



---

On a New Class of "Contagious" Distributions, Applicable in Entomology and Bacteriology

Author(s): J. Neyman

Reviewed work(s):

Source: *The Annals of Mathematical Statistics*, Vol. 10, No. 1 (Mar., 1939), pp. 35-57

Published by: [Institute of Mathematical Statistics](#)

Stable URL: <http://www.jstor.org/stable/2235986>

Accessed: 05/03/2013 03:16

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Institute of Mathematical Statistics* is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Mathematical Statistics*.

<http://www.jstor.org>

# ON A NEW CLASS OF "CONTAGIOUS" DISTRIBUTIONS, APPLICABLE IN ENTOMOLOGY AND BACTERIOLOGY

BY J. NEYMAN

## CONTENTS

	PAGE
1. Introduction.....	35
2. Distribution of larvae in experimental plots.....	36
3. Particular classes of the limiting distribution of $X$ .....	40
4. Certain general properties of the distributions deduced.....	43
5. Contagious distribution of type $A$ depending on two parameters.....	45
6. Contagious distribution of type $A$ depending on three parameters.....	48
7. Contagious distributions of types $B$ and $C$ .....	53
8. Illustrative examples and concluding remarks.....	54
9. References.....	57

1. **Introduction.** There are a number of fields in which experimental data cannot be treated with any success by means of the usual "Student's" test and—very probably—by means of the more general analysis of variance  $z$ -test of Fisher. It is known in fact [1] that the  $t$ -test, as applied to two samples, is only valid when the populations from which the samples are drawn have equal variances. As the  $z$ -test is of a nature similar to the  $t$ -test, with the difference that it is applied to detect differentiation in means of more than two populations, a similar conclusion seems very likely. Thus, whenever we have to compare means of populations with distinctly different variances, we have to look for some new tests. It may be useful to mention at once two instances in which the situation mentioned actually arises.

As a first instance we may quote certain entomological experiments. Suppose it is desired to test the efficiency of several treatments intended to destroy certain larvae on a field. The experiments are arranged in the usual way. The treatments compared are applied to particular plots with several replications and then the plots (or smaller parts of them) are inspected and all the surviving larvae are counted. Thus the observations represent the numbers of surviving larvae in several equal areas. It happens frequently that, while there is room for doubt as to whether there is any significant difference between the average number of survivors corresponding to particular treatments, there is no doubt whatever that the variability of the observations differs from treatment to treatment.

We have another similar case in bacteriology. The experiments I have in mind consist in determining the bacterial density by the so called "plating method." This consists in taking a number of samples of the analyzed liquid

and in spreading them separately on Petri plates. After a suitable period of time a number of colonies appear on the plates and their numbers represent the observational figures. I am informed that the variability of such observations does not depend very much on the technique of mixing the liquid and of taking the samples—when this technique is on a proper level—but does considerably depend on the kind and on the number of bacteria present in the liquid.

The above examples justify an effort to find some new and more appropriate test. The first step in this direction must consist in an analysis of the machinery behind the observable distributions and in deducing their analytical form. Once this problem is solved and repeated comparisons show a satisfactory agreement between the theory and the observation, we may proceed to the next step and deduce the appropriate tests.

The purpose of the present paper consists in deducing a family of distributions which provide a reasonably good fit in several cases in which they have been tested. It may be hoped that they will prove satisfactory also in many cases in the future.

**2. Distribution of larvae in experimental plots.** When the problem of the distribution of larvae in experimental plots first arose, attempts were made to fit the Poisson Law of frequency. These attempts, however, failed almost invariably with the characteristic feature that, as compared with the Poisson Law, there were too many empty plots and too few plots with only one larva. A similar circumstance is frequently, though not so regularly, observed in counts of microorganisms in single squares of a haemocytometer. These facts suggest that the distributions considered belong to a class which Pólya [3] proposed to call “contagious”: the presence of one larva within an experimental plot increases the chance of there being some more larvae. And it is not difficult to see the cause of this dependence. Larvae are hatched from eggs which are being laid in so-called “masses.” After being hatched they begin to travel in search of food. Their movements are slow and therefore, whenever in a given plot we find a larva, this means that the mass of eggs, from which it was hatched, must have been laid somewhere near, and this in turn means that we are likely to find in the same plot some more larvae from the same litter. Of course, there may be also others coming from other litters, too.

A similar explanation may apply also to microorganisms counted in single squares of a haemocytometer or to colonies on parallel plates. However, here the situation does not seem as clear as in the case of larvae. As far as the haemocytometer counts are concerned, also another cause of contagiousness may be suggested. Witnessing once the process of preparation of the experiment, I noticed that, immediately after the drop of liquid was deposited into the chamber of the haemocytometer and for some time after, the positions of cells seen under the microscope were not fixed. Some of them seemed to lie on the bottom and the others were floating downwards in an irregular movement. Trying to follow the movements of particular cells I had the impression

that they were slightly attracted by the cells already stationary or semi-stationary on the bottom of the chamber. If this impression of mine is justified, then the attraction of the floating cells by those already on the bottom could explain the contagiousness of the resulting distribution. It is known, however, that this contagiousness is always rather small and that frequently the distribution of cells in the squares of the haemocytometer does follow the Poisson Law very closely.

Owing to the fact that the cause of the contagiousness of the distribution of larvae in experimental plots is clear, we shall deal primarily with the distribution of larvae. Consequently, if the theoretical distributions that we shall deduce fit the empirical ones, we shall be more or less justified in assuming that we guessed the essential features of the actual machinery of movements of the larvae. On the other hand, if the same theoretical distributions appear also to fit satisfactorily the empirical counts of bacteria then in respect of these applications it will be safer to consider that we were lucky enough to find a sufficiently flexible interpolation formula.

After these preliminaries we may proceed to a more accurate specification of the conditions of the problem considered. The experimental plot in which the larvae are counted will be denoted by  $P$ . We shall make no restriction as to the shape of this plot, but we shall assume that its area, which we shall take as unity, is small compared with that of the experimental field,  $F$ . The latter will be assumed to possess  $M$  units of area. We shall further assume that the moths laying eggs on the field  $F$  select spots for this purpose in a purely random manner. This presupposes that the experimental field is uniform in many relevant respects, e.g. is sown in all its parts by the same kind of plant, etc. Denoting by  $\xi$  and  $\eta$  the coördinates of the mass of eggs laid by some particular moth on the field  $F$ , we shall treat them as random variables with the elementary probability law

$$(1) \quad p(\xi, \eta) = \frac{1}{M}$$

everywhere within  $F$  and zero elsewhere. After the larvae are hatched from the eggs there will be some mortality among them. Let us denote by  $n$  the number of larvae hatched from the same mass of eggs, surviving at the moment when the counts are made. We shall treat  $n$  as a random variable and denote by  $p(n)$  its probability law. At the present moment the writer has no information as to what may be the nature of the function  $p(n)$ . Consequently it will remain in our calculations in its general form and, wishing to obtain some formulae for immediate calculations, we shall have to substitute for  $p(n)$  hypothetical formulae which, on intuitive grounds, may seem plausible. If the larvae counted are all more or less of the same age, there is a possibility that  $p(n)$  does not differ very much from the Poisson Law, but this point might be verified experimentally and we shall not insist on its being necessarily true.

Consider now a single larva, survivor at the moment of observation, which

was hatched out at a point with coördinates  $\xi$  and  $\eta$ . Denote by  $x$  and  $y$  the coördinates of this larva at the moment of counts. We shall consider  $x$  and  $y$  as random variables. It is obvious that the probability law of  $x$  and  $y$  must depend on the values of  $\xi$  and  $\eta$ . We shall assume that the dependence is of a particular character; namely, that the probability law of  $x$  and  $y$  given  $\xi$  and  $\eta$  is a function of the differences  $x - \xi$  and  $y - \eta$ . We shall denote it by  $f(x - \xi, y - \eta)$ .

There is very little that we may consider as known about the function  $f(x - \xi, y - \eta)$ . It may be treated as describing the habits of travelling of the larvae. There are some indications that there are certain directions in which the larvae tend to travel rather than in others, but they are too vague to be taken into consideration. Only one thing is certain: during the period of time between the birth of the larvae and the moment that the counts are made the larvae are able to travel only at some limited distance. Consequently we shall assume that for sufficiently large values of  $|x - \xi|$  and  $|y - \eta|$  the function  $f(x - \xi, y - \eta)$  is identically zero. Otherwise we shall not make any further assumption concerning  $f(x - \xi, y - \eta)$ , and it will remain arbitrary in our calculations until we reach the final general formula.

While abstaining from making arbitrary assumptions concerning the habits of single larvae, we shall make one concerning the habits of several of them. This assumption, however, seems to be very plausible. We shall assume that the larvae have no social instincts, so that the random variables  $x$  and  $y$  corresponding to one larva are independent from those corresponding to any other—that is to say, apart from the possible dependence on the same pair of  $\xi$  and  $\eta$ .

Denote by  $N$  the total number of masses of eggs laid on the field  $F$  and let  $k_i$  be the number of larvae hatched from the  $i$ -th mass of eggs, surviving at the moment of observation and present within some particular experimental plot  $P$ . Finally let

$$(2) \quad X = \sum_{i=1}^N k_i$$

be the total number of larvae to be found within this plot. Our purpose will be to use the above hypotheses in order to determine the probability law of  $X$ . In doing so we shall first find that of any of the  $k_i$ 's. Obviously, when considering just one variable  $k_i$ , it would be useless to retain the subscript  $i$ , so that below we shall write simply  $k$  to denote the number of living larvae, to be found within  $P$ , all of which were hatched from the same mass of eggs, situated at some point  $(\xi, \eta)$ .

Let us first write the expression for the probability that one particular larva of that group will be found within  $P$ . This probability will be a function of  $\xi$  and  $\eta$  only, say

$$(3) \quad P(\xi, \eta) = \int \int_P f(x - \xi, y - \eta) dx dy.$$

Given that the number of survivors of the mass of eggs of the point  $(\xi, \eta)$  is  $n$ , the probability that exactly  $k$  of them will be found within  $P$  will be represented by the binomial formula, say

$$(4) \quad P\{k | n, \xi, \eta\} = \frac{n!}{k!(n-k)!} P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k}.$$

It will be noticed that in writing this formula we use the hypothesis that the larvae have no social instincts.

Multiplying (4) by the probability law of  $\xi$  and  $\eta$ , and integrating with respect to those variables over the whole field  $F$ , we shall obtain the probability,  $P\{k | n\}$  that out of the  $n$  survivors of a mass of eggs, laid anywhere within  $F$ , exactly  $k$  larvae will be found within  $P$ :

$$(5) \quad P\{k | n\} = \frac{n!}{k!(n-k)!} \frac{1}{M} \int_F \int_F P^k(\xi, \eta)(1 - P(\xi, \eta))^{n-k} d\xi d\eta.$$

Multiplying this result by  $p(n)$  and summing for all values of  $n$ , we shall obtain the absolute probability of  $k$  having any specified value.

However, before doing so, we must use the hypothesis about the function  $f(x - \xi, y - \eta)$  to deduce certain consequences concerning the integral in (5).

Originally we did not make any assumption as to the origin of coördinates on the field  $F$ . It will be now convenient to assume that it is located somewhere within the experimental plot  $P$ , for example in its center or in any other easily specified point. Owing to the particular property of the function  $f(x - \xi, y - \eta)$  it will now follow that, for sufficiently large values of  $\xi$  and  $\eta$ , the probability  $P(\xi, \eta)$  will be equal to zero. Let us denote by  $A$  the part of the experimental field where  $P(\xi, \eta) > 0$ . Obviously  $A$  denotes the set of points,  $a$ , in  $F$  such that, if a mass of eggs is laid in one of them, the distance of  $a$  from the plot  $P$  is not too large for the larvae hatched in  $a$  to reach the plot  $P$  before the moment of observation. Obviously also the plot  $P$  is included in  $A$ . Consequently the area of  $A$ , to be denoted by the same letter  $A$ , must be greater than unity. Owing to the lack of any precise knowledge of the nature of the function  $f(x - \xi, y - \eta)$  it is impossible to say anything about the shape of  $A$ .

Let us now turn to the integral in (5). The function under this integral changes its form according to whether the point  $(\xi, \eta)$  is within or without  $A$ . If  $k = 0$ , then the integral in (5) reduces to

$$(6) \quad \int_F \int_F (1 - P(\xi, \eta))^n d\xi d\eta = M - A + \int_A \int_A (1 - P(\xi, \eta))^n d\xi d\eta.$$

If however  $k > 0$ , then

$$(7) \quad \int_F \int_F P^k(\xi, \eta) (1 - P(\xi, \eta))^{n-k} d\xi d\eta = \int_A \int_A P^k(\xi, \eta) (1 - P(\xi, \eta))^{n-k} d\xi d\eta.$$

Now we can write

$$(8) \quad P\{k\} = \sum_{n \geq 0} p(n)P\{k | n\},$$



which gives in particular

$$(9) \quad P\{k = 0\} = 1 - \frac{A}{M} + \frac{1}{M} \int \int_A \sum_{n \geq 0} (1 - P(\xi, \eta))^n p(n) d\xi d\eta$$

and for  $k > 0$

$$(10) \quad P\{k\} = \frac{1}{M} \int \int_A \sum_{n \geq 0} \frac{n!}{k!(n-k)!} P^k(\xi, \eta) (1 - P(\xi, \eta))^{n-k} p(n) d\xi d\eta.$$

This is the general form of the probability law of  $k$ , which involves two unspecified functions  $p(n)$  and  $P(\xi, \eta)$ . We shall not analyze it but proceed to the calculation of the characteristic function  $\phi_k(t)$  of  $k$ , which will then be used to calculate that of  $X$ . We have

$$(11) \quad \phi_k(t) = \sum_{k \geq 0} e^{itk} P\{k\}$$

or, using (9) and (10), and after easy transformations

$$(12) \quad \phi_k(t) = 1 - \frac{A}{M} \left( 1 - \frac{1}{A} \int \int_A \sum_{n \geq 0} p(n) (P(\xi, \eta) e^{it} + 1 - P(\xi, \eta))^n d\xi d\eta \right).$$

Owing to the assumption that the larvae have no social instincts all the variables  $k_1, k_2, \dots, k_N$  in (2) must be considered as mutually independent. As the characteristic function of any of them has the same form (12), the characteristic function,  $\phi_X(t)$ , of their sum,  $X$ , will be represented by the  $N$ th power of the expression (12). Denoting by  $m$  the average number of masses of eggs per unit of area of the field  $F$ , so that  $N = Mm$ , we shall have

$$(13) \quad \begin{aligned} \phi_X(t) &= \phi_k^N(t) \\ &= \left\{ 1 - \frac{A}{M} \left( 1 - \frac{1}{A} \int \int_A \sum_{n \geq 0} p(n) (P(\xi, \eta) e^{it} + 1 - P(\xi, \eta))^n d\xi d\eta \right) \right\}^{Mm}. \end{aligned}$$

This will be the characteristic function of  $X$  for any value of  $M$ . If it is desired to put into effect the assumption that “ $M$  is large”, we shall have to consider the limit of (13) for  $M \rightarrow \infty$ . This will be denoted by  $\phi(t)$  and we shall have

$$(14) \quad \phi(t) = \exp \left\{ -Am \left( 1 - \frac{1}{A} \int \int_A \sum_{n \geq 0} p(n) (P(\xi, \eta) e^{it} + 1 - P(\xi, \eta))^n d\xi d\eta \right) \right\}.$$

In order to obtain the numerical value of the probability of  $X$  having any specified value  $X'$ , it remains only to specify the functions  $p(n)$  and  $P(\xi, \eta)$  and to use the familiar formula

$$(15) \quad P\{X = X'\} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \phi(t) e^{-itX'} dt.$$

**3. Particular classes of the limiting distribution of  $X$ .** Until we have some experimental evidence as to what might be the nature of the two functions  $p(n)$  and  $f(x - \xi, y - \eta)$  or  $P(\xi, \eta)$ , we may try a few guesses. If the results

obtained in this way agree with empirical distributions, we shall have some reason to think that the guesses are not altogether wrong.

In certain cases all the larvae considered are at the moment of observation approximately of the same age. Alternatively, we may count only larvae which are at the same stage of development. With such counts it is not unreasonable to try for  $p(n)$  either the binomial or the Poisson formula. Either of them will lead to easy calculations of (14). Writing

$$(16) \quad p(n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

with  $\lambda$  representing the average number of survivors at the moment of observation per unit mass of eggs, we shall get for  $\phi(t)$  the following expression

$$(17) \quad \phi(t) = \exp \left\{ -Am \left( 1 - \frac{1}{A} \int \int_A e^{\lambda P(\xi, \eta)(e^{it}-1)} d\xi d\eta \right) \right\}.$$

Substituting here for  $P(\xi, \eta)$  any suitable function we shall obtain a corresponding particular form of the characteristic function  $\phi(t)$ , so that (17) determines a whole family of distributions. Substituting in (14) instead of (16), say the binomial formula, we shall obtain another family of contagious distributions.

Strictly speaking, in order to obtain some particular distribution from the formula (17), we have to specify the function  $f(x - \xi, y - \eta)$ , then to calculate  $P(\xi, \eta)$  and substitute it in (17). Since however we have no knowledge of the properties of  $f(x - \xi, y - \eta)$  and have to select it only on intuitive grounds, we may as well select the function  $P(\xi, \eta)$ . It may be selected either by itself directly, in which case there will be no difficulty in substituting it in (17), or by some indirect method. In the other case we may find it more convenient to use another form of (17) which is obtained by expanding the exponential under the sign of the integral in (17) and by integrating term by term, which is obviously permissible. In this way we get

$$(18) \quad \log \phi(t) = Am \sum_{n=1}^{\infty} \frac{\lambda^n (e^{it} - 1)^n}{n!} P_n.$$

Where  $P_n$  stands for the expression

$$(19) \quad P_n = \frac{1}{A} \int \int_A P^n(\xi, \eta) d\xi d\eta$$

and has the form of a moment of  $n$ th order of a certain probability law which it is easy to determine.

We may consider for a moment the value of  $P(\xi, \eta)$  as a random variable  $Z$ . Its values cannot exceed the limits, zero and unity. Let  $z$  be any number between zero and unity and denote by  $AF(z)$  the measure of the set of points belonging to  $A$  where  $P(\xi, \eta) \leq z$ . Then the function  $F(z)$  will possess all the properties of the integral probability law of a variable  $Z$  which we may identify



with  $P(\xi, \eta)$  and the integrals  $P_n$  will be simply the moments of  $Z$  namely,  $P_n = \int_0^1 z^n dF$ , where, of course, the integral would be considered in the sense of Stieltjes. It is interesting to notice that  $P_1$  is always equal to  $A^{-1}$ . To see this consider the integral

$$(20) \quad AP_1 = \int \int_A P(\xi, \eta) d\xi d\eta$$

and substitute in it the expression of  $P(\xi, \eta)$  in terms of the function  $f(x - \xi, y - \eta)$ . We get

$$(21) \quad AP_1 = \int \int_A d\xi d\eta \int \int_P f(x - \xi, y - \eta) dx dy$$

$$(22) \quad = \iiint \int_W f(x - \xi, y - \eta) dx dy d\xi d\eta.$$

Where the four-dimensional region of integration  $W$  is defined as follows. (i) The variables  $x$  and  $y$  vary so that the point having them for its coördinates may have any position within, but cannot be outside, of the experimental plot  $P$ . (ii) When  $x$  and  $y$  are fixed in the above way, say  $x = x'$  and  $y = y'$ , then  $\xi$  and  $\eta$  may assume all those values for which the function  $f(x' - \xi, y' - \eta)$  is positive. Let us denote this system of values of  $\xi$  and  $\eta$  by  $B(x', y')$ . Then we can calculate  $AP_1$  as follows

$$(23) \quad AP_1 = \int \int_P dx dy \int \int_{B(x,y)} f(x - \xi, y - \eta) d\xi d\eta.$$

Now it is easy to see that the second integral in (23) is always equal to unity, whatever be  $x$  and  $y$  satisfying (i). To see this we have to recall the fundamental property of the function  $f(x - \xi, y - \eta)$ , due to the fact that it is the elementary probability law of  $x$  and  $y$ , namely that if  $\xi$  and  $\eta$  are fixed in one way or another, and it is integrated with respect to the other pair of variables, over all their values for which it is positive, the result will be equal to unity. In particular we shall have

$$(24) \quad \int \int_{f>0} f(u, v) du dv = 1.$$

Consider now the second integral in (23) and make the substitution

$$(25) \quad \xi = x - u, \quad \eta = y - v$$

so that, instead of  $\xi$  and  $\eta$  we shall now integrate for  $u$  and  $v$ . It will be seen that the result of this substitution is exactly the integral (24), equal to unity. Since it was assumed that the area of  $P$  is equal to unity, it follows that  $AP_1 = 1$ . This equality is thus the necessary condition that the function  $P(\xi, \eta)$  must satisfy. Besides, being a probability, it cannot be negative and cannot exceed unity. Whether any function having these properties may play the rôle of

$P(\xi, \eta)$  must be left for further inquiry. Assuming temporarily that this is so we can tentatively specify the probability laws belonging to the class determined by (18) by substituting in (18) instead of the  $P_n$ 's the corresponding moments  $M_n$  of any distribution function  $F(z)$  with its range between zero and unity, remembering only the interpretation of its first moment that we have found above, namely  $M_1 = P_1 = A^{-1}$ .

**4. Certain general properties of the distributions deduced.** Using the above result, we may substitute it in the formula (18) and get

$$(26) \quad \log \phi(t) = m\lambda(e^{it} - 1) + Am \sum_{n=2}^{\infty} \frac{\lambda^n (e^{it} - 1)^n}{n!} P_n.$$

Owing to the fact that the first term in the right hand side,  $m\lambda(e^{it} - 1)$ , represents the logarithm of the characteristic function of the Poisson Law,

$$(27) \quad p(x) = e^{-m\lambda} \frac{(m\lambda)^x}{x!}$$

for  $x = 0, 1, 2, \dots$  the formula (26) is especially interesting. Comparing the formulae

$$(28) \quad \begin{cases} P_1 = \int_0^1 z dF = A^{-1} \\ P_n = \int_0^1 z^n dF \end{cases}$$

we see that  $0 < P_n \leq A^{-1}$  so that  $AP_n \leq 1$ . This circumstance assures the absolute and uniform convergence of (26). Frequently the higher moments  $P_n$  will be much smaller than the first,  $P_1$ , and if this tends to zero, all the products  $AP_n$  for  $n \geq 2$  will do so too. In those cases  $\log \phi(t)$  will tend to  $m\lambda(e^{it} - 1)$  uniformly for all values of  $t$ . To see this take an arbitrary  $\epsilon > 0$  and select  $N$  so large that

$$(29) \quad m \sum_{n=N+1}^{\infty} \frac{(2\lambda)^n}{n!} < \frac{\epsilon}{2}.$$

Next let  $A_0$  be large enough for

$$(30) \quad AP_n < \frac{\epsilon}{2m} e^{-2\lambda}$$

for all  $n = 2, 3, \dots, N$  and for any  $A \geq A_0$ . For such values of  $A$  we shall have

$$(31) \quad \left| Am \sum_{n=2}^{\infty} \frac{\lambda^n (e^{it} - 1)^n}{n!} P_n \right| \leq Am \left( \sum_{n=2}^N \frac{\lambda^n |e^{it} - 1|^n}{n!} P_n + \sum_{n=N+1}^{\infty} \frac{\lambda^n |e^{it} - 1|^n}{n!} P_n \right) < \epsilon$$

independently of what is the value of  $t$ . This result may be formulated as

PROPOSITION I. *If the parameters  $m$  and  $\lambda$  remain constant but the probability law  $F(z)$  is changed so that all the products  $AP_n$  tend to zero for  $n = 2, 3, \dots$ , then  $\phi(t)$  tends to  $m\lambda(e^{it} - 1)$  uniformly for all values of  $t$  and, consequently, the corresponding probability law of  $X$  tends to that of Poisson, given by (27).*

The above proposition may be considered as an explanation of the circumstance that occasionally the distribution of larvae may be very close to that of Poisson. This may happen for instance when the larvae that we count are sufficiently old and have had a sufficient time to travel very far from the spot where they were hatched. In such cases  $A$  will be large and, if the function  $f(x - \xi, y - \eta)$  has some appropriate properties, all the products  $AP_n$  may be very small. But it is interesting to notice that there is a possibility of  $A$  increasing without the products  $AP_n$  tending to zero. Such will be for instance the case if  $P(\xi, \eta)$  could have within  $A$  only two values  $B_1(A)$  and  $B_2(A)$  changing with  $A$ , one close to unity and the other close to zero. If  $Ap$  and  $Aq$  are the areas of the parts of  $A$  where  $P(\xi, \eta)$  has those two different values, then we shall have

$$(32) \quad \begin{cases} P_1 = pB_1(A) + qB_2(A) = A^{-1} \\ P_n = pB_1^n(A) + qB_2^n(A) \end{cases}$$

and

$$(33) \quad AP_n = \frac{pB_1^n(A) + qB_2^n(A)}{pB_1(A) + qB_2(A)}$$

may tend to unity as  $A$  is increased. In such cases the probability law of  $X$  will not tend to (27). While calling attention to this possibility, it should be emphasized that it is not likely to occur in practice. In the cases of discontinuous  $F(z)$  considered below  $P\{X\}$  does tend to (17). The same is true also in such cases where it is assumed that

$$(34) \quad \begin{aligned} \frac{dF}{dz} &= a + bz \geq 0 && \text{for } 0 < 2 < c \leq 1 \\ &= 0 && \text{elsewhere} \end{aligned}$$

etc.

Before proceeding to specialize the expression (26) of the logarithm of the characteristic function, we shall show the connection existing between the  $P_n$ 's and the semi-invariants of  $X$ . To calculate the latter it is sufficient to differentiate (26) with respect to  $t$ , to put  $t = 0$ , and to divide the result by the appropriate power of  $i$ . Denoting by  $\gamma_k$  the  $k$ th semi-invariant, by  $\mu'_1$  the first moment about zero, and by  $\mu_k$  the  $k$ th central moment of  $X$  we easily get

$$(35) \quad \begin{cases} \mu'_1 = \gamma_1 = m\lambda \\ \mu_2 = \gamma_2 = m\lambda(1 + A\lambda P_2) \\ \mu_3 = \gamma_3 = m\lambda(1 + 3A\lambda P_2 + A\lambda^2 P_3) \\ \mu_4 - 3\mu_2^2 = \gamma_4 = m\lambda(1 + 7A\lambda P_2 + 6A\lambda^2 P_3 + A\lambda^3 P_4) \end{cases}$$

etc.

It will be seen that, in general, the  $k$ th semi-invariant depends on  $P_2, P_3, \dots, P_k$  only. Another property of the new distributions that we shall mention is that they are "stable".

PROPOSITION II. *If  $X_1, X_2, \dots, X_s$  are  $s$  independent random variables all following the same distribution with the logarithm of the characteristic function given by (26), then the sum  $Y = \sum_{i=1}^s X_i$  will follow the same probability law with the exception that instead of the parameter  $m$  it will depend on the product  $sm$ .*

In order to establish this proposition it is sufficient to notice that the logarithm of the characteristic function of the variable  $Y$  is equal to the expression (26) multiplied by  $s$ .

Lastly, it may be noticed that the family of distributions determined by (26) is different from the comparable distributions deduced by Pólya ([3], p. 153, formulae (40) and (41)). In fact the logarithms of the characteristic functions of the latter could be written as follows:

$$(36) \quad -a \log(1 - b(e^{it} - 1)) = ab(e^{it} - 1) + a \sum_{n=2}^{\infty} \frac{b^n(e^{it} - 1)^n}{n}$$

and

$$(37) \quad \frac{c(e^{it} - 1)}{1 - de^{it}} = \frac{c(e^{it} - 1)}{1 - d} + \frac{c}{1 - d} \sum_{n=2}^{\infty} \left(\frac{d}{1 - d}\right)^{n-1} (e^{it} - 1)^n$$

respectively and, even if the formal expansions in powers of  $(e^{it} - 1)$  converge, the identification of those expansions with (26) would require that  $P_n$  possess values exceeding unity, which is inconsistent with their essential property of being successive moments of a positive variable  $0 \leq Z \leq 1$ . Of course, the convergence of (36) and (37) would impose special restrictions on the constants that those formulae involve.

**5. Contagious distribution of type A depending on two parameters.** The simplest assumption that we can make concerning the function  $P(\xi, \eta)$  is that it possesses some constant positive value within  $A$  and is zero elsewhere. Owing to (20) this constant value must be equal to  $A^{-1}$ . Substituting this in (17) we immediately obtain, say

$$(38) \quad \phi_1(t) = \exp \left\{ -Am \left[ 1 - \exp \left( \frac{\lambda}{A} (e^{it} - 1) \right) \right] \right\}.$$

We could use the above formula directly to obtain the corresponding probability law. But before doing so, it may be useful to illustrate the machinery of the alternative method of obtaining the characteristic function of  $X$  and to calculate the same formula using (26).

If  $P(\xi, \eta)$  is equal to  $A^{-1}$  everywhere in  $A$ , this means that the function  $F(z)$  is a step function, which is equal to zero for any  $z < A^{-1}$  and is equal to unity

elsewhere. Accordingly we shall have  $M_n = A^{-n}$ . Substituting this into (26) instead of  $P_n$  we easily get

$$(39) \quad \log \phi_1(t) = Am \left( e^{\frac{\lambda}{A}(e^{it}-1)} - 1 \right)$$

which is equivalent with (38).

We shall now proceed to the calculation of the probabilities  $P\{x = k\}$  as determined by either (38) or (39). For this purpose it will be useful to notice that the characteristic function (38) depends really on two parameters only, which we shall denote by  $m_1$  and  $m_2$ ,

$$(40) \quad m_1 = Am, \quad m_2 = \lambda/A$$

In order to simplify the printing we shall further denote

$$(41) \quad z = m_1 e^{-m_2}$$

Expanding the two first exponentials of the three involved in (38), we may write

$$(42) \quad \phi_1(t) = e^{-m_1} \sum_{k=0}^{\infty} \frac{m_2^k}{k!} e^{ikt} \sum_{n=0}^{\infty} \frac{z^n}{n!} n^k.$$

This is the form of the characteristic function which is the most convenient when we have in mind applying the formula (15). In fact, it will be seen that we may multiply (42) by  $e^{-ix't}$  and then integrate the series term by term. Further, it will be noticed that, on integrating between the limits  $-\pi$  and  $+\pi$ , all the terms of the product will vanish except for the one which is independent of  $t$ . Consequently, the result of substituting (42) in the right hand side of (15) will be the coefficient of  $e^{ix't}$  in the expansion (42), so that

$$(43) \quad P\{X = k\} = e^{-m_1} \frac{m_2^k}{k!} \sum_{t=0}^{\infty} \frac{z^t}{t!} t^k.$$

As it is easy to verify, we have

$$(44) \quad P\{x = 0\} = e^{-m_1(1-e^{-m_2})}$$

and, for  $k \geq 1$

$$(45) \quad P\{X = k\} = e^{-m_1} \frac{m_2^k}{k!} \frac{d^k}{du^k} e^{m_1 e^{u-m_2}} \Big|_{u=0}.$$

This formula gives an easy check of the identity  $\sum_{n=0}^{\infty} P\{x = n\} = 1$ . In fact, the left hand side can be looked upon as a product of  $e^{-m_1}$  by the Taylor's expansion of the function differentiated in (45) taken at the point  $u = m_2$ , which gives identically unity.

Successive differentiations give in turn

$$(46) \quad P\{X = 1\} = e^{-m_1(1-e^{-m_2})} \frac{m_2}{1!} m_1 e^{-m_2}$$

$$(47) \quad P\{X = 2\} = e^{-m_1(1-e^{-m_2})} \frac{m_2^2}{2!} (m_1^2 e^{-2m_2} + m_1 e^{-m_2})$$

etc. Comparing the formulae (44), (46) and (47), the effect of the "contagiousness" of the distribution is easily seen.  $P\{x = 2\}$  differs from what it would have been, if the distribution was that of Poisson, by the additional term  $m_1 e^{-m_2}$  within the brackets.

Formulae (44), (46) and (47), and others which could be obtained by differentiating as indicated in (45), could be used for numerical calculations. However, these are greatly simplified by the use of the following elegant formula, deduced by Dr. Geoffrey Beall of the Dominion Entomological Experimental Station, Chatham, Ontario.

$$(48) \quad P\{X = n + 1\} = \frac{m_1 m_2 e^{-m_2}}{n + 1} \sum_{t=0}^n \frac{m_2^t}{t!} P\{X = n - t\}.$$

The correctness of this formula may be easily checked by calculating  $P\{X = n - t\}$  from (43) and by substituting it in (48). Simple rearrangements will then give what could be obtained from (43) by putting  $k = n + 1$ .

Substituting  $P_n = A^{-n}$  in formulae (35) and taking account of (40), we get

$$(49) \quad \mu'_1 = \lambda m = m_1 m_2$$

$$(50) \quad \mu_2 = \lambda m \left(1 + \frac{\lambda}{A}\right) = m_1 m_2 (1 + m_2).$$

Solving these equations for  $m_1$  and  $m_2$  we obtain the formulae

$$(51) \quad m_2 = (\mu_2 - \mu'_1)/\mu'_1, \quad m_1 = \mu'_1/m_2$$

If the moments  $\mu'_1$  and  $\mu_2$  are determined for an empirical distribution, these formulae may be used for estimating  $m_1$  and  $m_2$ . In cases which were tried, this process did give frequently a satisfactory fit. Sometimes, however, when the tail of the original empirical distribution was very irregular, this distribution was better approximated by calculating the moments  $\mu'_1$  and  $\mu_2$  not from itself but after a certain amount of smoothing of the tail. It follows that the method of fitting the new distribution to the empirical data requires some further study. At present it will suffice to mention that, whenever this distribution was tried on distributions of larvae which at the moment of counts were approximately of the same stage of development, the fit obtained was very satisfactory. It is hoped that a number of actual distributions fitted, together with the description of the method of counting, etc., will be soon published by Dr. Beall. As a matter of illustration one of his distributions is reproduced at the end of the present paper.



As for the distribution considered we have

$$(52) \quad \lim_{A \rightarrow \infty} AP_n = \lim_{A \rightarrow \infty} A^{-n+1} = 0, \quad n = 2, 3, \dots$$

It follows from the above theory that, as  $A \rightarrow \infty$ , the probability law (48) tends to that of Poisson, namely

$$(53) \quad \lim_{A \rightarrow \infty} P\{X = n\} = e^{-m_1 m_2} \frac{(m_1 m_2)^n}{n!}.$$

For this reason the distribution (48) could be perhaps called the generalized probability law of Poisson, but it seems that the term "contagious distribution of type A with two parameters" will be more descriptive. Further on we shall see what is the justification of the description "of type A".

It was stated at the outset of the present paper that, when comparing the distributions of larvae in two series of plots subjected to two different treatments, there is sometimes doubt whether the means of those distributions are equal or not, while the difference in variability is more or less obvious. The formulae (49) and (50) give us the explanation of these facts. It is seen from the formula (49) that the mean of the distribution is equal to the product of the mean number of masses of eggs per unit of area and of the mean number of larvae per mass of eggs surviving at the moment of counts. If the two treatments compared are of about the same efficiency of killing the larvae, then the values of  $\lambda$  for each of them will be approximately equal and, consequently, we shall obtain about the same values for the two means. But while being of an equal efficiency as far as the killing is concerned, the two treatments may annoy the larvae in an unequal way. For example if the first treatment is dummy (no treatment) and the other is in general ineffective, it may still spoil the taste of the leaves that the larvae feed on. In such a case they may be compelled to travel a little more than they would otherwise, which will lead to an increase in  $A$ . Looking at the formula (50), it is easy to see that this would lead to a decrease in the value of  $\mu_2$ . Alternatively the treatment may produce a temporary paralysis of the larvae which may reduce  $A$  and bring an increase of  $\mu_2$ .

These remarks were applied to moments (49) and (50) of the particular distribution (45), but looking at the formulae (35), it is easily seen that they are true in the general case also.

**6. Contagious distributions of type A depending on three parameters.** As mentioned before, in order to determine some particular contagious distribution contained in the class depending on equation (18) it is sufficient to substitute in it instead of the  $P_n$  the moments of any distribution with its range confined to the interval from zero to unity, with the only restriction that the reciprocal of the first moment should be equal to  $A$ . Obviously this could be done in an infinity of ways, all of which will give more or less different results. We shall select the following one, representing a natural generalization of the procedure adopted above and leading to very simple formulae.

Formerly we have assumed that  $P(\xi, \eta)$  possesses a constant value  $A^{-1}$  within the whole area  $A$ . At present we may assume that within this area it may possess one of two (three, four, etc.) values, say  $B_1$  and  $B_2$ . Considering again  $P(\xi, \eta)$  as a random variable  $Z$ , this will be equivalent to an assumption that  $Z$  may possess only one of the values  $B_1$  and  $B_2$  both positive and not exceeding unity. Again the probabilities of  $Z = B_i$  are at our disposal. We shall take that these probabilities are equal, i.e. equal to  $\frac{1}{2}$ .

Comparing these assumptions with what may be the actual situation, one may be led to think that they are rather artificial. This however is not so. There is no doubt that the value of  $P(\xi, \eta)$  does change within  $A$ , and it is also probable that the change is smooth. As we have no knowledge of the character of this function we first take its mean value within the area  $A$  and treat it as its first approximation. Next we divide the area  $A$  into two equal parts, say  $A_1$  and  $A_2$  and so that the greatest value of  $P(\xi, \eta)$  in  $A_1$  does not exceed any of the values in  $A_2$ . Then taking the average of  $P\{\xi, \eta\}$  within  $A_1$  and a similar average within  $A_2$  and denoting them by  $B_1$  and  $B_2$  respectively, we do obtain a better approximation to the actual values of  $P(\xi, \eta)$  assuming that it is equal to  $B_i$  everywhere in  $A_i$ . That is, in fact, the real meaning of the hypothesis formulated above and that we are going to accept in the following.

Denoting again by  $M_n$  the moments of  $Z$  we shall have

$$(54) \quad M_1 = \frac{1}{2}(B_1 + B_2) = A^{-1}$$

and generally

$$(55) \quad M_n = \frac{1}{2}(B_1^n + B_2^n).$$

Substituting (55) in (26) we get, say

$$(56) \quad \phi_2(t) = \frac{Am}{2} (e^{\lambda B_1(e^{it}-1)} + e^{\lambda B_2(e^{it}-1)} - 2).$$

We notice that this expression depends on three parameters, say

$$(57) \quad m_1 = Am, \quad m_2 = \lambda B_1, \quad m_3 = \lambda B_2.$$

In order to get the formulae for the probabilities of  $X$  having any specified values we could again apply the method used above when treating the more simple case. It may be useful however to illustrate a shorter way which easily leads to a generalization of Dr. Beall's recurrence formula. As we have noticed before, the probability  $P\{X = k\}$  is equal to the coefficient of  $e^{ikt}$  in the expansion of the characteristic function in powers of  $e^{it}$ . Substituting for simplicity  $z = e^{it}$ , so that  $t = -i \log z$ , we may say that, if  $\phi(t)$  is the characteristic function of a variable  $X_1$  which is able to possess only integer values, then  $P\{X = k\}$  is equal to the coefficient of  $z^k$  in the expansion of say  $\psi(z) = \phi(-i \log z)$ . Applying this rule to (56) we can write the following expression for the generating function  $\psi(z)$ ,

$$(58) \quad \psi(z) = e^{-m_1} e^{\frac{1}{2}m_1 \{e^{m_2(z-1)} + e^{m_3(z-1)}\}} = \sum_{k=0}^{\infty} z^k P\{X = k\}.$$

In other words

$$(59) \quad P\{X = 0\} = \psi_{(0)} = e^{-m_1} e^{\frac{m_1}{2} \{e^{-m_2} + e^{-m_3}\}}$$

$$(60) \quad P\{X = k\} = \frac{1}{k!} \left. \frac{d^k \psi}{dz^k} \right|_{z=0}, \quad k = 1, 2, \dots$$

But

$$(61) \quad \begin{aligned} \frac{d\psi}{dz} &= \frac{m_1}{2} \psi(z) \{m_2 e^{m_2(z-1)} + m_3 e^{m_3(z-1)}\} \\ &= \frac{m_1}{2} \psi(z) \chi(z) \quad (\text{say}) \end{aligned}$$

and it is easy to see that generally

$$(62) \quad \frac{d^k \chi}{dz^k} = m_2^{k+1} e^{m_2(z-1)} + m_3^{k+1} e^{m_3(z-1)}.$$

As the  $k$ th derivative of  $\psi(z)$  in (60) may be calculated by applying the familiar formula for the  $(k - 1)$ st derivative of the product  $\psi(z)\chi(z)$  in (61), we obtain

$$(63) \quad \left. \frac{d^{n+1} \psi}{dz^{n+1}} \right|_{z=0} = \frac{m_1}{2} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \left( \left. \frac{d^k \chi}{dz^k} \frac{d^{n-k} \psi}{dz^{n-k}} \right) \right|_{z=0}.$$

Using the formulae (60) and (62) we immediately obtain

$$(64) \quad P\{X = n + 1\} = \frac{m_1}{2(n+1)} \sum_{k=0}^n \frac{m_2^{k+1} e^{-m_2} + m_3^{k+1} e^{-m_3}}{k!} P\{X = n - k\}.$$

As whenever  $B_1 = B_2$  and consequently  $m_2 = m_3$ , the distribution considered now becomes identical with that considered formerly, depending on two parameters only, it is seen that the formula (64) represents a direct generalization of the formula (48). For purposes of successive calculation of the probabilities it will be probably more convenient to write (64) in the following form

$$(65) \quad \begin{aligned} P\{X = n + 1\} &= \frac{m_1 m_2 e^{-m_2}}{2(n+1)} \sum_{k=0}^n \frac{m_2^k}{k!} P\{X = n - k\} \\ &\quad + \frac{m_1 m_3 e^{-m_3}}{2(n+1)} \sum_{k=0}^n \frac{m_3^k}{k!} P\{X = n - k\}. \end{aligned}$$

This device of finding a recurrence formula for the probabilities will always succeed whenever there are no difficulties in finding the value of the  $n$ th derivative of the function  $\chi$ .

It may be easily shown that if  $m$  and  $\lambda$  remain fixed but  $A$  tends to infinity, then the distribution (60) tends to the Poisson Law of frequency. Owing to

the general result stated in Proposition I, in order to show this it is only sufficient to prove that for  $n \geq 2$

$$(66) \quad \lim_{A \rightarrow \infty} AM_n = \lim_{A \rightarrow \infty} \frac{B_1^n + B_2^n}{B_1 + B_2} = 0.$$

As both  $B_1$  and  $B_2$  must be included between zero and unity and their sum is equal to  $2A^{-1}$ , it follows that

$$(67) \quad 0 < B_1 \leq A^{-1} \leq B_2 < 2A^{-1}.$$

Therefore

$$(68) \quad 0 < AM_n < \frac{1 + 2^n}{2} A^{-n+1}$$

and (66) becomes obvious.

Substituting the values of  $M_2$  and  $M_3$  instead of  $P_2$  and  $P_3$  in the general expressions (35) of the moments, and taking into account the formulae (57), we obtain

$$(69) \quad \begin{cases} \mu'_1 = \frac{1}{2}m_1(m_2 + m_3) \\ \mu_2 = \frac{1}{2}m_1(m_2 + m_3 + m_2^2 + m_3^2) \\ \mu_3 = \frac{1}{2}m_1(m_2 + m_3 + 3(m_2^2 + m_3^2) + m_2^3 + m_3^3). \end{cases}$$

If it is desired to fit the distribution to some empirical one using the method of moments, then these formulae could be solved with respect to  $m_1$ ,  $m_2$  and  $m_3$ . We may proceed as follows. Write

$$(70) \quad a = 2\mu'_1, \quad b = 2(\mu_2 - \mu'_1), \quad c = 2(\mu_3 + 3\mu_2 + 2\mu'_1).$$

Then

$$(71) \quad m_1(m_2 + m_3) = a$$

$$(72) \quad m_1(m_2^2 + m_3^2) = b$$

$$(73) \quad m_1(m_2^3 + m_3^3) = c.$$

Multiplying the first of these equations by  $m_2$  and subtracting the result from the second and repeating the same process with the second equation and the third, we get

$$(74) \quad m_1m_3(m_3 - m_2) = b - am_2$$

$$m_1m_3^2(m_3 - m_2) = c - bm_2$$

and it follows

$$(75) \quad m_3 = \frac{c - bm_2}{b - am_2}$$

or

$$(76) \quad \frac{b}{a} (m_2 + m_3) - m_2 m_3 = \frac{c}{a}.$$

Again, dividing (73) by (71) we get

$$(77) \quad (m_2 + m_3)^2 - 3m_1 m_2 = \frac{c}{a}.$$

Multiplying (76) by 3 and subtracting from (77), we obtain

$$(78) \quad s^2 - 3bs/a - 2c/a = 0,$$

where  $s = m_2 + m_3$ . It follows that

$$(79) \quad s = \frac{3b}{2a} \pm \sqrt{\left(\frac{3b}{2a}\right)^2 - \frac{2c}{a}}$$

$$(80) \quad m_2 m_3 = p = \frac{bs - c}{a}$$

$$(81) \quad m_2 = \frac{1}{2}(s - \sqrt{s^2 - 4p})$$

$$(82) \quad m_3 = \frac{1}{2}(s + \sqrt{s^2 - 4p})$$

$$(83) \quad m_1 = a/s.$$

Following these steps we finally arrive to the values of all three parameters, given by the last three formulae.

If the values of the moments  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  were known without error, the above formulae would give accurate values of  $m_1$ ,  $m_2$  and  $m_3$ . If, however, the moments are estimated from a sample, then the reader must be prepared that, even if the observed variable follows exactly the law, occasionally the sampling errors in the moments will make it impossible to carry out all the calculations indicated. Especially this may easily happen when the true values of  $m_2$  and  $m_3$  are equal or nearly equal, so that the empirical distribution is close to that given by the contagious distribution with only two parameters. As it is seen from (81) and (82), in such a case the true values of  $s$  and  $p$  must satisfy the relation

$$(84) \quad s^2 - 4p = 0.$$

However, the sampling errors in the moments will ascribe to the left hand side of (84) a value only approximately equal to zero, which may be either positive or negative. In the latter case we shall not be able to use (81) and (82) to estimate  $m_2$  and  $m_3$ . As a matter of fact, the above circumstance actually arose in one case when it was tried to fit the three parameters distribution to a set of data which were excellently fitted by a simpler formula (45) involving only two parameters. As mentioned before, the problem of fitting the distributions which are deduced here requires further consideration.

Looking back on the method by which we have substituted a contagious distribution with three parameters  $m_1, m_2, m_3$  for the simpler one with only two parameters, it is easily seen that it can be carried further leading to distributions with four, five, etc. parameters. In each case we would mentally divide the area  $A$  in a number of parts of equal size so that the values of  $P(\xi, \eta)$  in the first never exceed those in the second, etc. Denoting the average values of  $P(\xi, \eta)$  in those areas by  $B_1, B_2, \dots, B_r$ , we shall obtain the moments

$$(85) \quad M_n = \frac{1}{r} \sum_{i=1}^r B_i^n,$$

substitute them in (26) and proceed more or less as we did above. All the distributions which may be obtained in this way possess certain common traits and I propose to call them "of type  $A$ ". If the number of parameters in such a distribution is sufficiently high, it seems practically certain that the function  $P(\xi, \eta)$  will be well approximated and we may hope to get an excellent fit. However, if a good fit may be attained only by introducing a great number of parameters, it usually means that the method of introducing those parameters is not very successful, and therefore it does not seem worth while to discuss in greater detail the distributions of type  $A$  with the number of parameters exceeding three. Instead we shall briefly indicate another class of distributions, built on another principle, which may be called of type  $B$  or  $C$ .

**7. Contagious distributions of types  $B$  and  $C$ .** As mentioned before, whenever the distributions of type  $A$  were tried on data, the character of which did not obviously contradict the basic assumptions of the theory (approximate equality of age of the larvae), the results were always satisfactory. However, our present experience is rather limited and it is well to anticipate the failures. We may expect that these will be caused by the over-simplified assumptions concerning the function  $P(\xi, \eta)$ .

In order to deal with such a case we may assume that for  $0 < z < 1$  the derivative of  $F(z)$  exists and is either a linear function of  $z$  or is equal to zero. Writing  $p(z) = dF/dz$  we shall put

$$(86) \quad \begin{aligned} p_1(z) &= \frac{1}{2}A \quad \text{for } 0 < z < 2A^{-1}, \quad A \geq 2 \\ &= 0 \quad \text{elsewhere.} \end{aligned}$$

Alternatively we may write, say

$$(87) \quad \begin{aligned} p_2(z) &= \frac{2A^2}{9} (3A^{-1} - z) \quad \text{for } 0 < z < 3A^{-1} \\ &= 0 \quad \text{elsewhere.} \end{aligned}$$

In the first case we shall obtain, say

$$(88) \quad M'_n = \frac{1}{n+1} \left( \frac{2}{A} \right)^n.$$



On the other hand, the moments of  $p_2(z)$  will be given by

$$(89) \quad M''_n = \frac{2(3A^{-1})^n}{(n+1)(n+2)}.$$

Substituting these expressions in (26) we shall easily obtain the two new forms of the characteristic function of  $X$ , say

$$(90) \quad \log \phi_3(t) = -m_1 + m_1 \frac{e^{m_2(e^{it}-1)} - 1}{m_2(e^{it} - 1)},$$

with

$$(91) \quad m_1 = Am \quad \text{and} \quad m_2 = 2\lambda/A.$$

Accordingly, the generating function of the probabilities will be, say

$$(92) \quad \psi_3(z) = e^{-m_1} e^{m_1 \frac{e^{m_2(z-1)} - 1}{m_2(z-1)}} = \sum_{n=0}^{\infty} z^n P\{X = n\}.$$

The distribution determined by (92) may be called of type  $B$ .

Using the moments (89) and substituting them in the usual way in (26), we obtain, say

$$(93) \quad \log \phi_4(t) = -m_1 + 2m_1 \frac{e^{m_2(e^{it}-1)} - 1 - m_2(e^{it} - 1)}{m_2^2(e^{it} - 1)^2},$$

with

$$(94) \quad m_1 = Am \quad \text{and} \quad m_2 = 3\lambda/A.$$

The probabilities of  $X$  having any specified value will be generated by the function, say

$$(95) \quad \psi_4(z) = e^{-m_1} e^{m_1 \frac{e^{m_2(z-1)} - 1 - m_2(z-1)}{m_2^2(z-1)^2}} = \sum_{n=0}^{\infty} z^n P\{X = n\}.$$

The probability law determined by (95) may be called of type  $C$ . The comparative merits of all those distributions could be judged by comparing them with the results of observation.

**8. Illustrative Examples and Concluding Remarks.** Any series of positive terms adding up to unity may be considered as determining a probability law of a discontinuous variable such as the  $X$  considered above. When trying to obtain probability laws fitting the empirical distributions of some particular origin, the distributions of the numbers of larvae in experimental plots, or the like, we could really start by considering series of some positive terms each depending on one or more parameters, say

$$(96) \quad u_0(m_1, m_2), u_1(m_1, m_2), u_2(m_1, m_2), \dots, u_n(m_1, m_2), \dots$$

and having the property that, whatever the values of those parameters,  $\sum_{n=0}^{\infty} u_n(m_1, m_2) = 1$ . Studying a considerable number of empirical distributions,

we could apply the "method" of trial and error to guess the form of dependence of the  $u_n(m_1, m_2)$  on the  $m$ 's so that for a broad class of empirical distributions there would be a system of values of the  $m$ 's, for which the series (96) would satisfactorily fit the data. If we succeed in this task we shall be entitled to a considerable satisfaction as the solution that we obtained would permit various further studies, e.g. the deduction of tests of significance applicable, or approximately applicable, in various cases, and so on.

Looking back at the history of statistics we shall find that the systems of frequency curves of Pearson, of Bruns-Charlier and others belong to the class of results just discussed. They are very important—and this especially applies to the Pearson curves—because of the empirical fact, that it is but rarely that we find in practice an empirical distribution, which could not be satisfactorily fitted by any of such curves. Consequently, wishing to deduce some test applicable in this or that case, we may usefully assume that the basic distribution is one of the Pearson system and, owing to the frequently continuous character of the connection between the conditions and the final results, our final formula will be approximately valid when applied to the data under consideration.

This point of view is not unfamiliar in pure mathematics. For example, we know that a broad class of functions may be approximated with any prescribed accuracy by means of polynomials. Wishing to prove a theorem applicable to this class of functions, we sometimes start by proving it for polynomials and then conclude that it is also true for the whole class. Here the rôle of polynomials is perfectly analogous to that of Pearson curves and could be described as that of good interpolation formulae.

But the problem of deducing theoretical distributions could be also considered from a slightly different point of view. Here again we require that the theoretical distribution fits satisfactorily the empirical data. But we may legitimately require something else: an "explanation" of the machinery producing the empirical distributions of a given kind. I have enclosed the word "explanation" in quotation marks so as not to suggest that I am attaching to it too much importance. Mathematics is always dealing with the conceptual sphere which is quite distinct from the perceptual and, at most, admits the possibility of establishing some correspondence. Therefore, however hard we try, we can never produce anything like a real mathematical explanation of any phenomena but instead only some "interpolation formula", some system of conceptions and hypotheses, the consequences of which are approximately similar to the observable facts. But this similarity may be differently placed. In the case of Pearson's curves it applies to the shape of these curves and to the shape of the empirical histograms. Otherwise it may apply to certain real features of the phenomena studied and to some mathematically described model of the same phenomena. And if the theoretical distributions deduced from the mathematical model do agree with those that we observe, and if that agreement is more or less permanent, we say that the mathematical model has "explained" the origin of the distributions.

If the problem of deducing interpolation formulae, sufficiently flexible to represent adequately a class of distributions, is of considerable interest, then that of producing similar formulae but involving an "explanation" of the phenomena studied, seems to be still more interesting. Of course, for it to be considered as successfully solved, the theoretical distributions deduced must fit the empirical ones, of a clearly specified kind, "practically always". At the

TABLE I

*Distribution of European corn borers in 120 groups of 8 hills each, (data provided by Dr. Beall), fitted by Poisson Law and by type A Law with two parameters*

No. of borers	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	5.0	24	22.6
1	16.0	16	16.7
2	25.3	16	18.3
3	26.7	18	16.4
4	21.1	15	13.4
5	13.4	9	10.3
6	7.1	6	7.5
7	3.2	5	5.2
8	1.3	3	3.5
9	.4	4	2.3
10	.1	3	1.5
11		0	
12		1	
Beyond		—	2.3
$m_1$	—	—	2.178
$m_2$	—	—	1.454
$P_{\chi^2}$	.000,000		.95

TABLE II

*Distribution of yeast cells in 400 squares of haemocytometer observed by "Student" (1907), fitted by Poisson Law and by type A Law with two parameters*

No. of cells	Frequency		
	Exp. P. L.	Observed	Exp. T. A.
0	202	213	214.8
1	138	128	121.3
2	47	37	45.7
3	11	18	13.7
4		3	3.6
5		1	.8
Beyond	2	—	.1
$m_1$	—	—	3.605
$m_2$	—	—	.189
$P_{\chi^2}$	> .02		> .1

present time we may quote a number of instances where it was possible to establish a mathematical probabilistic model of some class of phenomena determining probability laws which fit the empirical distributions with a remarkable accuracy. Perhaps the most important class of these phenomena is provided by the Mendelian theory; a number of other examples, although of a lesser importance but still interesting, have been mentioned elsewhere [2]. In all of them success-

ful checks and rechecks increase our confidence that the conclusions based on the mathematical model determining the theoretical distributions will satisfactorily apply to observational data and also that our interpretations of various constants is more or less correct.

Now, what is the situation with the contagious distributions deduced above? They do represent an attempt to give good interpolation formulae involving an "explanation" of the observable phenomena, and all the constants introduced have meanings which are easy to interpret. Owing to the fact that in the process of the larvae surviving and spreading over the field there are certain unknown features, the final general formula that we have deduced involves two arbitrary functions  $p(n)$  and  $P(\xi, \eta)$ . By substituting for them any appropriate functions that the intuition may suggest, we can obtain a number of distributions, each of which may or may not provide a satisfactory interpolation formula. Whether they do or not, must be empirically tested.

Up to the present time the contagious distributions of type *A* were tried on 12 distributions of larvae and on three distributions of yeast cells in squares of the haemacytometer, which did not quite agree with the Poisson Laws. The results of these trials were always the same: The type *A* distribution with two parameters provided an excellent fit, which was never worse than that of the more elaborate distribution with three parameters. This circumstance seems encouraging, but future experience may be less satisfactory and it would be very desirable to have some more empirical distributions and checks.

The following table gives two empirical distributions fitted with Poisson Law and with its generalization, as provided by the type *A* distribution with two parameters.

#### REFERENCES

- [1] P. L. HSU, "Contribution to the Theory of 'Student's' *t*-Test as Applied to the Problem of Two Samples." *Statistical Research Memoirs*, Vol. II, (1938) pp. 1-24.
- [2] J. NEYMAN: "Lectures and Conferences on Mathematical Statistics," published by the Graduate School of the United States Department of Agriculture, Washington, D. C., 1938.
- [3] G. PÓLYA: "Sur quelques points de la théorie des probabilités." *Annales de l'Institut Henri Poincaré*, Vol. 1, (1931) pp. 117-162.
- [4] "Student": "On the Error of Counting with a Haemacytometer," *Biometrika*, Vol. 5, (1907), pp. 351-364.

UNIVERSITY OF CALIFORNIA,  
BERKELEY, CALIFORNIA.