A SIMPLE STOCHASTIC EPIDEMIC

BY NORMAN T. J. BAILEY

Department of Medicine, University of Cambridge

CONTENTS

1. GENERAL INTRODUCTION

The mathematical theory of epidemics has usually been confined to the consideration of 'deterministic' models as, for example, in the work of Kermack & McKendrick (1927 and later) and Soper (1929). That is, it has been assumed that, for given numbers of susceptible and infectious individuals and given infection and removal rates, a certain definite number of fresh cases would arise in a given time. In fact, as is well known, a considerable degree of chance enters into the conditions under which fresh infections take place, and it is clear that for a more precise analysis we ought to take these statistical fluctuations into account. In short, we require 'stochastic' models to supplement existing deterministic ones.

Bartlett (1949) has emphasized this need and has devoted some discussion to various partial attacks already made on the problem. A brief reference has also been made by Bartlett (1946, pp. 52–3) to the simple stochastic epidemic process considered in greater detail in the present paper.

In deterministic treatments the total number of cases is a single-valued function of time, but in stochastic treatments the single point on the deterministic curve is replaced by a probability distribution. The usual deterministic *epidemic curve* gives the rate of change with respect to time of the total number of cases (regarded as continuous), while the most appropriate stochastic analogue is probably the curve of the rate of change with respect to time of the stochastic mean. The latter statement needs to be suitably qualified and is discussed in greater detail in the following section. In some processes stochastic means are identical with deterministic values, but this is not the case in epidemic processes. It is worth remarking in passing that the rather unexpected smoothness of *observed* epidemic curves is most likely to be due to the partial ironing out of statistical variations, by summation over finite periods of time (e.g. when quoting so many new cases per day or week) and by summation over relatively isolated epidemics occurring simultaneously in subgroups of the main population.

The present note deals with the very simplest case of the spread of a relatively mild infection, in which none of the infected individuals is removed from circulation by death, recovery or isolation. This is admittedly an over-simplification, but, apart from providing a possible basis for more extensive investigations, it may well represent the situation with, for example, some of the milder infections of the upper respiratory tract.

2. The importance of stochastic means in epidemics

It is usual to assume that the probability of a new case occurring in a small interval of time is proportional to both the number of susceptible and the number of infectious individuals. These assumptions are reasonable for areas small enough for homogeneous mixing to take place. This is clearly not so with large areas, for which it is well known that the overall epidemic can often be broken down into smaller epidemics occurring in separate regional subdivisions. These regional epidemics are not necessarily in phase and often interact with each other. Taking the process of subdivision a stage further we can consider a single town or district. Even here it is obvious that a given infectious individual has not the same chance of infecting each inhabitant. He will probably be in close contact with a small number of people only, perhaps of the order of 10–50, depending on the nature of his activities. The observed epidemic for the whole district will then be built up from epidemics taking place in several relatively small groups of associates and acquaintances. Although in practice the groups may overlap, we can employ the concept of an *effective* number of independent groups, and it may be possible to assume, as a first approximation, that they are equal in size. A typical model would involve a community of, say k independent groups each of size n. We can imagine an epidemic started by the simultaneous appearance or introduction of k primary cases, one for each group.

Stochastic means are often not very informative because of the large amount of variation associated with them. But in the model suggested, at any given time, the coefficient of variation of the total number of cases will be $1/\sqrt{k}$ times the coefficient of variation of any one of the groups. Thus the larger the value of k, the more nearly will the curve of the total number of cases approach in shape the curve of the stochastic mean for a population of size n; and we may expect the overall *epidemic curve* to approach in shape the curve of the rate of change with respect to time of the stochastic mean. In epidemic processes stochastic means are not the same as the deterministic values, so that special treatment is required.

Although the above model is rather over-simplified there is some justification for regarding the epidemic curve derived from the stochastic mean as the appropriate stochastic analogue corresponding to the classical deterministic epidemic curve.

3. Deterministic treatment of a simple epidemic

Let us consider a community containing n susceptibles into which a single infectious individual is introduced. We shall assume that the infection spreads by contact between the members of the community, and that it is not sufficiently serious for cases to be withdrawn from circulation by isolation or death; also that no case becomes clear of infection during the course of the main part of the epidemic. These are wide assumptions but, as already suggested, are probably nearly fulfilled with some of the milder infections of the upper respiratory tract.

Let y be the number of susceptibles at time t, and let β be the infection rate. Then the number of new infections in time dt is $\beta y(n-y+1)\,dt$. If we replace t by βt we shall find that our equation is dimensionless so far as the infection rate is concerned. It is easy to see that the deterministic differential equation is

$$\frac{dy}{dt} = -y(n-y+1), \tag{1}$$

satisfying the initial condition $\quad y = n \quad \text{when} \quad t = 0. \tag{2}$

Equation (1) above corresponds to equation (22) given by Bartlett (1946, p. 53). Bartlett's main variable is the number of infectious individuals, whereas ours is the number of susceptibles, and he presumably starts his process with one infectious individual and $n-1$ susceptibles.

The solution of (1) and (2) is $y = n(n+1)/\{n+e^{(n+1)\lambda t}\}$. (3)

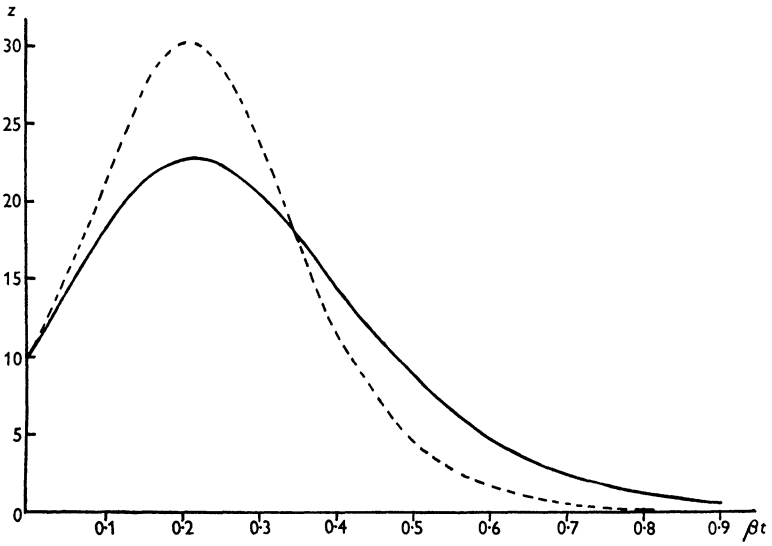

Fig. 1. Comparison of deterministic and stochastic *epidemic* curves for $n = 10$.
– – – – – deterministic curve; ——— stochastic curve.

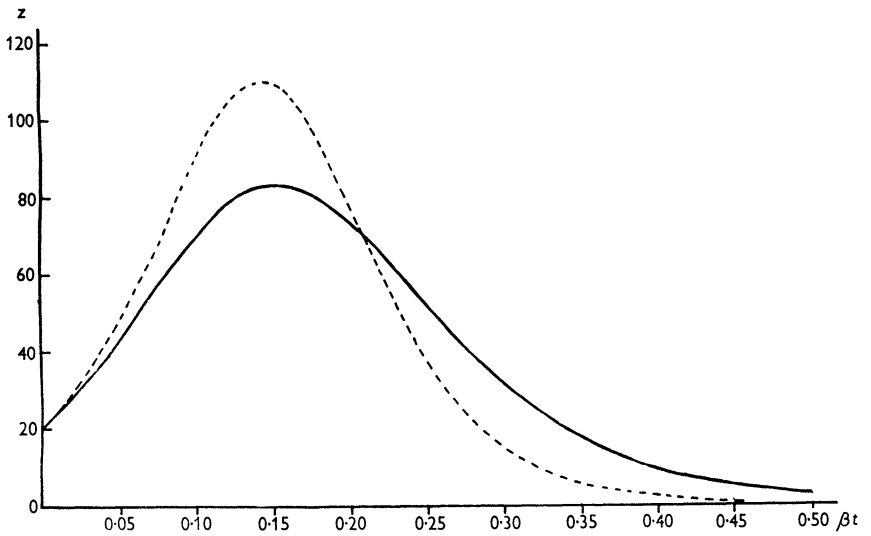

Fig. 2. Comparison of deterministic and stochastic *epidemic* curves for $n = 20$.
– – – – – deterministic curve; ——— stochastic curve.

Thus the deterministic epidemic curve is

$$z = -\frac{dy}{dt} = y(n-y+1) = \frac{n(n+1)^2 e^{(n+1)\lambda t}}{\{n+e^{(n+1)\lambda t}\}^2}. \tag{4}$$

The curve reaches a maximum when $t = \log n/(n+1)$, $y = \frac{1}{2}(n+1)$ and $z = \frac{1}{4}(n+1)^2$. It is clearly symmetrical about $t = \log n/(n+1)$.

The epidemic curve given by (4) is plotted for $n = 10$ and $n = 20$ in Figs. 1 and 2 respectively, where the corresponding curves for the stochastic cases are given for comparison.

Our solution given in (4) does not agree, making due allowance for the change of notation and variable, with Bartlett's solution (23). However, as the latter does not seem to satisfy the apparent initial conditions, there must be a misprint.

4. Stochastic treatment of a simple epidemic

(*a*) *Solution of stochastic differential-difference equations*

Let us use the same notation as in § 3. Then replacing t by βt as before, it is easy to see that on the assumption of homogeneous mixing the probability of one new infection taking place in the interval dt is $y(n-y+1)\,dt$. Now suppose that $p_r(t)$ is the probability that there are r susceptibles still uninfected at time t. Then the usual treatment shows that the epidemic process is characterized by the stochastic differential-difference equations:

$$\left.\begin{aligned} &\frac{dp_r(t)}{dt} = (r+1)(n-r)\,p_{r+1}(t) - r(n-r+1)\,p_r(t) \quad (r = 0, 1, 2, \ldots, (n-1)), \\ \text{and} \quad &\frac{dp_n(t)}{dt} = -np_n(t). \end{aligned}\right\} \tag{5}$$

I have to thank Mr D. G. Kendall for drawing my attention to a paper by Feller (1939). Feller's equation (19) is substantially the same as our (5), though it was obtained in a different context. Feller gives the solution of sets of equations of this type with generalized coefficients. On the right-hand side of our (5) every coefficient appears twice, leading to terms of the type ate^{-bt} as well as ce^{-dt} in the solution. We can use Feller's solution (21) if we apply the usual limiting procedure to the terms which have the form 0/0, but it is more convenient for our purpose to solve *ab initio* as follows.

Let us use the Laplace transform and its inverse with respect to time given by

$$\left.\begin{aligned} \phi^*(\lambda) &= \int_0^\infty e^{-\lambda t}\phi(t)\,dt \quad (R(\lambda) > 0), \\ \phi(t) &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{\lambda t}\phi^*(\lambda)\,d\lambda, \end{aligned}\right\} \tag{6}$$

where $\int_{c-i\infty}^{c+i\infty} \equiv \lim_{w\to\infty}\int_{c-iw}^{c+iw}$, and c is positive and greater than the abscissae of all the residues.

Taking transforms of the equations in (5) and using the boundary conditions

$$\left.\begin{aligned} p_r(0) &= 1 \quad (r = n) \\ &= 0 \quad (r < n), \end{aligned}\right\} \tag{7}$$

we obtain the recurrence relations

$$\left.\begin{aligned} &q_r = \frac{(r+1)(n-r)}{\lambda + r(n-r+1)}\,q_{r+1} \quad (r = 0, 1, 2, \ldots, (n-1)) \\ \text{and} \quad &q_n = \frac{1}{\lambda+n}, \end{aligned}\right\} \tag{8}$$

where

$$q_r = p_r^* = \int_0^\infty e^{-\lambda t}p_r(t)\,dt.$$

It follows from (8) that q_r is given by

$$q_r = \frac{n!(n-r)!}{r!}\Big/\prod_{j=1}^{n-r+1}\{\lambda + j(n-j+1)\} \quad (0 \leqslant r \leqslant n). \tag{9}$$

It is important to note that if $r \geqslant \frac{1}{2}(n+1)$, then the factors in the denominator are all different, while if $r < \frac{1}{2}(n+1)$ some of them are repeated. In the latter case we can write the denominator as

$$(\lambda+n)\{\lambda+2(n-1)\}\{\lambda+3(n-2)\}\ldots\{\lambda+r(n-r+1)\}^2\{\lambda+(r+1)(n-r)\}^2\ldots \times\{\lambda+(\tfrac{1}{2}n-1)(\tfrac{1}{2}n+2)\}^2\{\lambda+\tfrac{1}{2}n(\tfrac{1}{2}n+1)\}^2, \quad \text{for } n \text{ even}, \tag{10}$$

and

$$(\lambda+n)\{\lambda+2(n-1)\}\{\lambda+3(n-2)\}\ldots\{\lambda+r(n-r+1)\}^2\{\lambda+(r+1)(n-r)\}^2\ldots \times\left\{\lambda+\frac{(n-1)}{2}\frac{(n+3)}{2}\right\}^2\left\{\lambda+\frac{(n+1)^2}{4}\right\}, \quad \text{for } n \text{ odd}. \tag{11}$$

Thus all terms after the $(r-1)$th are squared, unless n is odd, in which case the last term is not squared.

We can now see the general character of the solution. To find p_r we merely express q_r in partial fractions and use the inverse of the Laplace transform. Terms in the denominator like $\{\lambda+r(n-r+1)\}$ and $\{\lambda+r(n-r+1)\}^2$ will give rise to $\exp\{-r(n-r+1)t\}$ and $t\exp\{-r(n-r+1)t\}$ respectively. The coefficients of the latter terms are simply the coefficients of the corresponding terms in the expansion of q_r in partial fractions.

In particular, we have

$$q_0 = (n!)^2/[\lambda(\lambda+n)^2\{\lambda+2(n-1)\}^2\ldots] \tag{12}$$

$$= \frac{1}{\lambda} + \sum_{r=1}\left\{\frac{k_r}{\{\lambda+r(n-r+1)\}^2} + \frac{l_r}{\{\lambda+r(n-r+1)\}}\right\}, \tag{13}$$

where $$k_r = q_0\{\lambda+r(n-r+1)\}^2\Big|_{\lambda=-r(n-r+1)} = -\frac{(n!)^2(n-2r+1)^2}{r!(r-1)!(n-r)!(n-r+1)!}. \tag{14}$$

Now if the probability generating function is

$$\Pi(x,t) = \sum_{r=0}^{n} x^r p_r(t), \tag{15}$$

then it can be seen from (5) that $\Pi(x,t)$ satisfies the partial differential equation

$$\frac{\partial\Pi}{\partial t} = (1-x)\left\{n\frac{\partial\Pi}{\partial x} - x\frac{\partial^2\Pi}{\partial x^2}\right\}, \tag{16}$$

with the boundary condition $$\Pi(x,0) = x^n. \tag{17}$$

The equations for the moment-generating function, $M(\theta,t)$, are derived from (16) and (17) by writing $x = e^\theta$. We find

$$\frac{\partial M}{\partial t} = (e^{-\theta}-1)\left\{(n+1)\frac{\partial M}{\partial\theta} - \frac{\partial^2 M}{\partial\theta^2}\right\}, \tag{18}$$

with the boundary condition $$M(\theta,0) = e^{n\theta}. \tag{19}$$

Our (18) is substantially the same as equation (20) given by Bartlett (1946, p. 53), making proper allowance for the change of notation and variable.

Suppose that the rth moment of the distribution of y is m_r', then we can substitute

$$M(\theta) = 1 + m_1'\theta + m_2'\frac{\theta^2}{2!} + \ldots \tag{20}$$

in (18), and equate coefficients of θ to give the following set of differential equations:

$$\left.\begin{aligned}
\frac{dm_1'}{dt} &= -\{(n+1)m_1' - m_2'\},\\
\frac{dm_2'}{dt} &= +\{(n+1)m_1' - m_2'\} - 2\{(n+1)m_2' - m_3'\},\\
\frac{dm_3'}{dt} &= -\{(n+1)m_1' - m_2'\} + 3\{(n+1)m_2' - m_3'\} - 3\{(n+1)m_3' - m_4'\},\\
&\text{etc.},
\end{aligned}\right\} \tag{21}$$

where the numerical coefficients on the right-hand side can be derived from binomial expansions.

Unfortunately, these equations, while capable of giving the higher moments in terms of m_1' when the latter has been found, e.g.

$$m_2' = (n+1)m_1' + \frac{dm_1'}{dt}, \tag{22}$$

are not so convenient for finding m_1' itself. They can obviously be used to develop a Taylor expansion for m_1' in powers of t by successive differentiation and substitution, since all the moments are known when $t = 0$. We have, in fact,

$$m_1' = n - nt - \frac{n(n-2)}{2!}t^2 - \frac{n(n^2-8n+8)}{3!}t^3 - \ldots. \tag{23}$$

However, the series does not converge rapidly enough to be very useful.

Thus we see that the usual method of equating coefficients of θ in the partial differential equation for the moment-generating function, which often leads to simple differential equations for at least the early moments, fails to be of service in the case of stochastic epidemic processes. In the next subsection we shall consider a different method of approach.

(*b*) *Stochastic mean values*

Let the transform of the probability-generating function be

$$\Pi^*(x, \lambda) = \sum_{r=0}^{n} x^r q_r. \tag{24}$$

Referring to equations (9) to (13), it is clear that we can write

$$\Pi^*(x, \lambda) = \frac{1}{\lambda} + \sum_{r=1} \left\{ \frac{f_r(x)}{\{\lambda + r(n-r+1)\}^2} + \frac{g_r(x)}{\{\lambda + r(n-r+1)\}} \right\}, \tag{25}$$

where $f_r(x)$ and $g_r(x)$ are polynomials in x. Thus the probability-generating function itself is of the form

$$\Pi(x, t) = 1 + \sum_{r=1} \{t f_r(x) + g_r(x)\} e^{-r(n-r+1)t}. \tag{26}$$

Therefore

$$m_1'(t) = \left.\frac{\partial \Pi}{\partial x}\right|_{x=1} = \sum_{r=1} \{t f_r'(1) + g_r'(1)\} e^{-r(n-r+1)t}, \tag{27}$$

where, in the expression on the right-hand side, primes are used to indicate differentiation with respect to x.

Now the transform of (16) shows that $\Pi^*(x, \lambda)$ satisfies the differential equation

$$x(1-x)\frac{\partial^2 \Pi^*}{\partial x^2} - n(1-x)\frac{\partial \Pi^*}{\partial x} + \lambda \Pi^* = x^n. \tag{28}$$

If we substitute (25) in (28), multiply by $\{\lambda+r(n-r+1)\}^2$ and then put $\lambda = -r(n-r+1)$, we obtain

$$x(1-x)f_r''-n(1-x)f_r'-r(n-r+1)f_r = 0. \tag{29}$$

The solution of (29) is

$$f_r(x) = CF\{-r, -n+r-1; -n, x\}, \tag{30}$$

where F is a terminating hypergeometric series and C an arbitrary constant. However, C is evidently the coefficient of $\{\lambda+r(n-r+1)\}^{-2}$ in the partial fraction expansion of q_0, i.e. $C = k_r$, whose value is given in (14). Substituting this value in (30), differentiating with respect to x and then putting $x = 1$ gives

$$\begin{aligned} f_r'(1) &= k_r \left.\frac{dF}{dx}\right|_{x=1} \\ &= -k_r \frac{r(n-r+1)}{n} F\{-r+1, -n+r; -n+1, 1\}. \end{aligned}$$

Therefore

$$f_r'(1) = \frac{n!(n-2r+1)^2}{(n-r)!\,(r-1)!}. \tag{31}$$

Specimen values of these coefficients, occurring in the expression for $m_1'(t)$ given by (27), are

r	$f_r'(1)$
1	$n(n-1)^2$
2	$n(n-1)(n-3)^2/1!$
3	$n(n-1)(n-2)(n-5)^2/2!$
4	$n(n-1)(n-2)(n-3)(n-7)^2/3!$
	etc.

(32)

To find the polynomials $g_r(x)$ we substitute (25) in (28), multiply by $\{\lambda+r(n-r+1)\}^2$, differentiate with respect to λ and then put $\lambda = -r(n-r+1)$. This gives the following equation:

$$x(1-x)g_r''-n(1-x)g_r'-r(n-r+1)g_r = -f_r. \tag{33}$$

From (33) we can derive a series solution for $g_r(x)$ in terms of the known $f_r(x)$. I have unfortunately been unable to find a simple and convenient general expression for $g_r(x)$, although it is easy to show that

$$g_1'(1) = n-n(n-1)\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\frac{1}{n-2}\right). \tag{34}$$

In view of the importance of the stochastic mean it was thought worth while to examine one or two cases in detail. We can calculate the probability-generating function as indicated above and from it derive the formula for the mean. The chief labour is in calculating the coefficients in the partial fraction expansions of expressions like (9), but it can, however, be materially reduced by suitably schematizing the computations.

The mean value has been found for the special cases, $n = 10$ and $n = 20$. The formula for $n = 10$ is given below explicitly, and in both cases the epidemic curves are plotted in Figs. 1 and 2, where they are compared with the corresponding deterministic curves.

The expressions for the stochastic mean and the epidemic curve in the special case, $n = 10$, are

$$\begin{aligned} m_1' = {} & e^{-10t}(810t-234\tfrac{17}{28})+e^{-18t}(4410t-902\tfrac{1}{4})+e^{-24t}(9000t-1247\tfrac{1}{7}) \\ & +e^{-28t}(7560t+126)+e^{-30t}(1260t+2268), \end{aligned} \tag{35}$$

$$\begin{aligned} z = -\frac{dm_1'}{dt} = {} & e^{-10t}(8100t-3156\tfrac{1}{14})+e^{-18t}(79{,}380t-20{,}650\tfrac{1}{2})+e^{-24t}(216{,}000t-38{,}931\tfrac{3}{7}) \\ & +e^{-28t}(211{,}680t-4032)+e^{-30t}(37{,}800t+66{,}780). \end{aligned} \tag{36}$$

It is evident that, for both $n = 10$ and $n = 20$, the stochastic epidemic curve has a somewhat different character from the deterministic curve. The latter is symmetrical about its maximum ordinate, whereas the former is not and falls more slowly than it originally rose. On the other hand, it may be noticed that the time at which the maximum occurs in the stochastic case does not differ very much from the time of the maximum in the deterministic case.

It is perhaps worth mentioning here that Feller's remark (1939, p. 22) about the stochastic mean always being less than the deterministic value is easily seen, from a comparison of our equations (1) and (22), to hold in the present case, provided we apply it to the mean number of infectious individuals (*not* the mean number of susceptibles)—as we should if the correct analogy is to be made with the process considered by Feller. In order to prevent confusion it should be remembered that Figs. 1 and 2 show the *epidemic* curve, i.e. the *rate of change* with respect to time of the mean number of infectious individuals.

(c) *Completion times*

Let us call an epidemic *complete* when all the available susceptibles have been exhausted; otherwise we shall say it is *incomplete*. It can be seen from (26) that $\Pi(x, \infty) = 1$; that is, the simple epidemic under consideration is always completed eventually. For more complicated types of epidemic this is not necessarily so, for the infected individuals may all be removed before the epidemic is complete.

Now $p_0(\tau)$ is the probability that the epidemic is complete at time τ. But since the number of susceptibles is a non-increasing function, $p_0(\tau)$ is also the chance that the epidemic has been completed in the interval from 0 to τ. Thus $p_0(\tau)$ is the distribution function and $dp_0/d\tau$ the frequency function for the completion time τ.

The moment-generating function for the completion time is

$$\begin{aligned} M_\tau(\theta) &= E\, e^{\theta\tau} \\ &= \int_0^\infty \frac{dp_0}{d\tau} e^{\theta\tau}\, d\tau \\ &= \left[p_0 e^{\theta\tau} \right]_0^\infty - \theta \int_0^\infty p_0 e^{\theta\tau}\, d\tau, \quad \text{integrating by parts,} \\ &= -\theta q_0(-\theta), \quad \text{for } \theta < 0, \end{aligned}$$

since $$p_0(0) = 0, \quad p_0(\infty) = 1.$$

Therefore
$$M_\tau(\theta) = -\theta q_0(-\theta). \tag{37}$$

Substituting for q_0 from (12) we obtain

$$\begin{aligned} M_\tau(\theta) &= (n!)^2 \Big/ \prod_{j=1}^{n} \{-\theta + j(n-j+1)\} \\ &= \prod_{j=1}^{n} \left\{ 1 - \frac{\theta}{j(n-j+1)} \right\}^{-1}. \end{aligned} \tag{38}$$

The cumulant-generating function is then given by

$$K_\tau(\theta) = -\sum_{j=1}^{n} \log\left\{ 1 - \frac{\theta}{j(n-j+1)} \right\} \tag{39}$$

and the rth cumulant is evidently

$$\kappa_r = (r-1)! \sum_{j=1}^{n} \frac{1}{j^r (n-j+1)^r}. \tag{40}$$

Each term on the right-hand side of (40) can be expanded in a series of partial fractions. If we collect together quantities with the same index we can write

$$\left.\begin{aligned} &\kappa_r = 2(r-1)!\sum_{p=1}^{r} a_p S_p, \\ \text{where}\quad &a_p = r(r+1)\dots(2r-p+1)/(r-p)!\,(n+1)^{2r-p} \quad (p<r), \\ &a_r = 1/(n+1)^r \\ \text{and}\quad &S_p = \sum_{u=1}^{n} \frac{1}{u^p}. \end{aligned}\right\} \tag{41}$$

Thus the first four cumulants are

$$\left.\begin{aligned} \kappa_1 &= \frac{2}{n+1} S_1, \\ \kappa_2 &= \frac{4}{(n+1)^3} S_1 + \frac{2}{(n+1)^2} S_2, \\ \kappa_3 &= \frac{24}{(n+1)^5} S_1 + \frac{12}{(n+1)^4} S_2 + \frac{4}{(n+1)^3} S_3, \\ \kappa_4 &= \frac{240}{(n+1)^7} S_1 + \frac{120}{(n+1)^6} S_2 + \frac{48}{(n+1)^5} S_3 + \frac{12}{(n+1)^4} S_4. \end{aligned}\right\} \tag{42}$$

I am indebted to Dr J. Wishart for pointing out to me that $S_p(n)$ is in general most easily computed by writing

$$\left.\begin{aligned} S_1(n) &= \psi(n+1) - \psi(1), \\ S_p(n) &= \frac{(-1)^{p-1}}{(p-1)!}\{\psi^{(p-1)}(n+1) - \psi^{(p-1)}(1)\} \quad (p>1), \end{aligned}\right\} \tag{43}$$

since values of the Polygamma Functions $\psi(x)$, $\psi^{(1)}(x)$, $\psi^{(2)}(x)$, ... are readily available from *Tables of the Higher Mathematical Functions* by Davis (1933, 1935).

For small n the cumulants are most easily calculated directly from (40) and for large n we can obtain asymptotic formulae by using the well-known expansions

$$\left.\begin{aligned} S_1(n) &\sim \log n + \gamma + \frac{1}{2n} - \frac{B_2}{2}\frac{1}{n^2} - \frac{B_4}{4}\frac{1}{n^4} - \dots, \\ S_p(n) &\sim \left\{\zeta(p) - \frac{1}{(p-1)\,n^{p-1}}\right\} + \frac{1}{n^p}\left\{\frac{1}{2} - \frac{B_2}{2}\binom{p}{1}\frac{1}{n} - \frac{B_4}{4}\binom{p+2}{3}\frac{1}{n^3} - \dots\right\}, \end{aligned}\right\} \tag{44}$$

where $\zeta(p)$ is Riemann's ζ-function, γ is Euler's constant and the B's are Bernoulli's numbers.

It is evident from (41), (42) and (44) that for large n

$$\left.\begin{aligned} \kappa_1 &= \frac{2}{(n+1)}\{\log n + \gamma + O(n^{-1})\}, \\ \kappa_r &= \frac{2(r-1)!}{(n+1)^r}\zeta(r)\{1 + O(n^{-1})\} \quad (r>1). \end{aligned}\right\} \tag{45}$$

Thus the coefficient of variation is asymptotically equal to

$$\pi/2\sqrt{3}\log n, \tag{46}$$

and the limiting values of γ_1 and γ_2, the coefficients of skewness and kurtosis, are given by

$$\left.\begin{aligned}\lim_{n\to\infty}\gamma_1 &= \frac{2^{\frac{1}{2}}\zeta(3)}{\{\zeta(2)\}^{\frac{3}{2}}} = 0{\cdot}806,\\ \lim_{n\to\infty}\gamma_2 &= \frac{3\zeta(4)}{\{\zeta(2)\}^2} = 1{\cdot}200.\end{aligned}\right\} \tag{47}$$

Values of $\bar{\tau}$, σ_τ, γ_1, γ_2 and the coefficient of variation are given in the following table for $n = 10, 20, 40, 80$ and ∞.

Some characteristics of the distribution of completion time

n	$\bar{\tau}$	σ_τ	γ_1	γ_2	c. of v. (%)
10	0·533	0·186	0·831	1·169	34·8
20	0·343	0·0938	0·774	1·081	27·4
40	0·209	0·0467	0·764	1·086	22·4
80	0·123	0·0231	0·771	1·114	18·9
∞	...	...	0·806	1·200	0

Thus it is evident that appreciable skewness and kurtosis remain even with large n. Furthermore, the coefficient of variation shows that for idealized communities in which the group size is 80 or less there will be considerable differences between groups with respect to the time taken for all the susceptibles of a group to become infected.

5. Summary

Classical mathematical investigations into the theory of epidemics have usually been deterministic, i.e. they have not taken probabilities into account. The present note attempts to make good this deficiency for a simple epidemic, where we have the spread of a relatively mild infection, in which none of the infected individuals is removed from circulation by death, recovery or isolation.

It is suggested that in general epidemic curves derived from *stochastic means* for the appropriate mathematical model would be likely to bear a close resemblance to the published returns for actual epidemics, because it is considered that the latter are in fact summed over a number of epidemics occurring in small groups of associates and acquaintances.

Curves of the stochastic means for the simple epidemic under consideration are given in the special cases when the group sizes are 10 and 20. The characteristics of the distribution of the time taken for the number of susceptibles to become exhausted are also discussed.

REFERENCES

Bartlett, M. S. (1946). *Stochastic Processes* (notes of a course given at the University of North Carolina in the Fall Quarter, 1946).

Bartlett, M. S. (1949). Some evolutionary stochastic processes. *J. R. Statist. Soc.*, Ser. B, **11**, 211.

Davis, H. T. (1933, 1935). *Tables of the Higher Mathematical Functions*, **1, 2**. Indiana: Principia Press.

Feller, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheoretica*, **5**, 11.

Kermack, W. O. & McKendrick, A. G. (1927 and later). Contributions to the mathematical theory of epidemics. *Proc. Roy. Soc.* A, **115**, 700; **138**, 55; **141**, 94.

Soper, H. E. (1929). Interpretation of periodicity in disease-prevalence. *J. R. Statist. Soc.* **92**, 34.