cdot8429 + \cdot1571)^{17\cdot6182}$
Lazio	$10(1\cdot2548 - \cdot2548)^{-12\cdot1646}$
Puglie	$10(1\cdot5111 - \cdot5111)^{-7\cdot0435}$
Veneto	$10(1\cdot3444 - \cdot3444)^{-1\cdot30645}$
Toscana	$10(2\cdot2667 - 1\cdot2667)^{-4\cdot26315}$
Calabria	$10(1\cdot3584 - \cdot3584)^{-24\cdot8305}$

of which only one (Emilia) approaches the conditions for an exponential distribution. If we test the totals at the foot of Table XIV, we find the result much to the advantage of the binomial, for which $P = \cdot902$ as against $\cdot714$ for the exponential.

(12) On Mortara's own showing nearly all the Q's of his numerous series are greater than unity, and very few of the binomials are positive. If we consider the distribution of Q's, given in his work omitting Table 13 (Deaths from Malaria) we find a range from $\cdot5$ to $3\cdot6$ with a mean Q' at

$$1\cdot2565 \pm \cdot0847,$$

while for the distribution of all the p's in the binomials we have determined, we find a range from $\cdot4$ to $15\cdot6$ with a mean p at $2\cdot5655 \pm \cdot3817$.

These results are sufficient to show that there is no real distribution of p round the value unity but the binomials have a distinct tendency to be *negative*.

(13) But the whole theory of Poisson's exponential law in the hands of Bortkewitsch and Mortara appears essentially vague. The binomial is built up on the assumption of the repetition n times of a number of independent events, of which the chance of occurrence is identical and equal to q. The population is n and the chance of occurrence q in the case of each individual. The mean frequency of occurrence is nq. But if q be very small we have seen that the series is

$$e^{-m}\left(1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots\right),$$

TABLE XIV.

Marriages of Uncle and Niece (1900—1909).

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16 & over
Marche	O.	7	3	—	—													
	M.	7·41	2·22	·33	·04													
	B.	7	3	—	—													
Umbria	O.	6	3	1	—	—												
	M	6·06	3·03	·76	·13	·02												
	B.	5·90	3·28	·73	·08	—												
Basilicata	O.	6	3	—	1	—												
	M	5·49	3·29	·99	·20	·03												
	B.	6·04	2·59	·92	·31	·10	·03	·01										
Sardegna	O.	2	5	3	—	—	—	—										
	M.	3·33	3·66	2·01	·74	·20	·05	·01										
	B.	2·01	4·96	3·03	—	—	—	—										
Emilia	O.	1	3	2	2	1	1	—										
	M.	1·11	2·44	2·68	1·97	1·08	·48	·18	·05	·01								
	B.	1·09	2·43	2·70	1·98	1·08	·47	·17	·05	·01								
Abruzzi	O.	—	3	1	3	2	—	1	—	—								
	M.	·61	1·70	2·38	2·23	1·56	·87	·41	·16	·06	·02							
	B.	·48	1·58	2·48	2·43	1·68	·87	·34	·11	·03	—							
Lazio	O.	1	1	2	3	—	2	—	1	—	—							
	M.	·45	1·40	2·17	2·24	1·73	1·07	·55	·25	·10	·03	·01						
	B.	·63	1·56	2·09	2·00	1·54	1·01	·59	·31	·14	·06	·03	·01	·01				
Puglie	O.	—	3	1	2	—	1	1	2	—	—	—	—	—				
	M.	·27	·98	1·77	2·13	1·91	1·38	·83	·42	·19	·08	·03	·01	—				
	B.	·55	1·30	1·77	1·80	1·53	1·14	·77	·49	·29	·16	·10	·05	·03	·01	·01		
Veneto	O.	1	—	1	1	3	1	—	1	2	—	—	—	—	—	—		
	M.	·11	·50	1·13	1·69	1·90	1·71	1·28	·82	·46	·23	·10	·04	·02	·01	—		
	B.	·21	·70	1·26	1·62	1·67	1·46	1·13	·79	·51	·30	·17	·09	·05	·02	·01	·01	
Toscana	O.	—	—	1	2	2	2	1	1	—	—	—	—	—	—	—	1	
	M.	·04	·24	·66	1·19	1·60	1·73	1·56	1·20	·81	·49	·26	·13	·06	·02	·01	—	
	B.	·31	·73	1·07	1·25	1·27	1·17	1·01	·83	·65	·50	·37	·27	·19	·13	·09	·06	·10
Calabria	O.	—	—	—	—	—	2	2	—	1	1	1	1	—	1	—	—	1
	M.	—	·01	·05	·16	·36	·64	·94	1·20	1·33	1·32	1·17	·95	·70	·48	·31	·18	·20
	B.	·00	·03	·11	·26	·48	·73	·96	1·16	1·21	1·17	1·04	·87	·69	·51	·37	·25	·16
Totals	O.	24	24	12	14	8	9	5	5	3	1	1	1	—	1	—	1	1
	M.	24·88	19·47	14·93	12·72	10·39	7·93	5·76	4·10	2·96	2·17	1·57	1·13	·78	·51	·32	·18	·20
	B.	24·22	22·16	16·46	11·73	9·35	6·88	4·98	3·74	2·84	2·19	1·71	1·29	·97	·67	·48	·32	·26

from which n has disappeared, and in this exponential we have seen that Bortkewitsch and Mortara suppose m small, i.e. 10 or under. We have seen that there is no reason why m should be absolutely small, and that the name given by Bortkewitsch to the Poisson-Exponential—i.e. the "Law of Small Numbers"—is misleading. But supposing the mean occurrence m to be small, it by no means follows that q need be small and n finite. For if $q = \cdot 2$ and $n = 4$, m would be "small"—and the sort of small number with which our authors deal, but the mere fact that the mean frequency of occurrence was 2 would not justify our using the Poisson-Exponential for

$$(\cdot 8 + \cdot 2)^4.$$

The fact is that when our authors speak of the deaths in a Prussian Army corps from the kick of a horse, or the suicides of schoolgirls, or the deaths from chronic alcoholism as being "small," they really mean small as compared with the number of persons exposed to risk. They had probably in mind all the men in the army corps, all school-girls or all individuals liable to death in the towns considered. But are all men in the army corps,—or only the cavalry, the artillery, etc.,—equally liable to death from the kick of a horse? Is every school-girl equally liable to commit suicide or only a very few morbid and unhealthy minded girls? Is every individual equally liable to die of chronic alcoholism, or only perhaps the 10 or 12 confirmed and aged drunkards in a town? The moment we realise these doubts, what is the population n to be considered? It is not m being small, but the smallness of m/n that leads us to believe that the binomial may have passed into an exponential. But if only six school-girls per year in a community are in the least likely to commit suicide, what is the justification for the "law of small numbers," if the average number of suicides be ·65? Further, if we pass to even a large community in which the tendency to commit suicide is graded—a very probable state of affairs—m might be small and n large, and yet since q is not constant, the binomial and its exponential limit would not be applicable; and this non-applicability would not depend on "lumpiness"—i.e. contagion or example in occurrence. Thus the probability might be:

$$(p_1 + q_1)(p_2 + q_2)(p_3 + q_3) \ldots (p_n + q_n)$$

with all the p's independent (as in spinning differently divided teetotums) and not correlated (as they would be in drawing successive non-returned cards from a pack). It would seem therefore that *a priori* we should not expect the conditions for the exponential to be fulfilled in most of the cases selected by Bortkewitsch and Mortara, although with perfect mixing we might expect it in the cases cited by "Student."

(14) In order to test this point on adequate numbers, the ages at death of all persons dying over 70 years of age were extracted for a period of three complete years from the notices of death in the *Times* newspaper for the years 1910—1912: see Table XV. These announcements of death are those of individuals in a fairly limited class, which may be considered stable in numbers for these three years.

TABLE XV.

Deaths per day of the Aged from the Times *newspaper.*

	Number of Deaths per diem	*70* Years and Over			*80* Years and Over			*85* Years and Over			*90* Years and Over		
		Observed	Binomial	Exponential	Observed	Binomial	Exponential	Observed	Binomial	Exponential	Observed	Binomial	Exponential
Men.	*0*	33	32·44	25·94	222	218·27	198·26	484	484·57	480·83	831	829·01	828·25
	1	110	107·50	97·11	339	332·78	338·99	391	391·78	396·16	225	230·70	232·00
	2	170	184·30	181·78	262	271·70	289·81	164	162·04	163·20	38	32·85	32·49
	3	246	217·68	226·84	151	157·71	165·18	45	45·69	44·82	2	3·44	3·26
	4	187	199·09	212·30	79	72·93	70·61	11	9·87	9·23			
	5	142	150·23	158 95	32	28·56	24·15	1	2·05	1·76			
	6	84	97·34	99·18	6	9·84	6·88						
	7	69	55·66	53·04	4	3·06	1·68						
	8	31	28·65	24·82	1	1·15	·44						
	9	19	13·47	10·36									
	10	4	5·86	3·88									
	11	1	3·78	1·80									
Women.	*0*	46	52·92	32·41	162	152·81	126·78	364	363·61	336·25	633	632·31	635·69
	1	140	139·03	114·11	267	274·37	273·47	376	375·20	397·30	350	350·88	346·27
	2	207	200·29	200·89	271	269·90	294·92	218	217·35	234·72	94	94·04	94·31
	3	221	209·30	235·78	185	192·49	212·04	89	93·12	92·45	17	16·20	15·31
	4	169	177·31	207·54	111	111·23	114·34	33	32·87	27·31	2	2·57	4·42
	5	119	129·17	146·15	61	55·25	49·33	13	10·11	6·45			
	6	87	83·89	85·77	27	24·45	17·73	2	2·81	1·27			
	7	44	49·73	43·14	8	9·88	5·46	1	·93	·25			
	8	35	27·38	18·99	3	3·70	1·47						
	9	18	14·17	7·43	1	1·92	·46						
	10	4	6·96	2·62									
	11	4	3·27	·84									
	12	1	1·48	·25									
	13	1	1·10	·08									

Table XVI shows that the announcements of deaths over 70 years of age only amount to 3·74 per day for males and 3·52 for females. These are certainly "small numbers," but "small" with regard to what? Are we to consider n as the number of the population which embraces, (i) all the individuals of the limited classes of the same range of ages as the defunct, (ii) all the individuals announced as dead on the same day, (iii) all the individuals of whatever ages of the class which announces deaths in the *Times*? Or, should we refer to all the individuals in the community of that range of ages, or the whole community at large, i.e. the chance that in a population of so many millions an individual over 70 or 80 as the case may be will die and have their death announced in the *Times* newspaper? Well, it really does not matter, because if for any one or all of these populations the binomial $(p + q)^n$ applied, we should get if q were small and n large, the Poisson series

$$e^{-m}\left(1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots\right),$$

and this quite regardless of the size of n. If therefore we did find a series in which q was very small and n large, we might not be able to say to which, if any of the above populations n applied. On the other hand the mere fact that m is small is no justification for the use of the "law of small numbers" as is sometimes implied. If it be argued that the small number of people who die over 80 and have their names recorded in the *Times* are drawn from a *small* population, we reply so it may be argued are the school children who commit suicide, the uncles who feel any inclination to marry their nieces, or the men liable to die of chronic alcoholism; and we can in the case of the announcement of deaths test the values of q and n on fairly adequate numbers. As a matter of fact we do not know, in attempting to apply the Poisson formula, what is the population from which we are drawing our individuals, and the justification of the Poisson formula lies only in showing that there actually does exist a binomial for which q is small and n large. We might imagine that as we got to the higher ages practically every person of that age would die, or that in our notation q would be 1 nearly and p be a very small quantity; thus an approach might be made to the Poisson-Exponential. But the approach to the Poisson-Exponential arises not through q approaching unity but from q becoming very small. Nor again in the lower age groups do we find ourselves left with a *positive* binomial.

In all cases except women over 90 years of age, we find that a negative binomial best fits the observations. Even in the case of the announcements of deaths of women over 90 years, we find that the approach of the binomial to the Poisson exponential depends on

$$\left(1 + \frac{1}{53{\cdot}3333}\right)^{53{\cdot}3333}$$

being measured with sufficient approximation by $e = 2{\cdot}71828$. But

$$(1{\cdot}01875)^{53{\cdot}333} = 2{\cdot}69323,$$

and is therefore not a very close approximation, a result shown when we use a binomial by the substantial improvement in the measure P of "goodness of fit." Even in this case we are not prepared to say what is the population for which the $q = \cdot01875$ in the case of these announcements of deaths of women over 90 years of age. It can scarcely be that there are only 29 women over 90 years

TABLE XVI.

Constants for Deaths of Aged.

Men.

Age over	p	q	Probable Error of q	n	Probable Error of n	m	Binomial P	Exponential P
70 years ...	1·12965	− ·12965	± ·03314	− 28·8747	± 7·3734	3·7436	·1355	·0045
80 years ...	1·12152	− ·12152	± ·03349	− 14·0703	± 3·8704	1·7099	·9358	·1129
85 years ...	1·01903	− ·01903	± ·02902	− 43·2996	± 67·5797	·8239	·9737	·9715
90 years ...	1·00654	− ·00654	± ·02934	− 42·8498	±192·3069	·2801	·6741	·6672

Women.

Age over	p	q	Probable Error of q	n	Probable Error of n	m	Binomial P	Exponential P
70 years ...	1·34012	− ·34012	± ·04161	− 10·3522	± 1·2307	3·5210	·8084	·0000
80 years ...	1·20770	− ·20770	± ·03294	− 10·4400	± 1·8309	2·1569	·9686	·0018
85 years ...	1·14507	− ·14507	± ·03077	− 8·1447	± 1·9627	1·1816	·9860	·1062
90 years ...	·98125	+ ·01875	± ·02779	+ 29·0573	±43·0634	·5447	·9848	·8116

of age living in the country, whose deaths are likely to be announced in the *Times* when they occur. Further the probable error of q is such that actually this case might equally well be a random sample from material following a negative binomial. Analysing our material we see that our first two cases of males and the first three of females are such that they could not possibly be random samples from positive binomials, the probable errors of q are too small. Next, seven cases out of the eight do give actually negative binomials and the eighth might, having regard to its probable errors, well be a negative binomial. Thus although our daily occurrences are certainly in Bortkewitsch and Mortara's sense "small numbers," they give no support to the use of a Poisson-Exponential.

If it be said that these "small numbers" differ in character from those used by our authors, the reply must be: we know in none of these cases the real population from which deaths are to be considered as drawn. The chances of death are certainly graduated with age, but the chances of suicide are graduated with temperament, and the same is true of alcoholism, or again the chance of

death by accident is graduated with occupation. At any rate until those who support the use of the "law of small numbers" demonstrate its application on material, where the probable errors are sufficiently small for us to measure the true value of q and n, no advance can be made. Nor until we have clear ideas of the population n in which the chance is q, is it possible to assert that it may be used for the suicides of school children, and the marriage of uncle and niece, and must not be used for the deaths of aged people, which certainly occur in "smaller" numbers.

In the illustrations of deaths we have taken, certainly the Poisson-Exponential is not the rule, although the distributions appear to approach it, as towards a limit, when the number of deaths approach zero. But our data which show the rule of the negative binomial appear to show it in no more marked manner than much of the data selected by Mortara himself indicate the negative binomial, although owing to the sparsity of his material his results are far more erratic and unreliable. Nor is Bortkewitsch much behind Mortara in the evidence he produces for a negative binomial being as reasonable a description—possibly owing to inherent lumpiness—as a positive binomial of these "small number" frequencies.

(15) *Conclusions.*

(*a*) The Poisson-Exponential gives a fairly reasonable method of dealing with the probable deviations of small sub-frequencies in the case of random sampling. When the average value of a sub-frequency is not more than 3 $^{\circ}/_{\circ}$ of a population, then Poisson's formula suffices in most practical cases to determine the range of error likely to be made. Tables are given to assist its use.

(*b*) The application of the Poisson-Exponential to various data by Bortkewitsch and Mortara has hardly been justified by those writers, for they have not tested whether the probability q is small and positive and the power n large and positive in the cases considered by them. When this is actually done, it is found that their hypotheses, having regard to the probable errors of q and n, are largely unjustified in the case of their illustrations. Even in such cases where it is justified, a binomial gives a better result as measured by the test for goodness of fit.

(*c*) Negative binomials repeatedly occur and give just as good fits, where they occur, as positive binomials. In the illustrations taken by Mortara, the frequency 10 used is so small that it is not possible to assert that either positive or negative binomials are demanded by the data. Still the average p of his results is very significantly in excess of unity.

(*d*) Mortara like Bortkewitsch cuts out of his data straight off all districts with, on the average, more than 10 cases in the year. But the q obtained from 20, 40, or even 100 cases in a population of 100,000 is a small q in the sense that the resulting binomial is adequately expressed by a Poisson-Exponential. There

appears to be no valid reason for such a procedure, except the experience that many such cases actually give negative binomials*. It seems to us theoretically unjustifiable to apply the exponential to 8 cases say in a district of 100,000, and not apply it to 12 cases in a district of 200,000. Actually p may be 1·4 in the first case and only 0·9 in the second.

(*e*) We consider that the reasonable method in every case is not to start with the Poisson-Exponential, which screens the truth or falsity of the *a priori* hypotheses, but to fit a binomial regardless of the magnitude of p. The fact that quite as good fits are obtained with negative as with positive binomials suggests that a new interpretation of these cases of "negative probability" is requisite. Several cases of the interrelation of "contributory cause groups" which provide a series represented by a negative binomial $(p-q)^{-n}$ have been recognised†. A general interpretation based on a very simple conception seems needed for these demographic cases in which the law of small numbers appears far more often to correspond to a negative than to a positive binomial.

This paper was worked out in the Biometric Laboratory, and I have to thank Professor Karl Pearson for his aid at various stages.

* Can we cite in addition perhaps, the fact that existing tables of $m^x e^{-m}/x!$ do not extend beyond $m=10$?

† Pearson, *Biometrika*, Vol. IV. p. 208.