

## ON A GENERAL CLASS OF "CONTAGIOUS" DISTRIBUTIONS

BY W. FELLER

*Brown University*

**1. Introduction.** In a paper of considerable interest, J. Neyman [11] recently discussed frequently occurring situations where the usual tests of significance fail. He discussed, in particular, experiences in entomology and bacteriology which cannot be described by the usual distribution functions and he constructed several new types of apparently contagious distributions. Now at first glance Neyman's investigation may seem of a rather specialized nature, and his distributions of a restricted applicability. It may therefore be useful to point out that they are intimately related to results obtained by various authors in connection with topics having so little apparent relation as accident statistics, telephone traffic, fire damage, sickness- and life-insurance, risk theory, and even an engineering problem. Viewed in the proper light of a general theory, Neyman's method is particularly closely related to some too little known considerations by Greenwood and Yule [6]. These authors were the first to find, and apply, the distribution which shortly afterwards was independently rediscovered by Eggenberger and Polya<sup>1</sup> [3, 4].

Greenwood and Yule discussed two types of what may conveniently be called contagion: with one type there is *true contagion* in the sense of Polya and Eggenberger, where each "favorable" event increases (or decreases) the probability of future favorable events; with the second type the events are, strictly speaking, independent and an *apparent contagion* is actually due to an inhomogeneity of the population. The two explanations are very different in nature as well as in practical implications. It is therefore most remarkable that Greenwood and Yule found their distribution assuming an *apparent contagion*; in their opinion this distribution *contradicts* true contagion. On the contrary, Polya and Eggenberger arrived at the *same* distribution assuming *true* contagion, while the possibility of an apparent contagion due to inhomogeneity seems not to have been noticed by them. The Greenwood-Yule-Polya-Eggenberger distribution has found many applications.<sup>2</sup> Therefore the possibility of its interpretation in two ways, diametrically opposite in their nature as well as in their implications is of greatest statistical significance. This fact is, incidentally, a justification for general theories in statistics.

We shall see that Neyman's contagious distributions belong to the *second* type and are related to the Polya-Eggenberger distribution only if the latter is

---

<sup>1</sup> The fact that the Polya-Eggenberger distribution is identical with the Greenwood-Yule distribution seems to be mentioned in the literature only in a Stockholm thesis by O. Lundberg [9].

<sup>2</sup> Of quite recent applications we mention Kitagawa and Huruya [8], Rosenblatt [15], O. Lundberg [9]. Only the latter seems aware of the double nature of the distribution.

interpreted in the sense of Greenwood and Yule. In Neyman's case as well as in the other cases referred to above we are concerned with inhomogeneous populations and there exists an extremely simple device to describe such situations appropriately. Once stated, this device will appear trivial. Nevertheless, a straightforward application of it would have avoided considerable mathematical difficulties in the literature and, occasionally, yielded better and simpler results. It seems also the simplest description of the mechanism behind many observed distributions, and therefore suited for a theory of tests<sup>3</sup>.

To start in a purely formal manner, consider an arbitrary cumulative distribution function (c.d.f.)  $F(x, a)$ , depending on a parameter  $a$ , and another c.d.f.  $U(a)$ . Then

$$(1.1) \quad G(x) = \int F(x, a) dU(a)$$

(the integration extending over the domain of variation of  $a$ ) is again a c.d.f. If, in particular,  $U(a)$  is a step function, (1.1) reduces to

$$(1.2) \quad G(x) = \sum p_i F(x, a_i),$$

where  $p_i$  is the weight attached to  $a_i$  (we have, of course,  $p_i \geq 0$ ,  $\sum p_i = 1$ ). Instead of (1.2) one can write more simply

$$(1.3) \quad G(x) = \sum p_i F_i(x),$$

where the  $F_i(x)$  are arbitrary c.d.f.'s. Of course,  $F(x, a)$  and  $U(a)$  may depend on additional parameters, and the procedure can be repeated.

The statistical meaning of (1.3) is clear. Consider a population made up of several subgroups  $A_1, A_2, \dots$ , mixed at random in proportions  $p_1:p_2:\dots$ . If  $F_i(x)$  is the c.d.f. of some character in  $A_i$ , then  $G(x)$ , as defined by (1.3), will represent the c.d.f. of that character in the total population, provided that the subgroups  $A_i$  are statistically independent. Similarly (1.1) describes an infinitely composite population. Postponing a discussion of the property of contagion to the last section, we shall first deduce a few properties of the *compound Poisson-distribution*, considered first by Greenwood and Yule. Neyman's "Contagious Distributions of Type A" as well as the Polya-Eggenberger distribution belong to this class. Our next example of a special case of (1.1) is what F. E. Satterthwaite [16] called the "Generalized Poisson Distribution." It has been independently discovered by many authors and represents heterogeneity of quite different a nature. Instead of further examples we shall, in the fourth section, show how Neyman's most general contagious distribution can be deduced by a repeated application of (1.1).

<sup>3</sup> Incidentally, attention may be drawn to an argument by Greenwood and Yule showing that the  $\chi^2$ -test when applied to the Poisson distribution is biased and tends to exaggerate the goodness of fit. The argument could be amplified from other experience.

*Notation:* If  $F(x)$  and  $G(x)$  are the c.d.f.'s of two independent variates  $X$  and  $Y$ , then their *convolution*, (that is to say the c.d.f. of  $X + Y$ ) will be denoted by  $F(x)*G(x)$ . Thus

$$(1.4) \quad F(x)*G(x) = \int_{-\infty}^{+\infty} F(x - y) dG(y).$$

More particularly we shall write

$$(1.5) \quad \begin{aligned} F(x)*F(x) &= F^{2*}(x), \\ F^{n*}(x)*F(x) &= F^{(n+1)*}(x). \end{aligned}$$

We shall denote by  $E(x)$  the *unitary c.d.f.*

$$(1.6) \quad E(x) = \begin{cases} 0 & \text{for } x < 1, \\ 1 & \text{for } x \geq 1, \end{cases}$$

so that  $E^{n*}(x) = 0$  for  $x < n$ , and 1 for  $x \geq n$ .

**2. The compound Poisson distribution.** Consider the well-known Poisson expression

$$(2.1) \quad \pi(n; a) = e^{-a} \frac{a^n}{n!},$$

where the parameter  $a > 0$  gives the expected number of “events”. We shall refer to (2.1) as the *simple Poisson distribution*. If different individuals of a population are associated with different values of  $a$ , and if the character  $a$  is distributed according to the cumulative probability law  $U(a)$ , the probability of  $n$  events in the total population will be given by

$$(2.2) \quad \pi_n = \int_0^{\infty} e^{-a} \frac{a^n}{n!} dU(a).$$

Following Greenwood and Yule we shall refer to (2.2) as the *compound Poisson distribution*. Referring for an interpretation to the last section, we first consider a few special cases.

a) If  $U(a)$  is a step function we are led to expressions of the form

$$(2.3) \quad \pi_n = \frac{1}{n!} \sum_i p_i e^{-a_i} a_i^n.$$

Such a distribution has been successfully applied by C. Palm [12] to problems of telephone traffic, and by O. Lundberg [9] to sickness statistics.

b) If  $U(a)$  is a Pearson Type III distribution

$$(2.4) \quad U'(a) = \left(\frac{1}{d}\right)^h \frac{1}{\Gamma\left(\frac{h}{d}\right)} e^{-\frac{a}{d}} x^{\frac{a}{d}-1}$$

(with  $d > 0, h > 0$ ), then

$$(2.5) \quad \pi_n = \frac{1}{n!} \frac{\Gamma\left(n + \frac{h}{d}\right)}{\Gamma\left(\frac{h}{d}\right)} (1+d)^{-\frac{d}{h}} \left(\frac{1}{1+d}\right)^n.$$

This is the *Polya-Eggenberger distribution* in its usual form, and has in this form (with a slight change of notations) been derived by Greenwood and Yule.

c) If  $a$  takes on the values  $kc$  only, where  $c > 0$  is a constant and  $k = 0, 1, \dots$ , and if  $a$  is distributed according to the Poisson law

$$(2.6) \quad \text{Prob}\{a = kc\} = e^{-\lambda} \frac{\lambda^k}{k!},$$

then

$$(2.7) \quad \pi_n = e^{-\lambda} \frac{c^n}{n!} \sum_{k=0}^{\infty} \frac{k^n}{k!} (e^{-c} \lambda)^k.$$

This is Neyman's *contagious distribution of type A* depending on two parameters (cf. section 4). If, instead,  $a$  is distributed according to a multiple Poisson law of form (2.3) we arrive at Neyman's *more-parametric distribution of type A*. They are, of course, essentially linear combinations of expressions of form (2.7).

It follows from the theory of Laplace transforms that *two compound Poisson distributions associated with different c.d.f.'s  $U(a)$  are never identical*.

The compound Poisson distribution gives a simple explanation of a phenomenon recorded by Neyman and observable in many instances. In the experiments described by Neyman "the attempts to fit the Poisson Law  $\dots$  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". It is easily checked in the literature that similar situations arise frequently. Now the Poisson distribution is usually fitted by the method of moments. Accordingly, the compound Poisson law (2.2) ought to be compared with the simple Poisson distribution with the same mean value. The mean value of (2.2) is

$$(2.8) \quad m = \int_0^{\infty} a dU(a),$$

so that (2.2) ought to be compared with the Poisson distribution  $\pi(n; m)$ . Now, whatever the c.d.f.  $U(a)$ , we have always

$$(2.9) \quad \pi_0 \geq \pi(0, m)$$

and

$$(2.10) \quad \frac{\pi_1}{\pi_0} \leq m = \frac{\pi(1, m)}{\pi(0, m)}.$$

As a matter of fact, using Lagrange’s form for the remainder in Taylor’s formula, we have

$$\begin{aligned}
 \pi_0 &= e^{-m} \int_0^\infty e^{m-a} dU(a) \\
 (2.11) \quad &\geq e^{-m} \int_0^\infty \{1 + (m - a)\} dU(a) = e^{-m} = \pi(0, m),
 \end{aligned}$$

which proves (2.9). Similarly

$$\begin{aligned}
 m\pi_0 - \pi_1 &= e^{-m} \int_0^\infty e^{m-a}(m - a) dU(a) \\
 (2.12) \quad &\geq e^{-m} \int_0^\infty (m - a) dU(a) = 0,
 \end{aligned}$$

which proves (2.10).

The above theorem shows that, *whenever the material under observation is not quite homogeneous so that the compound Poisson law applies instead of the simple one, there will be too many cases with “no event” and, as compared with these cases, too few with “one event”.* It should be noticed, however, that it is *not* strictly true that *always*

$$(2.13) \quad \pi_1 < \pi(1, m).$$

As a matter of fact, even in the numerical example given by Neyman, the computed value  $\pi_1$  exactly equals the observed value. Still, the inequality (2.13) will hold whenever the third moment about the mean of  $U(a)$  is smaller than twice the second. Writing

$$\begin{aligned}
 \sigma^2 &= \int_0^\infty (a - m)^2 dU(a), \\
 (2.14) \quad M &= \int_0^\infty (a - m)^3 dU(a),
 \end{aligned}$$

and using two more terms in the Taylor development of  $e^{m-a}$  than in (2.11) and (2.12) we see that

$$(2.15) \quad \pi_0 \geq e^{-m} \left\{ 1 + \frac{\sigma^2}{2} - \frac{1}{6} M \right\}$$

and

$$(2.16) \quad m\pi_0 - \pi_1 \geq e^{-m} \left\{ \sigma^2 - \frac{1}{2} M \right\}.$$

These inequalities are slightly sharper than (2.9) and (2.10), and often permit us to estimate the variance of  $U(a)$ .

We note furthermore that the *variance of the compound Poisson distribution is*

$$(2.17) \quad -\sigma^2 + m$$

as compared with the variance  $m$  of the corresponding simple Poisson distribution. Finally the following important property of the compound distribution may be mentioned: Consider two independent variates  $X$  and  $Y$  distributed according to two compound Poisson distributions  $\{\pi_n^{(1)}\}$  and  $\{\pi_n^{(2)}\}$  associated with the c.d.f.'s  $U_1(a)$  and  $U_2(a)$ , respectively. Then the variate  $X + Y$  is distributed according to a compound Poisson law  $\{\pi_n\}$  associated with the c.d.f.  $U(a) = U_1(a) * U_2(a)$  (cf. (1.4)).

It suffices to note that  $U_i(a) \equiv 0$  for  $a < 0$ , so that

$$U(a) = \int_0^a U_1(a-s) dU_2(s);$$

therefore, after a permitted change of the order of integration

$$\begin{aligned} \pi_n &= \int_0^\infty e^{-a} \frac{a^n}{n!} dU(a) \\ &= \int_0^\infty dU_2(s) \int_s^\infty e^{-a} \frac{a^n}{n!} dU_1(a-s) \\ &= \int_0^\infty dU_2(s) \int_0^\infty e^{-(s+t)} \frac{(s+t)^n}{n!} dU_1(t) \\ &= \sum_{k=0}^n \frac{1}{k!} \frac{1}{(n-k)!} \pi_k^{(1)} \pi_{n-k}^{(2)}; \end{aligned}$$

the last expression represents the convolution of  $\{\pi_n^{(1)}\}$  and  $\{\pi_n^{(2)}\}$ .

Neyman's distributions of type  $A$  with two parameters are special cases of a compound Poisson process where  $U(a)$  is a step function with jumps at equidistant places, the jumps being given by a simple Poisson distribution  $\{\pi(n; \lambda)\}$ . Now the convolution of two such distributions is again a simple Poisson distribution  $\{\pi(n; 2\lambda)\}$  with jumps at the same places; hence the convolution of two distributions of type  $A$  is again a similar distribution with one parameter doubled.

As mentioned before, the notion of a compound Poisson distribution is due to Greenwood and Yule [6]. The time dependent compound Poisson process has been the object of detailed investigations by J. Dubourdieu [2] and O. Lundberg [9]. The latter has discussed also the problem of fitting the compound Poisson process to empirical distributions.

**3. The generalized Poisson distribution.** Let  $F(x)$  be an arbitrary c.d.f. Then its  $n$ -fold convolution  $F^{n*}(x)$  (cf. (1.5)) may be considered as a c.d.f. depending on a parameter  $n$ . Choosing, for the latter, the simple Poisson distribution (2.1) and performing the operation indicated in (1.1), we arrive at the c.d.f. of the generalized Poisson law

$$(3.1) \quad G(x) = \sum_{n=0}^{\infty} e^{-a} \frac{a^n}{n!} F^{n*}(x).$$

If, in particular,  $F(x)$  is the unitary function (1.6), we have the ordinary Poisson law

$$(3.2) \quad \Pi(x) = \sum_{n=0}^{\infty} e^{-a} \frac{a^n}{n!} E^{n*}(x) = \sum_{n=0}^{\lfloor x \rfloor} e^{-a} \frac{a^n}{n!}$$

in its cumulative form.

The most frequently encountered application of the generalized Poisson distribution is to problems of the following type. Consider independent random events for which the simple Poisson distribution may be assumed, such as: telephone calls, the occurrence of claims in an insurance company, fire accidents, sickness, and the like. With each event there may be associated a random variable  $X$ . Thus, in the above examples,  $X$  may represent the length of the ensuing conversation, the sum under risk, the damage, the cost (or length) of hospitalization, respectively. To mention an interesting example of a different type, A. Einstein Jr. [5] and G. Polya [13, 14] have studied a problem arising out of engineering practice connected with the building of dams, where the events consist of the motions of a stone at the bottom of a river; the variable  $X$  is the distance through which the stone moves down the river.

Now, if  $F(x)$  is the c.d.f. of the variable  $X$  associated with a single event, then  $F^{n*}(x)$  is the c.d.f. of the accumulated variable associated with  $n$  events. Hence (3.1) is the probability law of the sum of the variables (sum of the conversation times, total sum paid by the company, total damage, total distance travelled by the stone, etc.).

In view of the above examples, it is not surprising that the law (3.1), or special cases of it, have been discovered, by various means and sometimes under disguised forms, by many authors. Quite recently Satterthwaite [16] was led to it (in the above simple form) from problems in insurance. Related (but less elegant) considerations may be found in a paper by W. G. Ackermann [1]. Simple as they are, the above considerations leading to (3.1) furnish a complete solution of the problem in all the cases mentioned. Unfortunately, the special features of the problems often so overshadow the essential point, that one is often led to unnecessarily complicated and incomplete solutions. As an example of the difficulties in considering special cases we mention that Polya [13, 14] was led to a partial differential equation of the hyperbolic type, which conceals the elementary nature of the problem.

If  $F(x)$  is itself a Poisson c.d.f. (3.1) reduces to (2.7). Thus Neyman's distribution of type A depending on two parameters is both a compound and a generalized Poisson distribution. We shall later on see that the generalized Poisson distribution plays an even more important rôle in Neyman's theory.

The main properties of (3.1) are easily derived using characteristic functions. If  $\varphi(z)$  is the characteristic function of  $F(x)$ , the characteristic function of  $G(x)$  is

$$(3.3) \quad \psi(z) = e^{a(\varphi(z)-1)}.$$

Accordingly the  $r$ -th semi-invariant of  $G(x)$  equals the  $r$ -th moment of  $F(x)$  multiplied by  $a^{-1}$ . Moreover it is readily seen that the  $r$ -th convolution of  $G(x)$  with itself is again a function of type (3.1), only with  $a$  replaced by  $ra$ . Neyman's Proposition II is a special case of this remark.

**4. Neyman's contagious distributions.** As an illustration of the general applicability of the operation (1.1) we shall consider the typical example treated by Neyman. Consider the distribution of larvae in a field. The field is divided into plots of equal areas and we are interested in the probability  $c_k$  that exactly  $k$  larvae are found in a certain plot. Now we assume with Neyman:

(i) The larvae may come from various litters. It is assumed that the probability that exactly  $\nu$  litters are represented on our plot is given by the simple Poisson distribution<sup>4</sup> (2.1). (ii) The probability that there are exactly  $n$  survivors is the same for all litters and will be denoted by  $p(n)$ . (iii) If, in any particular litter, there are exactly  $n$  survivors, the probability that  $k$  of them are found on the plot under observation is given by the binomial distribution. We shall write the latter in its cumulative form

$$(4.1) \quad B(x, n, u) = \sum_{k=0}^n \binom{n}{k} u^k (1-u)^{n-k} E^{k*}(x),$$

(cf. (1.6)). (iv) The parameter  $u$  in (4.1) is characteristic for any particular litter (and varies, in particular, with the position of the litter relative to the particular plot under observation). The c.d.f. of  $u$  (which characterizes the distribution of litters in the field) is supposed to be known and will be denoted by  $F(u)$ . The litters are statistically independent.

Now for any particular litter the probability that at most  $k$  survivors will be in the plot under observation is given by

$$(4.2) \quad L(k, u) = \sum_{n=0}^{\infty} p(n) B(k, n, u),$$

which is a special case of (1.2). Here  $u$  is the parameter for the litter picked out. Accordingly, the probability that at most  $k$  survivors from any one litter will be found on our plot is

$$(4.3) \quad L(k) = \int_0^1 L(k, u) dF(u),$$

and this is the second application of the operation (1.1). Since any number of litters may be represented on our plot, the final expression for the probability

---

<sup>4</sup> Actually Neyman at first assumes the number of litters in the field to be finite and considers therefore the binomial instead of the Poisson distribution. Later, however, a passage to the limit is performed which is equivalent to the above assumption. It will be seen that in the following consideration the Poisson distribution may be replaced by any other distribution.



that at most  $k$  larvae will be found on our plot is obtained in the form of a *generalized Poisson c.d.f.*

$$(4.4) \quad C(k) = \sum_{n=0}^{\infty} e^{-a} \frac{a^n}{n!} L^{n*}(k).$$

This is the desired c.d.f. For the desired probability  $c_k$  we have  $c_k = C(k) - C(k - 1)$ .

We specialize now with Neyman the assumption (ii) to the effect that the distribution function  $\{p(n)\}$  is a Poisson distribution

$$(4.5) \quad p(n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

The distribution (4.2) then becomes the c.d.f. of a generalized Poisson distribution, since  $B(x, n, u) = B^{n*}(x, 1, u)$ .

The simplest special case arises when all litters are characterized by the same value of the parameter, say  $u = u_0$ . Then  $F(u) = E\left(\frac{x}{u_0}\right)$ , and  $L(k) = L(k, u_0)$ .

Writing  $L'(k) = L(k) - L(k - 1)$  for the probability that exactly  $k$  survivors from any one litter will be found on our plot, we have

$$(4.6) \quad \begin{aligned} L'(k) &= \sum_{n=k}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} u_0^k (1 - u_0)^{n-k} \\ &= e^{-\lambda u_0} \frac{(\lambda u_0)^k}{k!}. \end{aligned}$$

The c.d.f. (4.4) then reduces to the form (2.7). Similarly, when  $F(u)$  is a step function we arrive at Neyman's more parametric distributions of type  $A$ .

If  $F(u) = u$  for  $0 < u < 1$  (rectangular distribution), then  $\int_0^1 B(k, n, u) dF(u)$  has only jumps of magnitude  $1/(n + 1)$ , and

$$(4.7) \quad L'(k) = \sum_{n=k}^{\infty} \frac{e^{-\lambda} \lambda^n}{(n + 1)!}.$$

This leads to Neyman's function of type  $B$ . The characteristic function of (4.7) is readily seen to be

$$l(z) = \frac{1}{\lambda} \frac{e^{\lambda(e^{iz} - 1)} - 1}{e^{iz} - 1},$$

so that the characteristic function of the final c.d.f.  $C(k)$  becomes

$$\exp \left\{ a \left( \frac{1}{\lambda} \frac{e^{\lambda(e^{iz} - 1)} - 1}{e^{iz} - 1} - 1 \right) \right\},$$

in agreement with Neyman's formula.

**5. The nature of contagion.** It is well known that the simple Poisson distribution describes mutually independent events; in other words, with a Poisson distribution the numbers of events in two non-overlapping time intervals are uncorrelated and the occurrence of an event has no influence on the probability of occurrence of further events. Accordingly, the compound Poisson process also applies to independent and *not* contagious events. With really contagious events (as, for example, with epidemics) the occurrence of each event increases (or decreases) the probability of further events. Greenwood and Yule [6] developed a very general scheme for such events but, due to the very generality, their formulas became too complex for practical applications. They considered the compound Poisson process, and, in particular, the Polya distribution (2.5), as an alternative hypothesis. Accordingly, they interpreted the good fit of that distribution to accident statistics as indicating that there was *no* contagion but that proneness to accidents varies with the person.

Considering a very similar problem, Polya and Eggenberger were later on led to consider a special model of true contagion. This turns out to be the simplest case of the general Greenwood-Yule scheme, but this had been overlooked by them. Curiously enough, Polya was led exactly to the distribution (2.5) which Greenwood and Yule found as an alternative to contagion. It is therefore seen that, contrary to a wide-spread opinion, *an excellent fit of Polya's distribution to observations is not necessarily indicative of any phenomenon of contagion in the mechanism behind the observed distribution.* In order to decide whether or not there is contagion, it is not sufficient to consider the distribution of events, but a detailed study of the correlation between various time intervals is necessary.<sup>5</sup>

The double interpretation of Polya's distribution leads to an understanding of the compound Poisson distribution. To the observer *the compound Poisson distribution will always appear "contagious"; however, this contagion is not inherent in any phenomenon in nature, but simply in our method of sampling.* As a matter of fact, with a compound Poisson distribution the parameter  $a$  is a random variable.<sup>6</sup> Its a priori c.d.f. in the total population is  $\text{Prob} \{a \leq x\} = U(x)$ . Now if, for any particular sample, the observed number of events is  $n$ , then the a posteriori c.d.f. of  $a$  in that sample is given by

$$\text{Prob} \{a \leq x\} = \frac{\int_0^x e^{-s} \frac{s^n}{n!} dU(s)}{\int_0^\infty e^{-s} \frac{s^n}{n!} dU(s)}.$$

<sup>5</sup> For such studies cf. Newbold [10] and Lundberg [9]. For some generalizations of the Polya-Eggenberger scheme see Kitagawa [17] and Rosenblatt [15].

<sup>6</sup> It will be noticed that here  $a$  is actually a random characteristic in the population and can be sampled. We are therefore not guilty of the absurdity which is usually connected with the unfortunate use of Bayes' theorem, when a constant is regarded as random variable. If the output of a machine is distributed according to a Poisson distribution, its parameter is a constant, characteristic of that machine. Regarding it as a random variable means to consider the collective of non-existing similar machines and making predictions for them, whereas we are interested in the one machine only.

This is additional information enabling us to make better predictions for the future or estimates of other properties of the sample. For example, if  $n$  is very large, there is a considerable probability that the mean of  $a$  in the sample exceeds that of the total population: accordingly, we shall expect that also in the future the number of events in our sample will be comparatively large. In other words, although the events themselves are strictly independent we have an apparent contagion due to our method of observation.

It is hardly necessary to point out that the contagion studied by Neyman is of the type just described. Any inhomogeneity of a population of type (1.1) will lead to such an apparent contagion. However, that the Polya-Eggenberger distribution is a member of our class of contagious distributions must be regarded as accident and due to the possibility of its being interpreted as a compound Poisson distribution.

REFERENCES

- [1] W. G. ACKERMANN, "Eine Erweiterung des Poissonschen Grenzwertsatzes und ihre Anwendung auf die Risikoprobleme der Sachversicherung," *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik*, Univ. Berlin, Vol. 4 (1939), pp. 211-255.
- [2] J. DUBOURDIEU, "Les fonctions absolument monotones et la théorie mathématique de l'assurance-accidents," *Comptes Rendus de l'Acad. Sc.*, Paris, Vol. 206, (1938), pp. 303-305, 556-557.
- [3] F. EGGENBERGER, "Die Wahrscheinlichkeitsansteckung," *Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker*, 1924, pp. 31-144.
- [4] F. EGGENBERGER and G. POLYA, "Über die Statistik verketteter Vorgänge," *Zeitschrift für Angewandte Mathematik und Mechanik*, Vol. 1, (1923), pp. 279-289.
- [5] A. EINSTEIN, JR., "Der Geschiebetrieb als Wahrscheinlichkeitsproblem," *Mitteilungen der Versuchsanstalt für Wasserbau an der Eidgenössischen Technischen Hochschule*, Zürich, 1937, pp. 3-112.
- [6] M. GREENWOOD and G. UDN YULE, "An inquiry into the nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents," *J. Roy. Stat. Soc.*, Vol. 83 (1920), pp. 255-279.
- [7] T. KITAGAWA, "The limit theorems of the stochastic contagious processes," *Mem. Faculty of Sc.*, Kyūsyū Imperial University, A, Vol. 1, (1941), pp. 167-194.
- [8] T. KITAGAWA and S. HURUYA, "The application of the limit theorems of the contagious stochastic processes to the contagious diseases," *Mem. Faculty of Sc.*, Kyūsyū Imperial University A, Vol. 1 (1941), pp. 195-207.
- [9] O. LUNDBERG, *On Random Processes and their Application to Sickness and Accident Statistics*, Thesis, University of Stockholm, 1940.
- [10] E. NEWBOLD, "Practical applications of the statistics of repeated events, particularly to industrial accidents," *J. Roy. Stat. Soc.*, Vol. 90 (1927), pp. 487-547.
- [11] J. NEYMAN, "On a new class of 'contagious' distributions, applicable in entomology and bacteriology," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 35-57.
- [12] C. PALM, "Inhomogeneous telephone traffic in full-availability groups", *Ericsson Technics*, 1937, no. 1. (Stockholm). pp. 1-36.

---

<sup>7</sup> A book by J. Dubourdieu, *Théorie de l'assurance-maladie*, Paris, 1939, has been announced, but was not available to the present writer. It presumably treats the compound Poisson distribution more fully than the short notes quoted above.

- [13] G. POLYA, "Zur Kinematik der Geschiebebewegung," *Mitteilungen der Versuchsanstalt für Wasserbau an der Eidgenössischen Technischen Hochschule*, Zürich, 1937.
- [14] G. POLYA, "Sur la promenade au hasard dans un réseau de rues. *Actualités Scientifiques et Industrielles*, No. 734, (1938), pp. 25-44.
- [15] A. ROSENBLATT, "Sur le concept de contagion de M. G. Pólya dans le calcul des probabilités. Applications à la peste bubonique au Pérou," *Actas Academ. Ciencias Lima*, Vol. 3 (1940), pp. 186-204.
- [16] F. E. SATTERTHWAITTE, "Generalized Poisson distribution," *Annals of Math. Stat.* Vol. 13 (1942), pp. 410-417.