

NUMERICAL STUDIES ON SOME
CONTAGIOUS DISTRIBUTIONS

BY C. C. SPICER

Central Public Health Laboratory, Colindale Avenue, London, N.W. 9

AND S. LIPTON

Rothamsted Experimental Station, Harpenden, Herts

(With 3 figures in the Text)

INTRODUCTION

Recently Bailey (1957) and Kendall (1956) have discussed the behaviour of certain distributions of epidemic size, especially in relation to the threshold theorem of Kermack & McKendrick (1927). The latter was put forward on the basis of a deterministic theory of the epidemic process, and suggested that the introduction of a small number of infected persons into a community would only result in a major epidemic if the density of population was above a certain threshold value. Numerical investigations by Bailey (1957) and Kendall (1956) suggested that a rough parallel with this existed for stochastic analogues of the determinist epidemic, and it seemed worth while to make a more detailed investigation of this question for stochastic epidemics using an electronic computer for the calculations required. Other reasons for carrying out such a study were that the general properties of the distributions of epidemic size for moderately large communities are not well understood, and are of some practical interest in connexion with epidemics in such units as nurseries, or school classes. Direct application to practical examples is not likely to be possible owing to the great heterogeneity that usually exists between communities, and the variations in susceptibility among their members. On the other hand, the broad picture may not be too badly disturbed by these factors and can be taken as a background against which observed results can be viewed.

Another use to which tables of epidemic size distribution might be put is in assessing the effectiveness of preventive or prophylactic measures. For example, comparisons of various policies of exclusion or treatment of symptomless carriers of infection in relation to the course of an epidemic in a small school or nursery will usually be based on highly skew or U-shaped distributions. This may vitiate the tests that are used in comparing effects on groups of individuals where no contagion exists.

A further reason for writing computer programmes is that once this has been done they are available for any other investigation that requires them. The computations are extremely laborious and time-consuming when performed on desk calculators, particularly for large community sizes.

METHODS AND SCOPE OF THE INQUIRY

The models of epidemic processes chosen for study were those due to Bailey (1953), Frost [see Bailey (1957), Chap. 6] and Greenwood (1931). Another class of general contagious distributions is that due to Neyman (1939), and there is also the well-known Polya–Eggenberger distribution (see Feller, 1950), a particular case of which was derived in relation to epidemic data by Irwin and discussed by Greenwood (Greenwood, 1949; Irwin, 1954). We have not examined the Neyman distributions but some consideration is given to the Polya–Eggenberger distribution in the discussion.

The programming of the three distributions for the computer has been based on the recurrence relations that exist between the probabilities of epidemics of different sizes in communities with different numbers of susceptibles. Contagious distributions of the kind considered can always be considered as being built up of a series of sub-epidemics as the number of susceptibles is successively reduced by infection, and the final epidemic size is the outcome of these sub-epidemics.

It is worth while giving a brief account of the distributions studied and their recurrence relations.

(a) *Continuous infection-removal model.* This model assumes that the infectiveness of a case is a continuous process, and that the probability per unit time of infecting an exposed susceptible is constant. At the same time the case has a constant probability per unit time of ceasing to be infectious. The ratio of the removal to the infection rate, ρ , is the only parameter of the distribution. In terms of this constant the probability $P(n, w)$ of an epidemic of size w in a community of size n , after the introduction of one infectious case, can be derived from the recurrence relation

$$f_{r, s} = \frac{\rho(s+1)}{s(r+\rho)} f_{r, s+1} + \frac{(r+1)(s-1)}{s(r+\rho)} f_{r+1, s-1},$$

where

$$P(n, w) = \rho f_{n-w, 1}.$$

The boundary conditions are such that

$$f_{n, 1} = \frac{1}{n+\rho}, \quad 1 \leq r+s \leq n+1, \quad 0 \leq r \leq 1, \quad 1 \leq s \leq n+1.$$

Full details can be found in Bailey’s book (Bailey, 1957).

(b) *Frost chain binomial.* This distribution assumes that the infectious process is a point event and that each case has a probability p of infecting any of the susceptibles exposed to it at the end of the incubation period. If r cases occur simultaneously then the probability that any susceptible exposed to them is infected is $\{1 - (1-p)^r\}$. The recurrence relation for this distribution is

$$P(n, r) = \binom{n}{r} (1-p)^{(r+1)(n-r)} G_r,$$

where

$$G_r = 1 - \sum_{s=0}^{r-1} P(r, s)$$

and

$$P(n, 0) = (1-p)^n.$$

This form of the recurrence was given by Lidwell & Sommerville (1951).

(c) *Greenwood's chain binomial*. This is a slight modification of the Frost model and assumes that when several cases become infectious simultaneously they are no more likely to infect an exposed susceptible than a single case. The main reason for examining this distribution was to see what effect this change of assumptions had in differentiating the behaviour of the Frost and Greenwood models. The recurrence relation used for calculating the probabilities was

$$P(n, r) = \sum_{k=0}^{r-1} B(n, r-k) P(n-r+k, k),$$

where

$$B(n, m) = \binom{n}{m} p^m (1-p)^{n-m}$$

the usual Bernoullian probability term.

It will be noticed that this procedure involves having available the relevant binomial probabilities. A very fast programme for these is possible using the basic recurrence relation of the binomial distribution.

The electronic computer used for this investigation was the Elliott-N.R.D.C. 401 computer at Rothamsted Experimental Station (for a description of this machine see Lipton (1955)). In order to keep the running-time of the programmes within reasonable bounds it was decided to impose an upper limit on the community size examined. In this type of work the computing time increases very rapidly as the community size increases. For the Frost and Greenwood models the highest value for the number exposed to risk was fixed at 40, while for the Bailey model only integral values of ρ greater than 1 were permissible and n was never allowed above 40. Speed of computation was the main consideration in programming the three models and for this reason several techniques, such as floating arithmetic, which would have eased the programming work, were not adopted. A further limitation, though not one which seriously affected the present study, is imposed by the storage space available in the computer. The storage required is proportional to the square of the community size and so, like the running-time, increases rapidly as larger communities are examined. Although it would be simple to alter our programmes to handle a larger community size, say up to 60, it would be very uneconomical in machine time. In any event the present investigation was intended to study the general nature of the epidemic size distribution rather than to produce specific numerical values for some particular set of observations. We believe the upper bounds chosen are sufficient for this purpose.

Some examples of various times (in minutes) taken by the computer are:

Community size/model	Bailey	Greenwood	Frost
5	1 ($\rho = 5$)	1	1
10	4 ($\rho = 10$)	4	4.5
20	15 ($\rho = 20$)	6	6

The comparatively short times for the Greenwood model are due mainly to a modification of the computer itself which was made after work on the other two

models had been completed. All these times include the printing of results, which in the case of the Greenwood and Frost models was over 50% of the total running time.

The programmes were checked by verifying that the probabilities of the terms of the distributions added up to unity and, in some cases, by independent calculations on a desk machine.

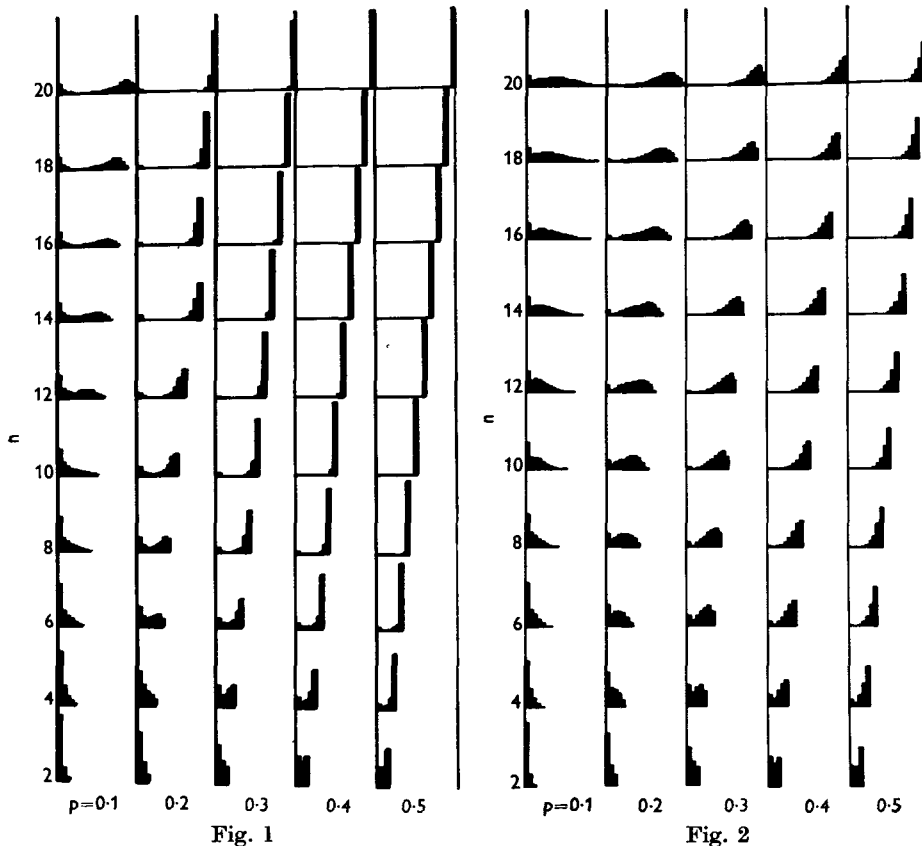


Fig. 1. Distribution of epidemic size according to Frost's mathematical model in communities of 2-20 susceptibles (n), and various degrees of infectiousness (p).

Fig. 2. Distribution of epidemic size according to Greenwood's model. Same values of parameters as for Fig. 1.

RESULTS

The results are summarized graphically in Figs. 1, 2 and 3. The numerical values on which these graphs are based can be given to those who may find them useful, but the general behaviour is best demonstrated pictorially.

Fig. 1 shows the behaviour of the Frost distribution for values of p (the chance of a single infection of one exposed by one primary case) from 0.1 to 0.5 and for communities of 2-20 susceptibles. When p is 0.1 or less there is a slow transition, as the size of community is increased, from a J-shaped distribution with a maximum at zero to one J-shaped in the opposite sense. The rapidity of the transition increases sharply with increase in p , and for $p > 0.5$ the distribution is

always negatively skewed and, as the size of community increases, becomes concentrated in a single ordinate corresponding to infection of all the exposed susceptibles.

The behaviour of the Greenwood distribution is shown in Fig. 2 for the same values of n and p . This is mathematically related to the Frost binomial but the infectious process is much less intense and the transition to virtual certainty of a 'pandemic' with increasing numbers of susceptibles is much slower.

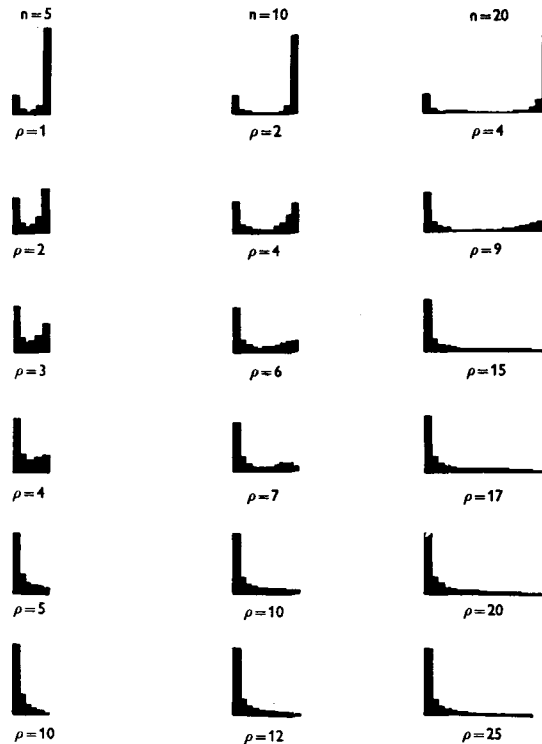


Fig. 3. Distribution of epidemic size for Bailey's continuous stochastic model for 5, 10 and 20 susceptibles (n), and various values of ρ , the ratio of removal, to infection rate in an affected case.

A few representative examples of Bailey's distribution are given in Fig. 3. They are set out to show how the form of the distribution alters with increasing values of ρ , the ratio of the removal, to the infection rate. It was suggested as a working rule by D. G. Kendall (1956) that when ρ was less than n the distribution was U-shaped, while at about the point where $\rho = n$ there was a transition from a U- to a J-shape, most epidemics petering out without affecting more than a fraction of the community. This rule is fairly accurate for small values of n , as is illustrated in Fig. 3 by the case $n = 5$. As n becomes larger the transition is much less abrupt and occurs at values of ρ less than n . This can be seen in Fig. 3 for $n = 20$ where the distribution actually becomes J-shaped at about $\rho = 17$.* Furthermore, as was noted by Bailey, the distribution goes through a phase in which the right-hand mode moves to the left and ceases to correspond to a 'pandemic'.

* This is not quite apparent from Fig. 3 due to the small size of the diagrams.

DISCUSSION

The most obvious conclusion to be drawn from the present study is that, where these mathematical models are applicable, any moderately infectious disease is likely to spread to almost all susceptible members in communities of 10 or more.

By a moderately infectious disease is meant one in which the probability of an initial case infecting a single exposed susceptible is more than about 0.1. This is the parameter p of the Frost and Greenwood distributions, and in Bailey's model is found by calculating

$$1/(\rho + 1) = p.$$

The general parallelism between the two discrete processes represented by the Greenwood and Frost distributions and the continuous case studied by Bailey can be seen from the figures. The threshold for the Bailey distribution occurs roughly at $\rho = n$, when there is a transition from a J-shaped to a U-shaped distribution as n increases. If ρ and p are connected by the relation $\rho = (1 - p)/p$ then, for a given p , we should expect a transition to occur in the Greenwood and Bailey distributions at about the point where $n = (1 - p)/p$. In fact, this transition takes place at rather higher values of n in the range of p considered, though a rough analogy with Kendall's rule exists.

In those diseases for which reliable estimates of p exist its value is so high as to ensure that moderate-sized communities of susceptibles will all be infected. There is also evidence that some diseases, for example, poliomyelitis and Sonne dysentery, are widely spread in nursery or school outbreaks, where mixing is considerable and there are ample opportunities for transmitting infection. If any of the mathematical models considered here are approximately applicable to these diseases this effect of aggregation would often appear in communities of 10–15 or more, since their values of ρ are very small.

The range of applicability of the results given above is difficult to assess, but their main features are probably not badly distorted by departures from the basic assumptions. The most likely source of disagreement with observation would be the presence of wide variations in susceptibility in the community. This would manifest itself particularly when an appreciable number of those exposed were not susceptible, so that the infective contacts of a case were greatly reduced.

The distributions studied differ fundamentally from the contagious distributions of Neyman's types, and from the Polya–Eggenberger distribution. These two distributions have no behaviour analogous to a threshold, and in the case of the Polya–Eggenberger distribution it appears that the main effect of increasing the size of the population at risk is to cut down rather than to increase the proportion of the community affected. However, for the less extreme forms of the epidemic distributions, we have found empirically that the Polya–Eggenberger distribution often gives a good fit and could not be distinguished by any ordinary set of observational data.

SUMMARY

A numerical investigation of three contagious distributions characteristic of epidemic data has been made using an electronic computer.

All three show a modified threshold effect such that the proportion of a community affected tends to 100% as its size increases though there is no abrupt change of form.

We would like to thank Mr M. J. R. Healy for assistance with one of the programmes, and Miss Molly Fretton for her help with the computing.

REFERENCES

- BAILEY, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika*, **40**, 177.
- BAILEY, N. T. J. (1957). *The Mathematical Theory of Epidemics*. London: Griffin and Co.
- FELLER, W. (1950). *Probability Theory and its Applications*. New York: John Wiley and Sons.
- GREENWOOD, M. (1931). On the statistical measure of infectiousness. *J. Hyg., Camb.*, **31**, 336.
- GREENWOOD, M. (1949). The infectiousness of measles. *Biometrika*, **36**, 1.
- IRWIN, J. O. (1954). A distribution arising in the study of infectious diseases. *Biometrika*, **41**, 266.
- KENDALL, D. G. (1956). Deterministic and stochastic epidemics in closed populations. *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 149. Berkeley and Los Angeles, University of California Press.
- KERMACK, W. O. & MCKENDRICK, A. G. (1927). Contributions to the mathematical theory of epidemics. *Proc. Roy. Soc. A*, **115**, 700.
- LIDWELL, O. M. & SOMMERVILLE, T. (1951). Observations on the incidence and distribution of the common cold in a rural community during 1948 and 1949. *J. Hyg., Camb.*, **49**, 365.
- LIPTON, S. (1955). A note on the electronic computer at Rothamsted. *Math. Tab., Wash.*, **9**, 69.
- NEYMAN, J. (1939). On a new class of contagious distributions applicable in entomology and bacteriology. *Ann. Math. Statist.* **10**, 35.

(*MS. received for publication 25. VII. 58*)