Applications of Poisson's Probability Summation

By FRANCES THORNDIKE

SYNOPSIS: The applicability of Poisson's exponential summation to a variety of actual data is illustrated by thirty-two examples of actual frequency-distributions to which the Poisson distribution is a fairly good approximation. The comparison of actual and theoretical distributions is made graphically, using as a background new probability curves showing Poisson's exponential summation with a logarithmic scale for the average. To suggest possible explanations of the observed deviations from the theoretical Poisson distribution consideration is given to the effect on the theoretical distribution of certain modifications in the underlying assumptions, corresponding to conditions under which much actual data must be obtained.

IN an earlier number of THE BELL SYSTEM TECHNICAL JOURNAL there were published two sets of curves showing Poisson's exponential summation.[1] These charts, which are shown on a reduced scale in Figs. 1 and 2, give the relation between a, the average number of occurrences of an event in a large group of trials, the number of trials being very great compared with the average a, and the probability P that the actual number of occurrences in any such group of trials will equal or exceed any given number c. The purpose of this paper is to facilitate the use of these curves by making clear the characteristics of the Poisson summation, especially the assumptions on which it is based, and the precautions which must be observed in applying it, these points being illustrated by a number of actual frequency-distributions for which the Poisson distribution furnishes a fairly good working approximation.

POISSON'S EXPONENTIAL SUMMATION

Three assumptions underlie the mathematical treatment of Poisson's exponential summation

$$P=1-\left[1+\frac{a}{1!}+\frac{a^2}{2!}+\frac{a^3}{3!}+\ .\ .\ .\ +\frac{a^{c-1}}{(c-1)!}\right]e^{-a}$$

and its application to practical problems. The first is that the quantity measured is the number of occurrences of a particular event which always definitely happens or fails to happen, so that the actual number of occurrences c is either zero or a positive integer. The second assumption is that we may imagine the group of trials con-

[1] Figs. 1 and 2 of "Probability Curves Showing Poisson's Exponential Summation," by G. A. Campbell, *Bell System Technical Journal*, Vol. 2, No. 1, pp. 95–113, January, 1923.

stituting the sample in question to be repeated an infinite number of times, independently and uniformly, with an average number of occurrences per sample equal to a, so that we may speak of a as the average number of occurrences for the sample in question. The

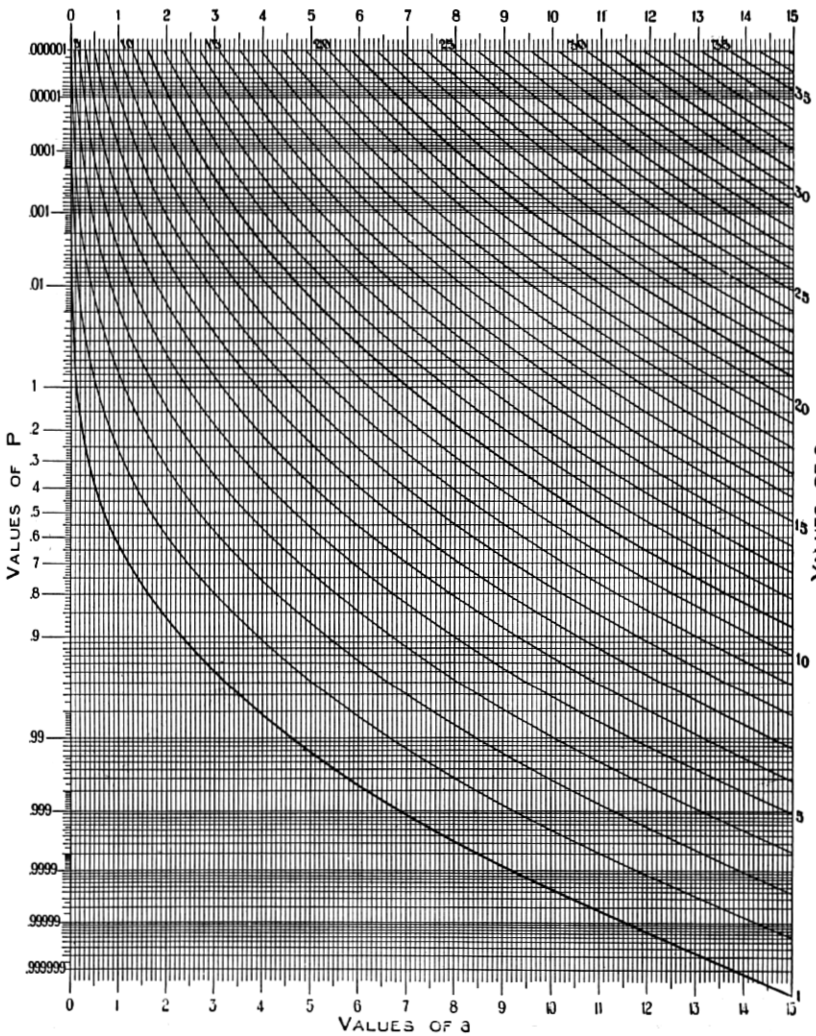


Fig. 1—Probability curves showing Poisson's exponential summation

$$P = 1 - \left[1 + \frac{a}{1!} + \frac{a^2}{2!} + \ldots + \frac{a^{c-1}}{(c-1)!}\right] e^{-a}$$

for the probability P that an event occur at least c times in a large group of trials for which the average number of occurrences is a. A scale proportional to the normal probability integral is used for P, a linear scale for a

third assumption is that, while the sample has a finite average number of occurrences, it consists of an infinite number of independent, uniform trials, so that the possible number of occurrences in a sample is infinite, and the probability that the event occur in a single trial

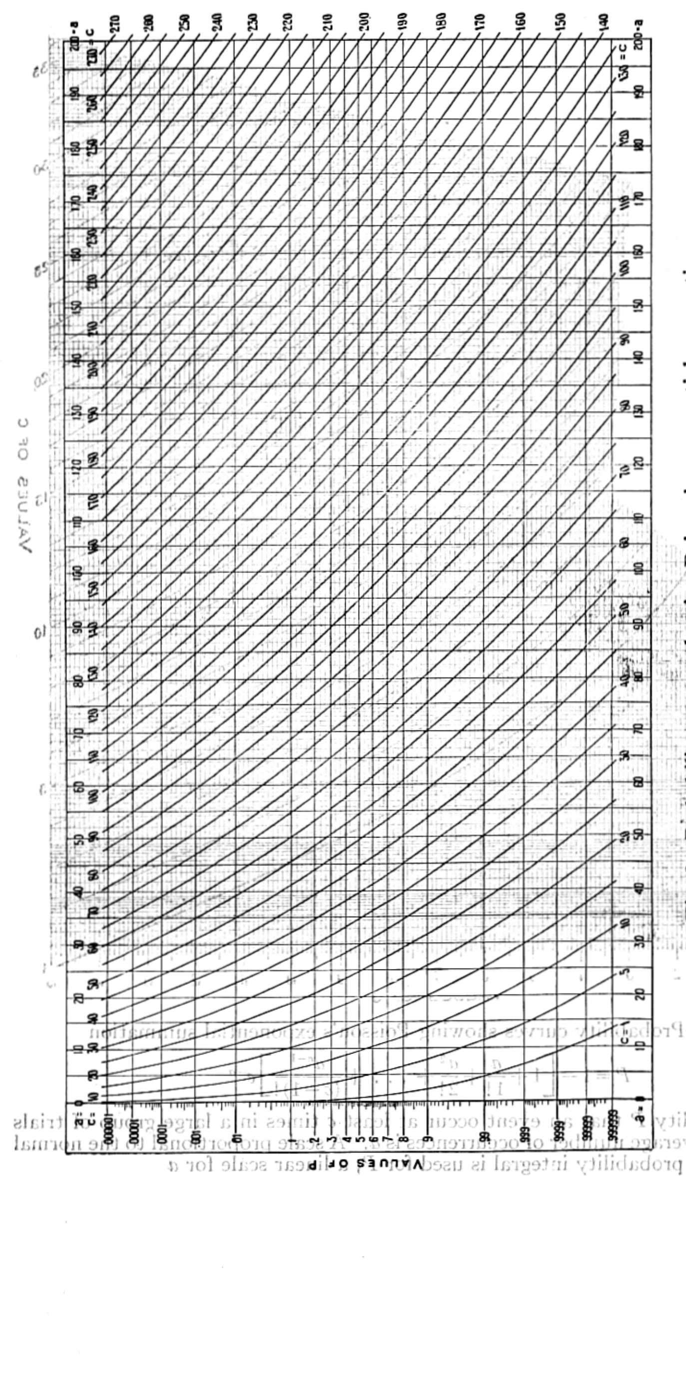


Fig. 2—Probability curves showing Poisson's exponential summation

$$P = 1 - \left[1 + \frac{a}{1!} + \frac{a^2}{2!} + \dots + \frac{a^{c-1}}{(c-1)!}\right] e^{-a}$$

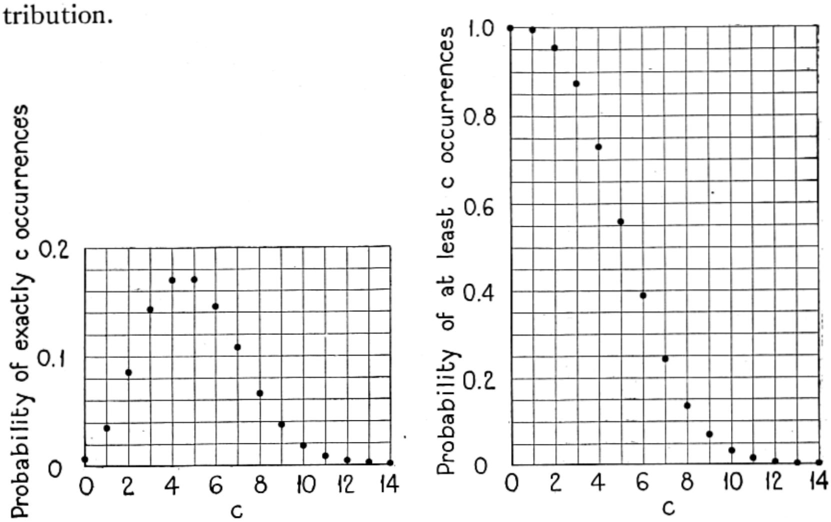
for the probability P that an event occur at least c times in a large group of trials for which the average number of occurrences is a. A scale proportional to the normal probability integral is used for P, a linear scale for a

is infinitely small. The term "uniform" applies, of course, not to the results of the trials (or samples) but to the essential conditions under which they are obtained, and "independent" is used with the meaning that the result of one trial (or sample) does not affect the occurrence of the event in any other trial (or sample). The first and third assumptions, translated into exact mathematical language, define a particular kind of probability function, which can be derived by taking the limit, as n becomes infinite and pn remains finite, of the point binomial $(p+q)^n$ for the probability of any number of occurrences of a given event in a group of n independent, uniform trials, when the probability that the event occur in a single trial is p. The second assumption is required in order that we may pass from the abstract idea of a probability function to the concrete idea of a frequency-distribution.

Throughout this discussion the summation form of the frequency-distribution, giving the probability of at least c occurrences, is used rather than the individual term form, giving the probability of exactly c occurrences. One reason for the use of the summation form is its more direct applicability to many practical problems in which the chance of exceeding a certain limit, rather than the chance of obtaining any one particular value, is of practical importance. Secondly, as Fig. 3a shows, the individual term form gives in general two possible values of c for any pair of values of a and P, whereas the summation form is single-valued and introduces no such ambiguity.

Fig. 3 also calls attention to some of the outstanding characteristics of the Poisson distribution, its discontinuity and skewness, in particular. That the Poisson distribution must be a series of discrete points and not a continuous curve is a direct result of the assumption that c represents a number of occurrences. That the distribution is skew follows from the fact that the possible number of occurrences is much larger, in fact infinitely larger, than the average number of occurrences. This skewness is quite marked even in the Poisson distribution with $a=5$, which is shown in Fig. 3, and it becomes more pronounced as a is decreased toward zero. If, for example, the average number of occurrences in a million trials is one, in any particular group of a million trials it is equally likely that there will be no occurrence of the event or one occurrence, and it is almost 1.4 times as likely that there will be no occurrence as that there will be two or more occurrences, though zero and two are equally removed from the average. A third important characteristic of the Poisson exponential, which is not brought out by this figure, is its extreme simplicity. The distribution is entirely determined by the value given to a single parameter, the average a; its standard deviation is $\sqrt{a}$, its skewness is $1/\sqrt{a}$, and its kurtosis is $3+1/a$.[2]

One consequence of this simplicity is that there is no difficulty in deciding on a definition of the *corresponding Poisson distribution* with which any other distribution should be compared. It is naturally the Poisson distribution having the same average as the given distribution.



(a) (b)

Fig. 3—Poisson distribution with the average $a=5$ shown (a) in the individual term form and (b) in the summation form

Poisson Probability Curves

Another advantage is that it is possible to represent the whole family of Poisson distributions graphically by a chart such as Fig. 1 or Fig. 2, in which the value of the average a is read on the horizontal scale, the value of the probability P on the vertical scale, and the number of occurrences c on the individual curves of the set. Any two of these three variables may then be taken as the independent variables and the values assigned to them will determine the value of the third variable, which can be read off at once. The only ambiguity occurs

[2] The standard deviation (σ), skewness (k), and kurtosis (β_2) of any distribution are defined as

$$\sqrt{\frac{\Sigma(x_i-a)^2}{N}},\quad \frac{\Sigma(x_i-a)^3}{\sigma^3},\ \text{and}\ \frac{\Sigma(x_i-a)^4}{\sigma^4},$$

respectively, N being the number of samples in the series, and x_i the actual number of occurrences in the ith sample. For any point binomial

$$\sigma=\sqrt{npq},\quad k=\frac{q-p}{\sqrt{npq}},\quad \beta_2=3+\frac{1-6pq}{npq}.$$

when a and P are the independent variables. The point determined by their values will, in general, fall between two of the c curves and the interpretation of P must be known to determine which of the two values of c should be taken. The desired value of c is read from the lower curve if P means *a probability of P or more*, from the upper curve if P means *a probability of not more than P*.

These charts may then be used conveniently in place of unwieldy double-entry tables to obtain theoretical values needed either for comparison with experimental data or to take the place of experimental data. Examples of such uses of the Poisson exponential are discussed in detail by Karl Pearson,[3] W. A. Shewhart,[4] and E. C. Molina.[5] The use of these curves in the study of telephone trunking, letting a represent the average number of simultaneous calls from a large group of subscribers, $c-1$ the number of trunks provided for them, and P the probability that all the trunks will be in use when a subscriber attempts to make a call, is suggested by Mr. Molina's paper. Other possible applications might be found in connection with the control of errors in service, defects in a manufactured article, the stock on hand of staple articles such as ink, shoe-polish, or spark plugs, or the number of copies of reference books in a library serving a large number of people. Still others may be suggested by Table I, which is a summary of the actual data now brought together for the first time for comparison with the theory.

The comparison of any actual distribution with the corresponding Poisson distribution may easily be made graphically, using these curves as a background. In fact the charts will often be found useful as coordinate paper on which to plot any frequency-distribution, theoretical or observed, provided the values of the variate are inherently limited to the positive integers and zero.

When the curves are used in this way the corresponding Poisson distribution is represented by the points in which the vertical line for the observed value of a cuts the c curves, or for convenience simply by the vertical line itself. The other distribution may then be plotted with c and P as the independent variables, and the horizontal deviations of these points from the vertical line serve as a measure of the discrepancy between the two distributions.[6] If the comparison is to be made with an observed frequency-distribution the values used

[3] Introduction to "Tables of the Incomplete Gamma Function," London, 1922.

[4] "Some Applications of Statistical Methods to the Analysis of Physical and Engineering Data," *Bell System Technical Journal*, Vol. 3, No. 1, pp. 43–87, January, 1924.

[5] "The Theory of Probabilities Applied to Telephone Trunking Problems," *Bell System Technical Journal*, Vol. 1, No. 2, pp. 69–81, November, 1922.

[6] The distributions might be plotted in other ways, e.g., letting P or c be the dependent variable, but the method used here is the simplest.

TABLE I

N = number of samples
aN = total number of occurrences
a = average number of occurrences per sample

	Series	N	aN	a
a 1	Alpha particles	2608	10097	3.87
a 2	Alpha particles	1304	10094	7.74
a 3	Deaths of aged	1096	903	0.82
a 4	Deaths of aged	1096	2364	2.16
a 5	Telephone lines in use	>1000	>4315	4.32
a 6	Bacilli	1000	1927	1.93
b 1	Yeast cells	400	720	1.80
b 2	Yeast cells	400	1872	4.68
b 3	Lost articles	423	439	1.04
b 4	Number *12*	500	421	0.84
b 5	Fires	364	9487	26.1
b 6	Incorrect reports	506	138	0.27
b 7	Cutoffs	506	1057	2.09
b 8	Double connections	506	1760	3.48
b 9	Calls for wrong number	506	2520	4.98
c 1	Deaths from kick of horse	200	122	0.61
c 2	Number *12*	250	251	1.00
c 3	Calls from group of two coin-box telephones	145	172	1.19
c 4	Calls from group of four coin-box telephones	140	384	2.74
c 5	Calls from group of two coin-box telephones	141	212	1.5[illegible]
c 6	Calls from group of six coin-box telephones	138	468	3.39
c 7	Cutoffs	267	557	2.09
c 8	Double connections	267	906	3.39
c 9	Calls for wrong number	267	1351	5.06
c10	Connections to wrong number	267	2334	8.74
c11	Party lines	300	1981	6.60
c12	"Lost and found" advertisements	209	7051	33.7
d 1	Number *12*	100	421	4.21
d 2	Number *12*	50	421	8.42
d 3	Comets	100	258	2.58
d 4	Particles in emulsion	50	46	0.92
d 5	Particles in emulsion	50	106	2.12

for the probability P are the values of the observed relative frequency F, which are calculated as indicated in Table II, and the observed distribution is represented by an irregular series of dots, as in Fig. 4.

A third set of curves, Fig. 5, supplementary to Figs. 1 and 2, has now been drawn using a logarithmic scale for a. This chart shows the individual c curves up as far as $a = 30$ and it shows more clearly than does Fig. 1 the range $0.1 \leqq a \leqq 2$. It may also be used as a background in the same way as Figs. 1 and 2, with the additional advantage of making the distances of the plotted points from the vertical line proportional to the percentage deviations rather than proportional to the absolute values of the deviations, so that the fit of a distribution having a small average can be compared directly by eye with that of a distribution having a large average, since it is more often the relative than the absolute value of the deviation which is significant.

Practical Applications

In applying the Poisson summation to any concrete problem, or in comparing any observed distribution with the corresponding Poisson distribution, it is necessary to bear in mind several practical conditions which must work against any perfect agreement between the

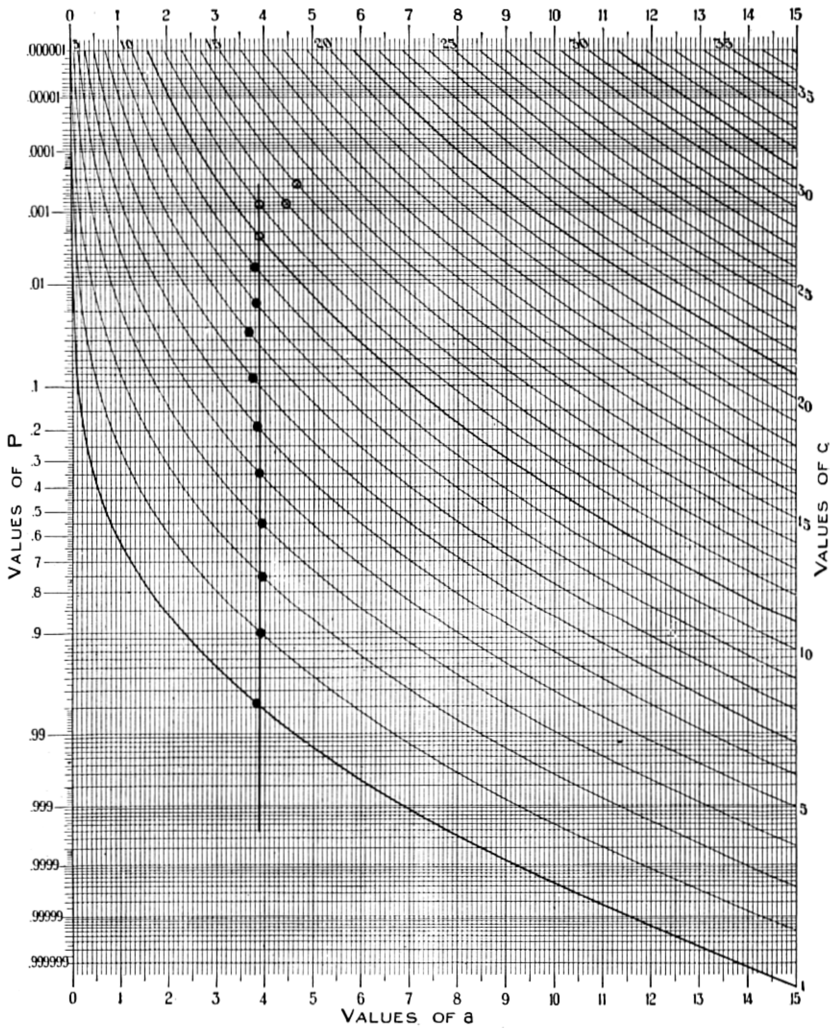


Fig. 4—Comparison of an observed distribution of the number of α particles emitted with the corresponding Poisson distribution, showing the method of using Fig. 1 or Fig. 2 as a background for plotting actual distributions. The Poisson distribution is shown by a vertical line, the observed distribution by dots

observed distribution and the corresponding Poisson distribution. In the first place, the sample considered will necessarily consist of a finite number of trials instead of an infinite number as assumed in the mathematical theory, and the trials may not be completely independent or entirely uniform. Secondly, even if the individual sample possessed the ideal characteristics assumed in the mathematical formulation, the actual series of samples must be finite and the samples may be interdependent and far from uniform. The size of the samples relating to the economic, geographic, and time divisions ordinarily used in statistical work generally varies considerably. The effect of modifying the original mathematical assumptions to correspond with some of these actual conditions is illustrated by Figs. 6–8, which show various theoretical frequency-distributions plotted on Fig. 1 or Fig. 2 for comparison with the corresponding Poisson distributions.

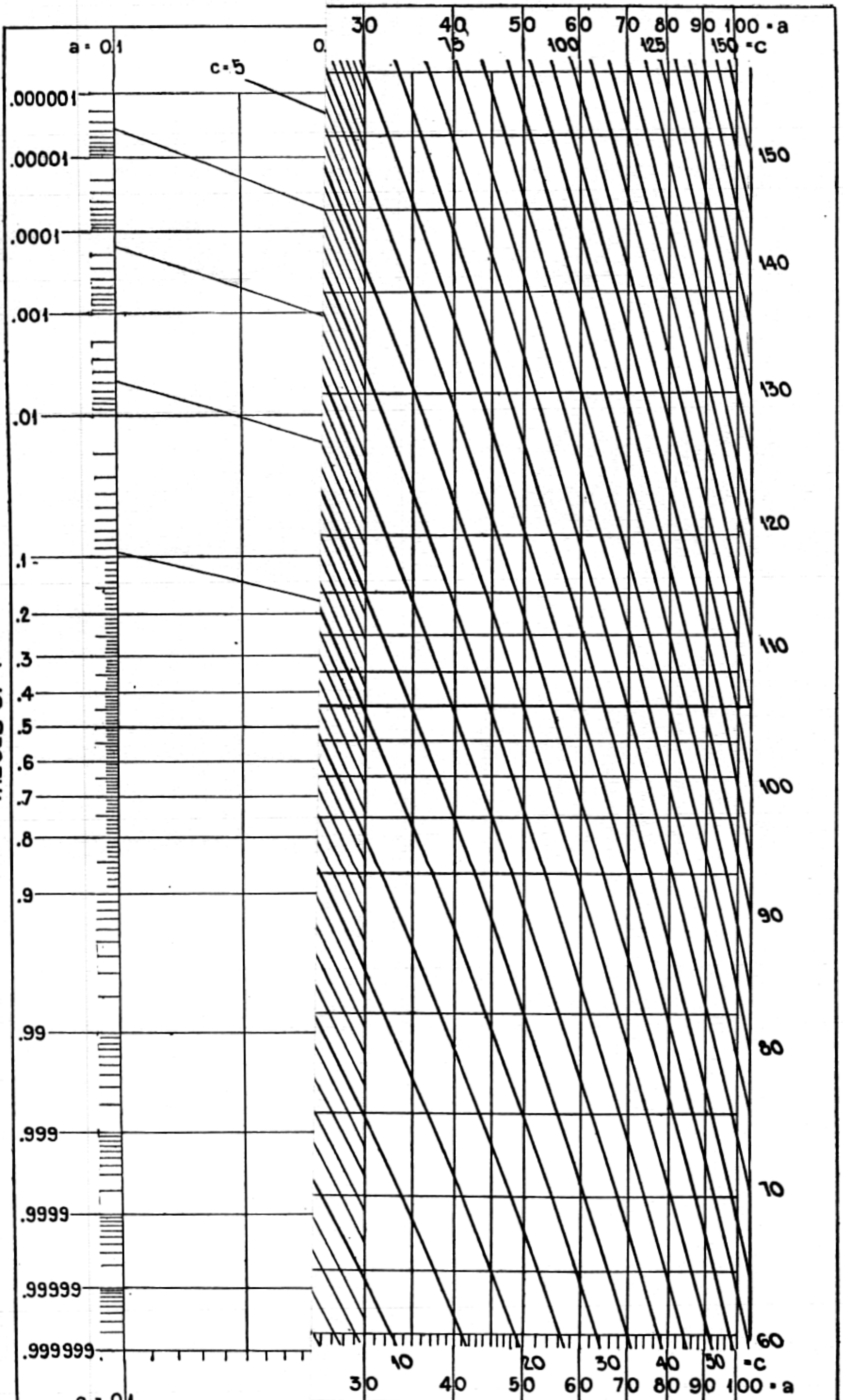
The finiteness of the number of trials n not only makes impossible the occurrence of values of c greater than the value of n, but also tends to produce a general trend away from the Poisson distribution. This is illustrated by the four typical finite binomial distributions shown in Fig. 6, which have a definite curve and slope toward the left which becomes more pronounced as n is decreased.[7] Interdependence of the trials constituting a sample will also tend to give the resulting distribution a slant, to the right if the correlation is positive, to the left if the correlation is negative.[8] Thirdly, even though the trials are independent, if they are not uniform, there will be a tendency for the distribution to slant to the left.

The requirement that N, the number of samples in the actual series, be finite introduces a somewhat different kind of deviation from the theoretical Poisson distribution. The observed relative frequency F, which is compared with the theoretical probability P, is an integral multiple of $1/N$, so that, since N is finite, the points representing the observed distribution (except those at $P=0$ and $P=1$, for which the ordinates are plus and minus infinity, and which, therefore, never appear on the graph) are all in the finite range between the two horizontal lines $P=1/N$ and $P=1-1/N$. Not only is the occurrence of points outside this range impossible, but the points near its extremes, being determined by a comparatively small number of samples, are of less significance than those near the center.

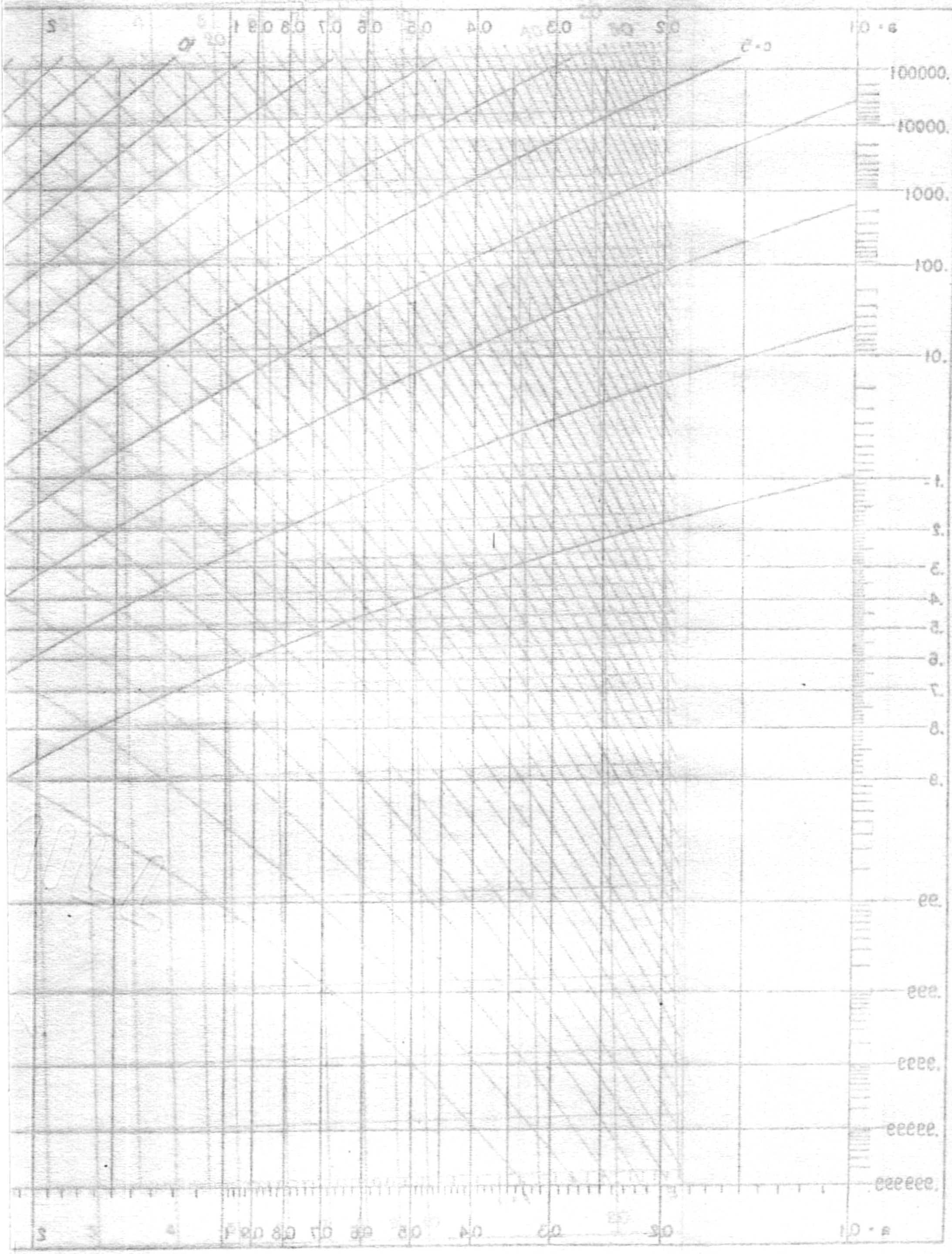
To call attention to these facts all observed distributions shown here have been represented, as in Fig. 4, with the vertical line rep-

[7] A more detailed discussion of the effect of finite sampling will be found in the paper by G. A. Campbell previously referred to.

[8] See "Explanation of Deviations from Poisson's Law in Practice," by "Student," *Biometrika*, Vol. 12, pp. 211–215, 1919.



tional to the



resenting the corresponding Poisson distribution terminated at $P=1/N$ and $P=1-1/N$, and with the observed points in the range $P=10/N$ to $P=1-10/N$ shown as solid black dots and the points outside this range shown as circles with white centers. This secondary division is quite arbitrary, for the increase in reliability of the points as the center of the range is approached is gradual. There will, of course, be irregularities due to sampling even in the center as long as the number of samples is finite.

Non-uniformity of the samples of the series may introduce a definite trend away from the Poisson distribution, a slant to the right such as is shown in Figs. 7 and 8. Such trends result when the value of a varies from sample to sample of the series. Fig. 7 shows three theoretical distributions of this sort, each having the same average $a=75$. Series (a) is made up of two equal sub-series having $a=50$ and $a=100$, respectively, (b) of two unequal sub-series, in the ratio of 3:1, having $a=60$ and $a=120$, respectively, and (c) of three equal sub-series having $a=15$, $a=60$, and $a=150$, respectively.[9] Fig. 8 shows the effect on the distribution of letting a vary continuously and uniformly between the limits 5 and 15, the compound series (b) made up of two equal sub-series with averages 5 and 15 being also shown for comparison.[10] Since in practical time series a usually increases or decreases with the time, this kind of distribution may be expected to occur frequently. It should be noted that in all these cases it is immaterial whether a changes because of a change in the number of trials in the sample, or because of a change in the probability of the event's happening at a single trial, or because of both; if a is constant throughout the series a Poisson distribution will be obtained, and if a varies the tendency to slope to the right will be introduced. Various devices may be employed to keep the average constant in an actual series, some of which will be illustrated by the examples given below.

In selecting the following examples of the Poisson summation only two general rules were followed: that there must be some reason to

[9] In a compound distribution

$$P=\sum \frac{N_i}{N} P_i$$

where N_i is the number of samples with the average a_i, and $P_i=P(c, a_i)$.

[10] If a varies uniformly and continuously from a_1 to a_2

$$P=\int_{a_1}^{a_2} \frac{P(c, a)}{a_2-a_1}\, da$$

$$=1-\frac{1}{a_2-a_1} \sum_{i=0}^{c} [P(i, a_2)-P(i, a_1)].$$

suppose the possible number of occurrences n to be at least thirty times the average a and at least 25, and that N, the number of samples in the series, must be at least 50. This last requirement excludes from our list a number of series which have previously been presented

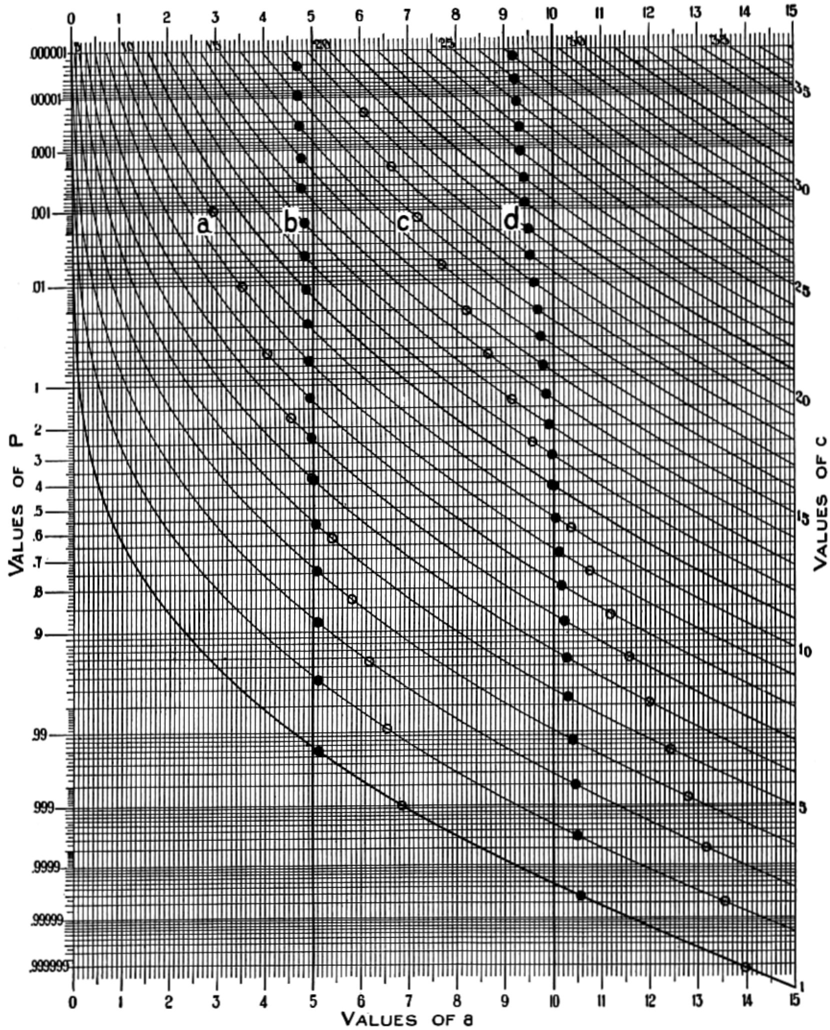


Fig. 6—Typical finite binomial distributions for the probability that an event occur at least c times in a group of n trials for which the average number of occurrences is $a = np$

(a) $a = 5$, $n = 10$
(b) $a = 5$, $n = 100$
(c) $a = 10$, $n = 20$
(d) $a = 10$, $n = 100$

as examples of the Poisson exponential, in particular those of Mortara [11] and all but one of those given by Bortkewitsch.[12]

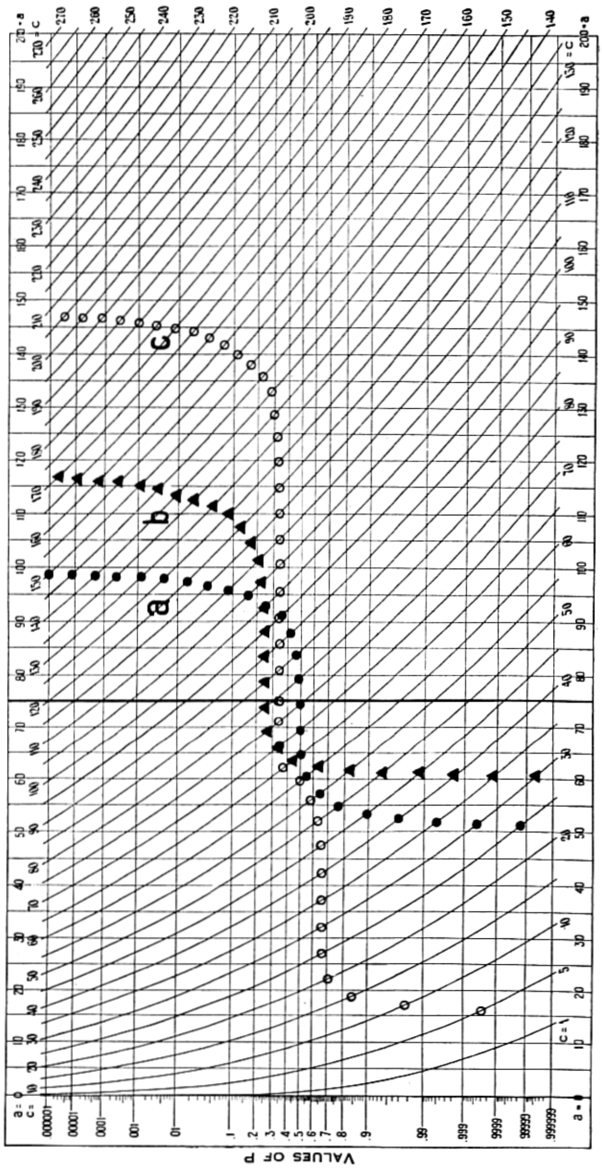


Fig. 7—Theoretical distributions for series in which

(a) $a=50$ for one half of the samples and $a=100$ for the other half

(b) $a=60$ for three quarters of the samples and $a=120$ for the other quarter

(c) $a=15$ for one third of the samples, $a=60$ for one third, and $a=150$ for the other third

[11] "Sulle Variazione di Frequenza di Alcuni Fenomeni Demografici Rari," by Giorgio Mortara, *Annali di Statistica*, Series V, Vol. 4, pp. 5–61, 1912.

[12] "Das Gesetz der kleinen Zahlen," by L. von Bortkewitsch, Leipzig, 1898.

Each of the thirty-two actual distributions shown in Fig. 9 has been plotted using Fig. 5 as the background, so that the percentage deviations in all distributions may be compared directly by inspection without regard to the magnitude of the average. The examples are divided into four groups according to the number of samples in the series, and are arranged in each group roughly in order of decreasing agreement of the observed with the theoretical distributions. A

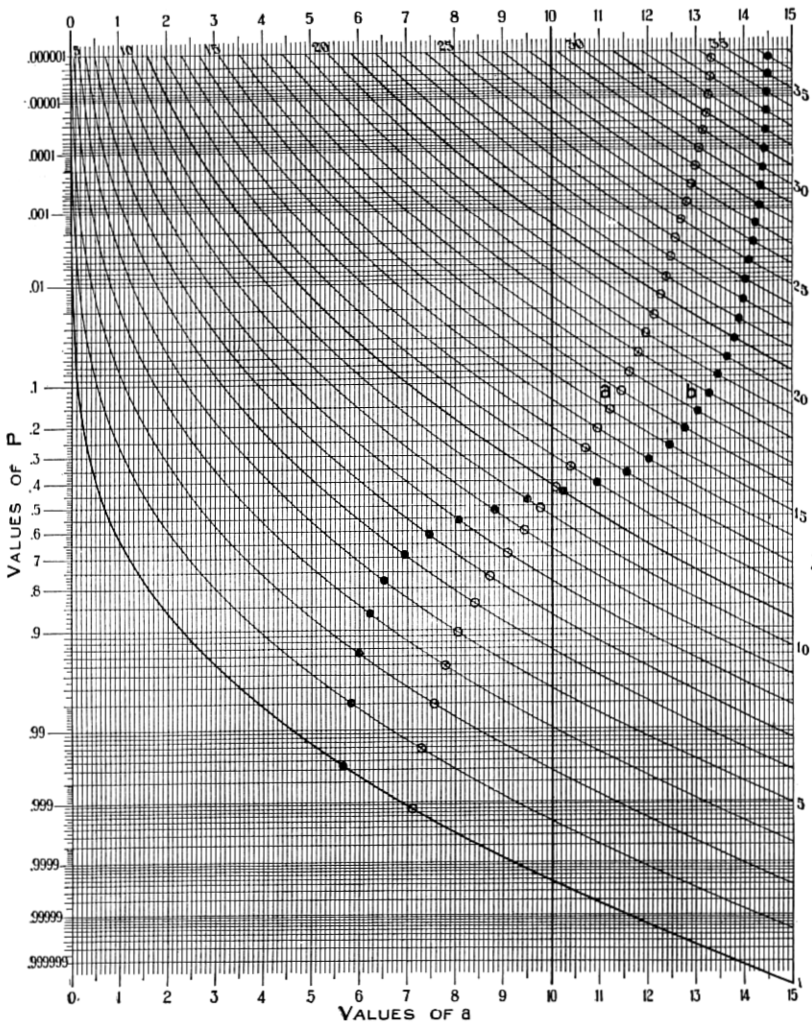


Fig. 8—Theoretical distribution for a series in which the average a varies continuously and uniformly from 5 to 15

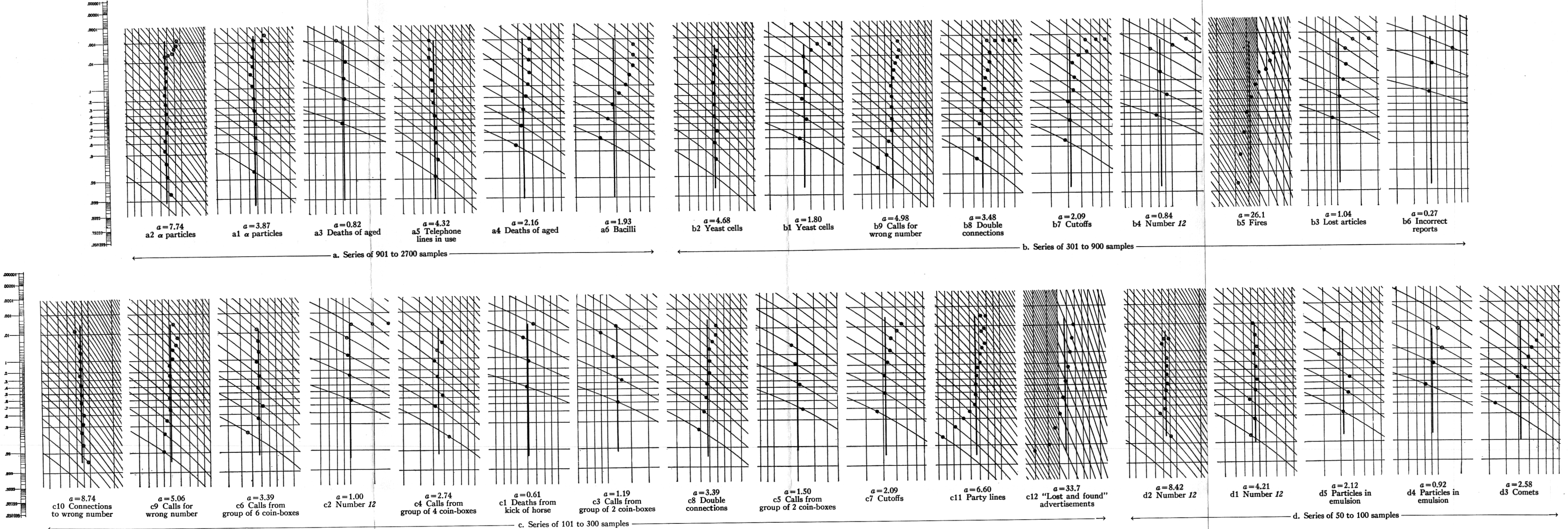


Fig. 9—Comparison of observed distributions with the corresponding Poisson distributions, using Fig. 5 as a background

summary of the data used is given in Table I and the observed distributions are given in full in Table II.

The distributions shown in the first group are taken from the work of Rutherford and Geiger, Whitaker, Holm, and Greenwood and White. Rutherford and Geiger observed the collision with a small screen of an α particle emitted from a small bar of polonium placed at a short distance from the screen. The number of such collisions in each of 2608 eighth-minute intervals was recorded, the distance between bar and screen being gradually decreased so as to compensate for the decay of the radioactive substance. From this record two frequency-distributions were calculated, that of the number of particles striking the screen in an eighth-minute interval, and in a quarter-minute interval.[13] These are distributions (a1) and (a2), respectively. Distributions (a3) and (a4) are based on a count of the number of death notices in the London *Times* on each day for three consecutive years.[14] The distribution of deaths of men over 85 years of age (a3) and that of deaths of women over 80 (a4) are shown here. The next (a5) is a frequency-distribution of the number of telephone lines simultaneously in use, from measurements on a group of 100 subscribers.[15] The last distribution of this group (a6) was obtained from a count of the number of bacilli in each of 1,000 phagocytes, or white blood cells, in the same solution and as far as possible under the same conditions, and is typical of a large number of distributions of the number of tubercle bacilli ingested per cell.[16]

The first two examples in the second group are due to "Student" and the remaining seven are new. Distributions (b1) and (b2) show the results obtained from two different solutions of yeast cells by counting the number of cells per square of a haemacytometer slide on which the solution had been spread as uniformly as possible after it had been thoroughly shaken to break up any clumps of cells.[17] The next example (b3) was obtained from the records of the "lost and found" office of the Telephone and Telegraph Building, 195 Broadway, New York City. The number of lost articles found in the building

[13] "The Probability Variations in the Distribution of α Particles," by Ernest Rutherford and Hans Geiger, *Phil. Mag.*, Vol. 20, pp. 698–707, October, 1910.

[14] "On the Poisson Law of Small Numbers," by Lucy Whitaker, *Biometrika*, Vol. 10, pp. 36–71, 1914. Six other similar distributions are given.

[15] "Calculation of Blocking Factors of Automatic Exchanges," by Ragnar Holm, *P. O. E. E. J.*, Vol. 15, pp. 22–38, April, 1922.

[16] "A Biometric Study of Phagocytosis with Special Reference to the 'Opsonic Index'," by M. Greenwood and J. D. C. White, *Biometrika*, Vol. 6, pp. 376–401, 1908–1909. Fourteen other distributions are given.

[17] "On the Error of Counting with a Haemacytometer," by "Student," *Biometrika*, Vol. 5, pp. 351–360, 1906–1907. Two other distributions are given.

and turned in to the office on each day except Sundays and holidays was recorded and tabulated for the period from November 1, 1923 to September 30, 1925, inclusive, excluding June, July, and August of each year, when there might be considerable variations in the population of the building. Distribution (b4) shows the result of a count of the number of times that the number *12* appeared as the last two digits of a ten-place logarithm in a sample consisting of a column of 100 logarithms in Duffield's table,[18] and (b5) shows the number of fires per day in New York City in 1924, as reported daily in *The New York Times*, the figures for July 4 and for Election Day being discarded for obvious reasons. The last four examples in this group were taken from telephone company records of local service observations. A sample consisted of the calls observed at one central office in one month, and the series of samples used was selected from a complete record for all the central offices in a large city by the requirement that the number of calls per sample be not less than 450 nor more than 550. Distribution (b6) was obtained for the number of incorrect reports, (b7) for the number of cutoffs, (b8) for the number of double connections, and (b9) for the number of calls for the wrong number.

Group three is headed by Bortkewitsch's classical example of the Poisson exponential.[19] He found from the records of the Prussian army the number of men killed by the kick of a horse in each of 14 corps in each of 20 successive years, and, after discarding the records for 4 corps which were considerably larger than the others, treated the rest as one series of samples. This is distribution (c1). Series (c2) is similar to (b4), except that the samples of 100 two-place numbers were obtained from several different sources, logarithmic tables, trigonometric tables, and numbers listed in a telephone directory. Examples (c3), (c4), (c5), and (c6) show the variation in the number of telephone messages recorded per five-minute interval for certain groups of coin-box telephones in a large transportation terminal. The number of calls registered for each of 23 such telephones in each of about 20 five-minute intervals between noon and 2 p.m. was recorded on each of seven days (no Saturdays or Sundays included) but as the telephones are arranged in groups the distribution of the number of calls per interval was calculated for each group rather than for the individual telephones. These shown here are for a group of two telephones (c3), a group of four (c4), another group of two (c5), and a group of six (c6). The next four examples are

[18] "Logarithms, Their Nature, Computation, and Uses," by W. W. Duffield, Washington, 1897.

[19] Bortkewitsch, *op. cit.*

similar to examples (b6)–(b9), except that the limits of the number of calls per sample were 515 ± 25. Distribution (c7) was obtained for the number of cutoffs, (c8) for the number of double connections, (c9) for the number of calls for the wrong number, and (c10) for the number of connections to the wrong number. The next distribution (c11) was obtained from a count of the number of party-line subscribers listed per page of a large telephone directory and the last distribution of the group (c12) from a count of the number of advertisements in the "lost and found" column of *The New York Times* on each of the week-days from January 1, 1924 to August 31, 1924.

The fourth group contains only five examples, three of which are new. The first two of these present the same material used for example (b4) differently arranged. The 50,000 logarithms used are divided into 100 groups of 500 logarithms each for example (d1), and into 50 groups of 1,000 logarithms each for example (d2). The third (d3) is the distribution of the number of comets observed per year for the years 1789 to 1888 inclusive.[20] The other two distributions have been given by Perrin as typical of the data obtained when, in order to determine the density of the particles of an emulsion at a given depth, he restricted his field of vision to a tiny part of that layer, small enough so that the average number of particles visible was only one or two, and then made a large number of observations of the number of particles in that space at regular intervals.[21]

As was to be expected, these observed distributions have not only irregularities due to finite sampling but also in some cases what appear to be definite trends away from the corresponding Poisson distributions. In some cases there is an explanation ready at hand. For example, in series (b3), which gives the number of articles lost in the Telephone and Telegraph Building, the average number of articles lost per day might be expected to increase as the population of the building increased in this period following the completion of an addition, and the observed slant to the right is what would be expected. Also in series (d3), which gives the number of comets observed per year, the average would naturally increase steadily as a result of the continual improvement of telescopes and other instruments from 1789 to 1888. The curve toward the left in examples (c3) and (c5) might also be predicted because of the fact that the number of calls which could possibly be made in five minutes from a group of two telephones is certainly finite and probably rather small, and in examples (d4) and (d5) because it is difficult to judge by eye the number

[20] "Handbook of Astronomy," by G. F. Chambers, 4th ed., Oxford, 1889.

[21] "Brownian Movement and Molecular Reality," by Jean Perrin, London, 1910.

of particles visible simultaneously if that number is more than three or four.

In several cases special measures have been taken to reduce the variation of a and the resulting trend away from the corresponding Poisson distribution. In general, a is made as nearly constant as possible by making n and p constant throughout. In examples (b6)–(b9) and (c7)–(c10), for instance, each sample consists of approximately the same number of calls, and in example (c1) four corps were rejected because they were considerably larger than the others. In these examples it is assumed that p is practically constant and that by making n constant a constant average will be obtained. A somewhat different adjustment to keep a constant is illustrated by examples (a1) and (a2), where, as the decay of the radioactive substance decreases the average number of α particles emitted in a given solid angle per unit of time, the screen on which the particles strike is moved so that it intercepts a greater angle. In some cases n may be controlled much more easily than p, or vice versa, and a may be kept constant by letting one factor vary and adjusting the other to compensate, rather than by keeping both constant.

Summary

These examples of distributions which can be described by the Poisson exponential are of a dozen quite different kinds. They include eleven distributions found in published work on biometrics or statistics and twenty-one which are new. The agreement between the observed and the theoretical distribution is, in general, fairly good, and the applicability of the Poisson summation to a great variety of data is clearly indicated. The practical importance of some of these cases has been discussed above.

The use of the probability curves showing Poisson's exponential summation in place of double-entry tables as a source of data is shown to be simple, and their convenience as a background for plotting and comparing frequency-distributions is illustrated by Figs. 4 and 6–9. The new chart with a logarithmic scale for a (Fig. 5) is convenient in comparing distributions of different averages. It also shows the complete set of curves up to $a=30$ instead of only to $a=15$, and it makes it possible to read with considerable accuracy values of the variables in the range $0.1 \leqq a \leqq 2$, which is not clearly shown in Fig. 1 or Fig. 2.

TABLE II

c = number of occurrences of the event per sample.
m = number of samples with exactly c occurrences.
f = number of samples with at least c occurrences.
F = relative frequency of at least c occurrences per sample.

a1 *Alpha particles*
total = 10097
average = 3.87

c	m	f	F
0	57	2608	1.000
1	203	2551	.978
2	383	2348	.900
3	525	1955	.753
4	532	1440	.552
5	408	908	.348
6	273	500	.192
7	139	227	087
8	45	88	.034
9	27	43	.0165
10	10	16	.0061
11	4	6	.0023
12	0	2	.00077
13	1	2	.00077
14	1	1	.00038

a2 *Alpha particles*
total = 10094
average = 7.74

c	m	f	F
0	0	1304	1.0000
1	3	1304	1.0000
2	17	1301	.9977
3	46	1284	.9847
4	99	1238	.949
5	126	1139	.873
6	151	1013	.777
7	187	862	.661
8	180	675	.518
9	173	495	.380
10	131	322	.247
11	75	191	.146
12	44	116	.089
13	35	72	.055
14	16	37	.028
15	14	21	.0161
16	1	7	.0054
17	1	6	.0046
18	2	5	.0038
19	1	3	.0023
20	1	2	.00153
21	1	1	.00077

a3 *Deaths of aged*
total = 903
average = 0.82

c	m	f	F
0	484	1096	1.000
1	391	612	.558
2	164	221	.202
3	45	57	.052
4	11	12	.0109
5	1	1	.00091

a4 *Deaths of aged*
total = 2364
average = 2.16

c	m	f	F
0	162	1096	1.000
1	267	934	.852
2	271	667	.609
3	185	396	.361
4	111	211	.193
5	61	100	.091
6	27	39	.036
7	8	12	.0109
8	3	4	.0036
9	1	1	.00091

a5 *Telephone lines in use**
total = ?
average = 4.32

c	M	F
0	.013	1.000
1	.045	.987
2	.125	.942
3	.185	.817
4	.187	.632
5	.186	.445
6	.126	.259
7	.071	.133
8	.036	.062
9	.018	.026
10	.005	.008
11	.002	.003
12	.001	.001

a6 *Bacilli*
total = 1927
average = 1.93

c	m	f	F
0	219	1000	1.000
1	267	781	.781
2	219	514	.514
3	129	295	.295
4	70	166	.166
5	50	96	.096
6	26	46	.046
7	13	20	.020
8	5	7	.007
9	2	2	.002

b1 *Yeast cells*
total = 720
average = 1.80

c	m	f	F
0	75	400	1.000
1	103	325	.813
2	121	222	.555
3	54	101	.253
4	30	47	.118
5	13	17	.043
6	2	4	.0100
7	1	2	.0050
8	0	1	.0025
9	1	1	.0025

b2 *Yeast cells*
total = 1872
average = 4.68

c	m	f	F
0	0	400	1.000
1	20	400	1.000
2	43	380	.950
3	53	337	.843
4	86	284	.710
5	70	198	.495
6	54	128	.320
7	37	74	.185
8	18	37	.093
9	10	19	.048
10	5	9	.023
11	2	4	.010
12	2	2	.005

b3 *Lost articles*
total = 439
average = 1.04

c	m	f	F
0	169	423	1.000
1	134	254	.600
2	74	120	.284
3	32	46	.109
4	11	14	.033
5	2	3	.0071
6	0	1	.0024
7	1	1	.0024

b4 *Number 12*
total = 421
average = 0.84

c	m	f	F
0	231	500	1.000
1	150	269	.538
2	92	119	.238
3	24	27	.054
4	1	3	.006
5	1	2	.004
6	1	1	.002

b5 *Fires***
total = 9487
average = 26.1

c	m	f	F
0		364	1.0000
5		364	1.0000
10		363	.9973
15		346	.951
20		286	.786
25		185	.508
30		103	.283
35		53	.146
40		22	.060
45		18	.049
50		8	.022
55		4	.0110

b6 *Incorrect reports*
total = 138
average = 0.27

c	m	f	F
0	388	506	1.000
1	102	118	.233
2	12	16	.032
3	4	4	.0079

b7 *Cutoffs*
total = 1057
average = 2.09

c	m	f	F
0	75	506	1.000
1	126	431	.852
2	141	305	.603
3	73	164	.324
4	50	91	.180
5	29	41	.081
6	6	12	.024
7	2	6	.0119
8	3	4	.0079
9	0	1	.0020
10	0	1	.0020
11	1	1	.0020

b8 *Double connections*
total = 1760
average = 3.48

c	m	f	F
0	21	506	1.000
1	63	485	.958
2	98	422	.834
3	97	324	.640
4	85	227	.449
5	61	142	.281
6	42	81	.160
7	18	39	.077
8	11	21	.042
9	6	10	.0198
10	3	4	.0079
11	0	1	.0020
12	0	1	.0020
13	0	1	.0020
14	0	1	.0020
15	1	1	.0020

b9 *Calls for wrong number*
total = 2520
average = 4.98

c	m	f	F
0	10	506	1.000
1	20	496	.980
2	45	476	.941
3	60	431	.852
4	85	371	.733
5	92	286	.565
6	73	194	.383
7	55	121	.239
8	28	66	.130
9	18	38	.075
10	9	20	.040
11	5	11	.022
12	3	6	.0119
13	2	3	.0059
14	1	1	.0020

c1 *Deaths from kick of horse*
total = 122
average = 0.61

c	m	f	F
0	109	200	1.000
1	65	91	.455
2	22	26	.130
3	3	4	.020
4	1	1	.005

c2 *Number 12*
total = 251
average = 1.00

c	m	f	F
0	90	250	1.000
1	95	160	.640
2	46	65	.260
3	15	19	.076
4	3	4	.016
5	0	1	.004
6	0	1	.004
7	1	1	.004

c3 *Calls from group of two coin-box telephones*
total = 172
average = 1.19

c	m	f	F
0	44	145	1.000
1	48	101	.697
2	38	53	.366
3	13	15	.103
4	1	2	.0138
5	1	1	.0069

c4 *Calls from group of four coin-box telephones*
total=384
average=2.74

c	*m*	*f*	*F*
0	5	140	1.000
1	33	135	.964
2	24	102	.729
3	38	78	.557
4	23	40	.286
5	9	17	.121
6	4	8	.057
7	4	4	.029

c5 *Calls from group of two coin-box telephones*
total=212
average=1.50

c	*m*	*f*	*F*
0	27	141	1.000
1	49	114	.809
2	39	65	.461
3	19	26	.184
4	7	7	.050

c6 *Calls from group of six coin-box telephones*
total=468
average=3.39

c	*m*	*f*	*F*
0	8	138	1.000
1	13	130	.942
2	20	117	.848
3	37	97	.703
4	24	60	.435
5	20	36	.261
6	8	16	.116
7	5	8	.058
8	2	3	.022
9	1	1	.0072

c7 *Cutoffs*
total=557
average=2.09

c	*m*	*f*	*F*
0	44	267	1.000
1	62	223	.835
2	71	161	.603
3	43	90	.337
4	25	47	.176
5	14	22	.082
6	4	8	.030
7	2	4	.0150
8	2	2	.0075

c8 *Double connections*
total=906
average=3.39

c	*m*	*f*	*F*
0	14	267	1.000
1	33	253	.948
2	48	220	.824
3	56	172	.644
4	43	116	.434
5	34	73	.273
6	22	39	.146
7	8	17	.064
8	4	9	.034
9	3	5	.0187
10	2	2	.0075

c9 *Calls for wrong number*
total=1351
average=5.06

c	*m*	*f*	*F*
0	3	267	1.000
1	12	264	.989
2	23	252	.944
3	31	229	.858
4	45	198	.742
5	50	153	.573
6	37	103	.386
7	29	66	.247
8	13	37	.139
9	12	24	.090
10	4	12	.045
11	4	8	.030
12	3	4	.0150
13	1	1	.0037

c10 *Connections to wrong number*
total=2334
average=8.74

c	*m*	*f*	*F*
2	1	267	1.0000
3	5	266	.9963
4	11	261	.978
5	14	250	.936
6	22	236	.884
7	43	214	.801
8	31	171	.640
9	40	140	.524
10	35	100	.375
11	20	65	.243
12	18	45	.169
13	12	27	.101
14	7	15	.056
15	6	8	.030
16	2	2	.0075

c11 *Party lines*
total=1981
average=6.60

c	*m*	*f*	*F*
0	7	300	1.000
1	9	293	.977
2	14	284	.947
3	17	270	.900
4	21	253	.843
5	40	232	.773
6	46	192	.640
7	42	146	.487
8	32	104	.347
9	17	72	.240
10	22	55	.183
11	12	33	.110
12	6	21	.070
13	10	15	.050
14	1	5	.0167
15	3	4	.0133
16	0	1	.0033
17	1	1	.0033

c12 *"Lost and found" advertisements***
total=7051
average=33.7

c	*m*	*f*	*F*
0		209	1.0000
5		209	1.0000
10		208	.9952
15		207	.9904
20		199	.952
25		182	.871
30		144	.689
35		93	.445
40		51	.244
45		21	.100
50		7	.033
55		2	.0096

d1 *Number 12*
total=421
average=4.21

c	*m*	*f*	*F*
0	2	100	1.00
1	6	98	.98
2	18	92	.92
3	13	74	.74
4	16	61	.61
5	19	45	.45
6	13	26	.26
7	5	13	.13
8	5	8	.08
9	2	3	.03
10	1	1	.01

d2 *Number 12*
total=421
average=8.42

c	m	f	F
3	1	50	1.00
4	5	49	.98
5	2	44	.88
6	6	42	.84
7	6	36	.72
8	5	30	.60
9	7	25	.50
10	6	18	.36
11	4	12	.24
12	5	8	.16
13	1	3	.06
14	0	2	.04
15	2	2	.04

d3 *Comets*
total=258
average=2.58

c	m	f	F
0	19	100	1.00
1	19	81	.81
2	17	62	.62
3	14	45	.45
4	13	31	.31
5	8	18	.18
6	4	10	.10
7	2	6	.06
8	3	4	.04
9	1	1	.01

d4 *Particles in emulsion*
total=46
average=0.92

c	m	f	F
0	22	50	1.00
1	16	28	.56
2	7	12	.24
3	4	5	.10
4	1	1	.02

d5 *Particles in emulsion*
total=106
average=2.12

c	m	f	F
0	6	50	1.00
1	11	44	.88
2	12	33	.66
3	14	21	.42
4	6	7	.14
5	1	1	.02

* M is the relative frequency of exactly c occurrences per sample. Holm does not state the actual number of samples from which this was calculated, but it was evidently at least 1000.

** Since in the range $a>30$ the curves are drawn only for every fifth value of c, in these two distributions which extend beyond $a=30$ the values of f and F are tabulated only for every fifth value of c, and the values of m, which are meaningless unless the complete series is given, are omitted.