Adv. Appl. Prob. **6**, 607–621 (1974)
Printed in Israel

THE MAXIMUM SIZE OF A CLOSED EPIDEMIC

H. E. DANIELS, *University of Birmingham*

Abstract

An approximation is found to the distribution of the maximum number of infectives present at any time during the course of a closed epidemic. The technique used is applicable to a commonly occurring type of random walk problem where there is a curved absorbing boundary which is far from the mean path except over a narrow range.

EPIDEMICS; RANDOM WALKS; CURVED BOUNDARY

1.

An important feature of an epidemic is its *maximum size*, i.e., the maximum number of infectives ever present at anytime during the course of the epidemic. A knowledge of how large this is likely to be should influence the planning of adequate medical facilities for dealing with such epidemics.

In this paper an approximation is obtained to its distribution in the case of a closed epidemic with removals, by considering the diffusion approximation to the paths of the imbedded random walk. The problem is attacked by a "scanning" technique, developed elsewhere by the author, which yields a higher order approximation than is found by a simple passage to the Brownian limit. The mathematical treatment is to some extent heuristic.

2.

Initially there is a population of ξ susceptibles into which a small number η of infectives is introduced (typically $\eta = 1$). After time t there are x susceptibles, y infectives and z removals by death or recovery. Since the population is closed, $x+y+z = \xi+\eta$. Assuming a unit infection rate and a relative removal rate ρ, the possible transitions have probabilities

$$\begin{aligned} &\Pr(x, y \to x-1, y+1) &&= xy\delta t + o(\delta t), \\ &\Pr(x, y \to x, y-1) &&= \rho y \delta t + o(\delta t), \qquad (2.1) \\ &\Pr(x, y \to x, y) &&= 1 - (\rho + x)y\delta t + o(\delta t). \end{aligned}$$

Received in revised form 17 January 1974.

Consider first the maximum number of infectives attained in the deterministic model. The equations

$$\frac{dx}{dt} = -xy, \qquad \frac{dy}{dt} = xy - \rho y, \qquad \frac{dz}{dt} = \rho y \tag{2.2}$$

yield the deterministic path

$$y = \xi + \eta - x - \rho \log(\xi / x) \tag{2.3}$$

in the (x, y) plane. This attains its maximum

$$\hat{y} = \xi + \eta - \rho\{1 + \log(\xi/\rho)\} \tag{2.4}$$

at $\hat{x} = \rho$, provided $\xi > \rho$; otherwise there is no epidemic and $y \leqq \eta$.

Turning now to the stochastic model, it is sufficient to consider the imbedded random walk in the (x, y) plane with transition probabilities

$$\begin{aligned} \Pr(x, y \to x-1, y+1) &= x/(\rho + x), \\ \Pr(x, y \to x, y-1) &= \rho/(\rho + x). \end{aligned} \tag{2.5}$$

There is an absorbing boundary at $y = 0$ on which the epidemic terminates. If we insert a second absorbing boundary at $y = y_0$, the probability $\Pr(Y \geqq y_0)$ that the path is absorbed on it will give the required distribution of the maximum size y_0. The exact calculation of this probability is difficult, though numerical computation is feasible for ξ up to about 1000, using either the backward or forward equation for the random walk.

3.

If ξ is large the epidemic process in its early stages is well known to approximate to a simple birth-death process in y alone with transition probabilities

$$\Pr(y \to y+1) \sim \xi y \delta t, \qquad \Pr(y \to y-1) \sim \rho y \delta t$$

(Bartlett (1966), p. 141). The imbedded chain is then a simple random walk starting at $y = \eta$, and

$$\Pr(Y \geqq y_0) \sim \frac{1 - (\rho/\xi)^{\eta}}{1 - (\rho/\xi)^{y_0}}. \tag{3.1}$$

Since we are assuming that $\xi > \rho$ this gives information on the part of the distribution where y_0 is small, arising from paths where an epidemic failed to develop, the probability of which is $(\rho/\xi)^{\eta}$ in this simplified model.

The main difficulty lies in approximating to $\Pr(Y \geqq y_0)$ in the region corresponding to genuine epidemic paths. If ξ/ρ is not near 1 these paths separate off from the others, and given that an epidemic has developed, with probability

$1-(\rho/\xi)^{\eta}$, one can to an acceptable approximation ignore the lower absorbing boundary.

A more convenient formulation of the problem is in terms of *removals* z and *new cases* $m = \xi - x$, the transition probabilities being

$$\text{(3.2)}\qquad \begin{aligned} \Pr(z, m \to z+1, m) &= \rho/(\rho+\xi-m) = p_m, \\ \Pr(z, m \to z, m+1) &= (\xi-m)/(\rho+\xi-m) = 1-p_m. \end{aligned}$$

The upper boundary $y = y_0$ becomes a boundary at $z = m - y_0 + \eta$; the lower boundary, which we ignore, becomes $z = m + \eta$. The paths start at $m = 0$, $z = 0$.

The deterministic path is the solution of $dz/dm = p_m/(1-p_m)$ starting at the origin. It is

$$\text{(3.3)}\qquad z = \rho \log\{\xi/(\xi - m)\}$$

which is another form of (2.3). In the random walk, the step in the z direction at "time" m has a geometric distribution with mean $p_m/(1-p_m)$ and variance $p_m/(1-p_m)^2$. In passing to the diffusion limit the mean path of the random walk tends to the deterministic path (3.3). We transform to a new "time" scale

$$\text{(3.4)}\qquad n = \int_0^m \frac{p_m dm}{(1-p_m)^2} = \frac{\rho^2}{\xi - m} - \frac{\rho^2}{\xi} + \rho \log\left(\frac{\xi}{\xi - m}\right)$$

for which the rate of increase of var z is unity. We also write

$$\text{(3.5)}\qquad w = \rho \log\{\xi/(\xi - m)\} - z$$

which reduces the mean path to zero and produces an upper curved absorbing boundary at

$$\text{(3.6)}\qquad w = \rho \log\{\xi/(\xi - m)\} - m + y_0 - \eta = c_n, \qquad \text{say,}$$

except for the range $0 \leqq m \leqq y_0 - \eta$ where the natural boundary at $z = 0$ replaces it by $w = \rho \log\{\xi/(\xi - m)\}$. It has a minimum at $\hat{m} = \xi - \rho$ at which it takes the value $\hat{w} = \hat{y} - y_0$, $\hat{y}$ being the maximum (2.4) of the deterministic path. The boundary is shown in Figure 1, as a function of n.

However, there is an anomalous feature of (3.4) and (3.6) which has to be dealt with. It will be observed that for large n, $w \sim \rho \log n$, which implies that the boundary will be crossed with probability 1 whatever the value of y_0. This is obviously not true — for example, when $y_0 = \xi + \eta$ the probability is $\xi!\,\rho!/(\rho+\xi)!$. One reason for the anomaly is that if the actual random walk in the (x, y) plane crosses $x + y = y_0$ for some $x > 0$ it cannot cross the boundary: in fact, $x + y = y_0$ is an absorbing boundary. The situation is also complicated by the fact that the diffusion approximation breaks down for large n, as m approaches ξ.

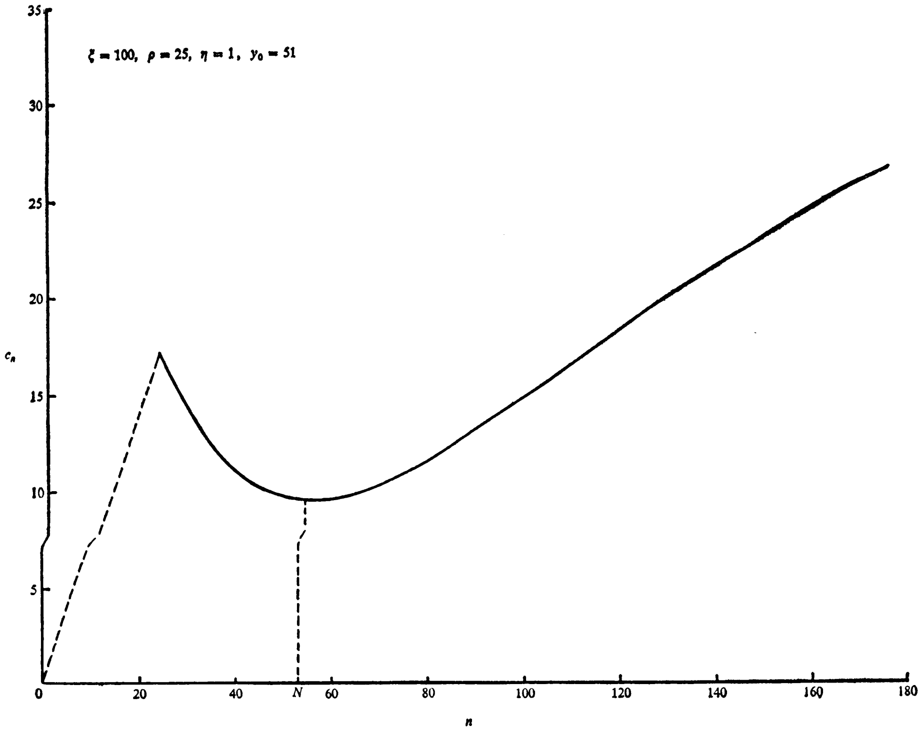


Figure 1.

One way of avoiding these difficulties is to truncate the boundary (3.6) at some suitable value of n beyond which it can be shown that the exact probability of crossing the boundary is negligible. This is a straightforward matter and we shall assume that it has been done.

4.

Consider what happens when ξ and ρ become large. The value of n at the minimum of the boundary (3.6) is[1]

$$N = \rho(\xi - \rho)/\xi + \rho \log(\xi/\rho). \tag{4.1}$$

Provided that $\xi - \rho$, ξ and ρ are all of comparable order, N is large and of the same order as ξ and ρ. It is convenient to use N to scale the process.

It follows from the work of Nagaev and Startsev (1970) that if

$$n = Nu, \quad w = N^{\frac{1}{2}}W(u), \quad c_n = N^{\frac{1}{2}}f(u), \tag{4.2}$$

[1] For typographical reasons in the ensuing formulae, N has been used rather than $\hat{n}$.

$W(u)$ tends to a standard Wiener process as $N \to \infty$ except for those values of n, excluded by the truncation of the boundary, for which $\xi - m$ is relatively small. Also the boundary $f(u)$ has its minimum at $u = 1$, near which it behaves like

$$f(u) \sim f(1) + \frac{N^{\frac{3}{2}}}{8\rho}(u-1)^2. \tag{4.3}$$

Since the coefficient of $(u-1)^2$ is of order $N^{\frac{1}{2}}$, $f(u)$ rises rapidly away from $f(1)$ on either side of $u = 1$. As a first approximation $f(u)$ may be replaced by a vertical "shutter" rising from $f(1) = N^{-\frac{1}{2}}(y_0 - \hat{y})$ to ∞, to which it tends as $N \to \infty$. This gives a simple approximation for the probability that an epidemic path crosses the boundary,

$$P(y_0) \sim 1 - \Phi\{N^{-\frac{1}{2}}(y_0 - \hat{y})\} \tag{4.4}$$

where $1 - P(y)$ is just the limiting distribution function of the number y of infectives at $\hat{x} = \rho$, the threshold number of susceptibles.

However, $f(u) - f(1)$ becomes large only outside a range $u - 1 = O(N^{-\frac{1}{4}})$ and (4.4) cannot be expected to be accurate unless N is very large. This is borne out by the calculations of Table 1. Exact distributions have been computed from the

TABLE 1

ξ	ρ	μ	σ	$\hat{y}$	$N^{\frac{1}{2}}$	$\hat{y}+(4\rho)^{\frac{1}{2}}$
50	10	28.2	4.5	24.9	4.9	28.3
100	10	71.0	5.1	68.0	5.7	71.4
100	25	45.8	6.7	41.4	7.3	46.0
100	50	22.7	6.9	16.3	7.7	22.2
1000	200	488.1	21.7	479.1	22.0	488.4

backward equations of the imbedded chain for the values of ξ and ρ listed, all with $\eta = 1$. The exact means μ and standard deviations σ are given for comparison with $\hat{y}$ and $N^{\frac{1}{2}}$. (In the case of $\xi = 100$, $\rho = 50$ the epidemic and non-epidemic parts of the distribution were not completely separated and μ, σ were estimated from a probit line fitted to the epidemic part.) A noticeable feature is the marked bias in the mean even for $\xi = 1000$, $\rho = 200$. For practical purposes a better approximation is required, and we therefore approach the problem in a different way which leads naturally to a higher order of approximation.

An essential feature of the problem is that under the normalization (4.2) most of the boundary is far from the mean path. Limit theorems for curved boundaries in such cases have been developed in considerable depth and detail by A. A. Borovkov (see Borovkov (1972) for a review and references). An alternative approach using martingales has also been discussed by Barbour (1972). I shall adopt what may be called a "scanning" technique (see Daniels (1945), (1963), (1967)) which seems particularly well suited to problems of this kind. The method is first developed in general and is then applied to the epidemic problem in Section 7.

5.

We have an upper absorbing boundary $w = c_n$, in our case defined by (3.4), (3.6), which takes its minimum value c_N at some large N. Instead of (4.2) we write

$$n = Nu, \qquad w = Nv, \qquad c_n = Nc(u), \tag{5.1}$$

where $c(u)$ is assumed twice differentiable, so that $c'(1) = 0$, $c''(1) > 0$ at the minimum $u = 1$. This normalization is appropriate because the boundary $v = c(u)$ is to be non-degenerate as $N \to \infty$. It is assumed that $c(u)$ is truncated beyond a suitably remote value U of u for the diffusion appoximation to be applicable. The unconditional probability that the path reaches v at time u is approximately

$$d\Phi\{v(N/u)^{\frac{1}{2}}\} = \exp\{-\tfrac{1}{2}Nv^2/u\}\,(N/2\pi u)^{\frac{1}{2}}dv \tag{5.2}$$

with a relative error of $O(N^{-\frac{1}{2}})$, where dv replaces the discrete step N^{-1} in v.

If $(u_1, c(u_1))$ is a point on the boundary, and $g(u)$ is the first passage time density, we have, to the same order, the renewal equation

$$\begin{aligned}\exp\left\{-\frac{Nc^2(u_1)}{2u_1}\right\}\left(\frac{N}{2\pi u_1}\right)^{\frac{1}{2}} &= \int_0^{u_1}\exp\left\{\frac{-N\{c(u_1)-c(u)\}^2}{2(u_1-u)}\right\} \\ &\qquad \times\left(\frac{N}{2\pi(u_1-u)}\right)^{\frac{1}{2}} g(u)du.\end{aligned} \tag{5.3}$$

This is now used to find an approximation to $g(u)$ and hence to the probability of crossing $c(u)$:

$$P = \int_0^u g(u)\,du. \tag{5.4}$$

It is helpful to write $g(u)$ in the form

$$g(u) = \exp\left\{-\frac{Nc^2(u)}{2u}\right\}\left(\frac{N}{2\pi u}\right)^{\frac{1}{2}} h(u) \tag{5.5}$$

since $h(u)$ turns out to vary slowly over the interesting range of u. (For example, if $c(u) = \alpha + \beta u$ it is known that $h(u) = \alpha/u$ which varies slowly except near $u = 0$.) With this substitution (5.3) reduces to

$$1 = \int_0^{u_1}\exp\left\{-\frac{Nu_1}{2u(u_1-u)}\left\{u\frac{c(u_1)}{u_1}-c(u)\right\}^2\right\}\left(\frac{Nu_1}{2\pi u(u_1-u)}\right)^{\frac{1}{2}} h(u)du. \tag{5.6}$$

The exponential part of the integral in (5.6) has sharp peaks at those roots of

$$c(u) = uc(u_1)/u_1 \tag{5.7}$$

which lie in $0 < u \leqq u_1$ and it is negligible elsewhere. The only appreciable contributions to the integral arise from values of u near these roots. One of the roots is $u = u_1$, but there may be smaller ones at the intersections of $v = uc(u_1)/u_1$ with $v = c(u)$ (see Figures 2,3). By varying u_1 the position of the roots can be moved and extra roots can be included. In this way, the whole range of u can be scanned to give information about $h(u)$.

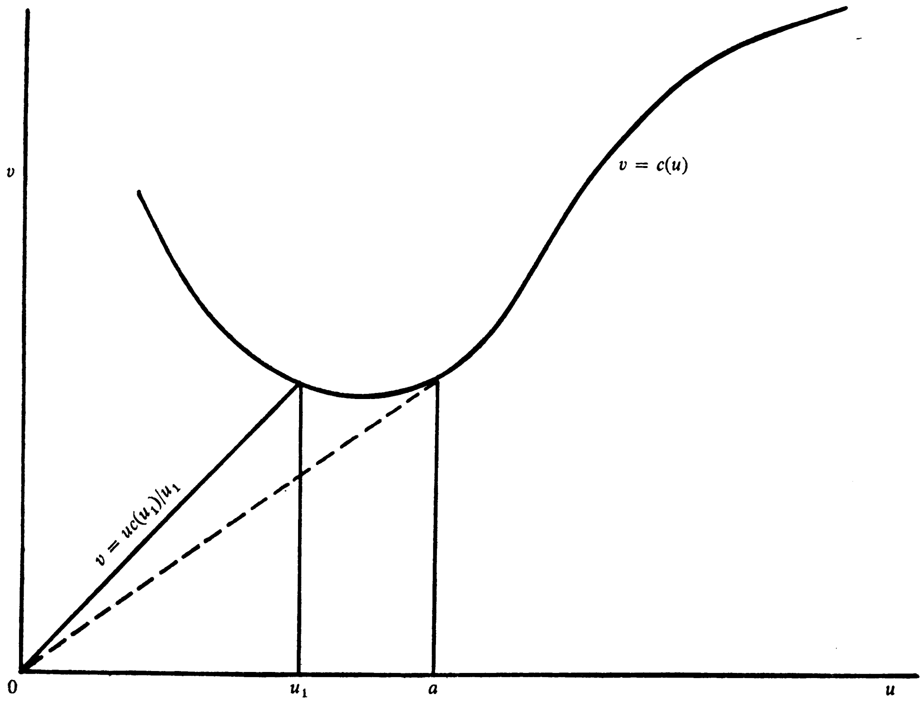


Figure 2.

There are special values of u where the line $v = uc(u_1)/u_1$ touches $v = c(u)$; they satisfy

$$c'(u) = c(u)/u. \tag{5.8}$$

In our case there are two such values a, b, $(a < b)$ at which $c(u)/u$ has a minimum and a maximum respectively.

(i) We first consider the range $0 < u_1 \leqq a - \varepsilon$ where ε is independent of N. Here u_1 is the only root of (5.7) in $0 < u \leqq u_1$. The integrand in (5.6) is negligible except when $u_1 - u$ is $O(N^{-1})$, and (5.6) approximates to

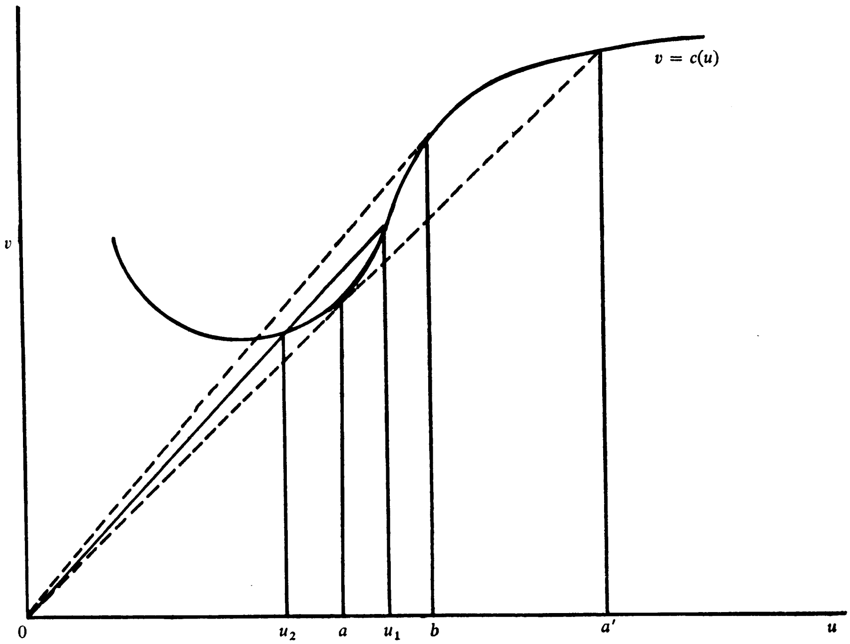


Figure 3.

$$(5.9)\quad 1 \sim \int_{-\infty}^{u_1} \exp\{-\tfrac{1}{2}N\{c(u_1)/u_1 - c'(u_1)\}^2(u_1 - u)\}\left(\frac{N}{2\pi(u_1 - u)}\right)^{\frac{1}{2}} h(u)du.$$

Assuming that $h(u)$ can be expanded in powers of $u - u_1$ it is easily found that

$$(5.10)\qquad 1 \sim h(u_1)\{c(u_1)/u_1 - c'(u_1)\}^{-1} + O(N^{-1})$$

and because u_1 can have any value in $0 < u \leqq a - \varepsilon$ it follows that

$$(5.11)\qquad h(u) \sim c(u)/u - c'(u) + O(N^{-1}), \qquad 0 < u \leqq a - \varepsilon.$$

The error $O(N^{-1})$ in the approximation was deduced from (5.9), but one also has to consider the effects of the passage from (5.6) to (5.9) and the original normal approximation (5.2). Detailed study of the terms neglected indicates that the remainder $O(N^{-1})$ incorporates the errors of these other approximations also.

(ii) Avoiding a transitional region near a, to be discussed shortly, we next consider the range $a + \varepsilon \leqq u_1 \leqq b - \varepsilon$ for which there are two roots of (5.7), $u_1 \geqq a + \varepsilon$ and $u_2 \leqq a - \varepsilon$. The neighbourhood of each root provides a contribution to the integral in (5.6). Near $u = u_1$ the previous argument shows that the contribution is $h(u_1)\{c'(u_1) - c(u_1)/u_1\}^{-1} + O(N^{-1})$ since $c'(u_1)$ now exceeds $c(u_1)/u_1$. Near $u = u_2$ the integrand is appreciable only when $u - u_2 = O(N^{-\frac{1}{2}})$, and the contribution from this region is approximately

$$h(u_2)\int_{-\infty}^{\infty}\exp\left\{-\frac{Nu_1}{2u_2(u_1-u_2)}\{c(u_2)/u_2-c'(u_2)\}^2(u-u_2)^2\right\}\left(\frac{Nu_1}{2\pi u_2(u_1-u_2)}\right)^{\frac{1}{2}}du$$

(5.12)

$$= h(u_2)\{c(u_2)/u_2-c'(u_2)\}\sim 1$$

by the previous result (5.10) with u_2 replacing u_1. The error is again $O(N^{-1})$ because odd powers of $u-u_1$ in the higher order terms of the expansions contribute nothing. It follows that $h(u)\sim O(N^{-1})$ over the range $a+\varepsilon\leqq u\leqq b-\varepsilon$.

When u_1 exceeds $b+\varepsilon$ the number of roots increases to 3, but the same argument indicates that $h(u)$ remains $O(N^{-1})$, and this can also be shown to hold throughout $b-\varepsilon\leqq u\leqq b+\varepsilon$. However, when u_1 passes beyond the value a' at which $v=uc(a)/a$ intersects $v=c(u)$ again, the number of roots drops to 1 and the situation returns to (i). We can therefore say that

(5.13) $$h(u)\sim O(N^{-1}),\ a+\varepsilon<u<a'-\varepsilon,$$

(5.14) $$h(u)\sim c(u)/u-c'(u)+O(N^{-1}),\ a'+\varepsilon<u.$$

The results (5.11), (5.13), (5.14) have an intuitive interpretation. The integrand of (5.6) can be regarded as the first passage time density of the "tied down" process passing through $(u_1, c(u_1))$. The roots of (5.7) are the intersections of its mean path with the boundary, and (5.12) or (5.14) with $u=u_1$ are the known exact results when the boundary is replaced by its tangent at u_1, there being no other root in $0<u\leqq u_1$. In the case of (5.13), nearly all the absorption takes place near where the mean path first hits the boundary at $u_2<u_1$.

(iii) There is a transitional region near $u=a$ where $h'(u)$ changes smoothly from nearly $-c''(a)$ when $u<a$ to 0 when $u>a$. As u_1 approaches a, $c(u_1)/u_1-c'(u_1)$ becomes comparable with u_1-u in magnitude and it is necessary to include the quadratic term when expanding the exponent in (5.6). So we write

$$uc(u_1)/u_1-c'(u)=\{c(u_1)/u_1-c'(u_1)\}(u-u_1)$$
$$-\tfrac{1}{2}c''(u_1)(u-u_1)^2+O(u-u_1)^3,$$

(5.15) $$c(u_1)/u_1-c'(u_1)=c''(a)(a-u_1)+O(a-u_1)^3,$$

$$c''(u_1)=c''(a)+O(a-u_1)$$

and (5.6) becomes

$$1\sim\int_{-\infty}^{u_1}\exp\{-\tfrac{1}{2}N\{c''(a)\}^2(u_1-u)\{a-u_1+\tfrac{1}{2}(u_1-u)\}^2\}$$

(5.16)

$$\times\left(\frac{N}{2\pi(u_1-u)}\right)^{\frac{1}{2}}h(u)du,$$

the first neglected terms involving $u_1 - u$ and $N(u_1 - u)^4$. The important values of u and u_1 are such that $u_1 - u$ and $a - u_1$ are $O(N^{-\frac{1}{3}})$, and the neglected terms are then seen to be $O(N^{-\frac{1}{3}})$. The substitution

$$(5.17)\qquad \begin{aligned} x &= N^{\frac{1}{3}}\{c''(a)\}^{\frac{2}{3}}(a - u_1), \\ y &= N^{\frac{1}{3}}\{c''(a)\}^{\frac{2}{3}}(u_1 - u), \end{aligned}$$

then leads to the result that

$$(5.18)\qquad h(u) \sim N^{-\frac{1}{3}}\{c''(a)\}^{\frac{1}{3}}F(N^{\frac{1}{3}}\{c''(a)\}^{\frac{2}{3}}(a - u))$$

where $F(\cdot)$ satisfies the integral equation

$$(5.19)\qquad 1 = \int_0^\infty F(x + y)\exp\{-\tfrac{1}{2}y(x + \tfrac{1}{2}y)^2\}\,(2\pi y)^{-\frac{1}{2}}dy.$$

An explicit solution for $F(x)$ has not been found and $F(x)$ has to be tabulated numerically. However, something is known about its asymptotic behaviour. When $|x|$ is large one would expect that $F(x) \sim (x)^+$. In fact, when $x \gg 0$ there is an asymptotic expansion

$$F(x) \sim x + \tfrac{1}{2}x^{-2} - 2x^{-5} + \cdots$$

in powers of x^{-3r+1}; when $x \ll 0$, calculations so far show that $F(x) \sim o(|x|^{-8})$ and it may well be exponentially small.

There is another type of transitional behaviour near $u = a'$ similar to that discussed in Section 6 which need not concern us.

6.

Armed with this information about $h(u)$ we can now discuss the probability (5.4) of crossing $c(u)$ which we write in the form

$$(6.1)\qquad P = \int_0^u \exp\left\{-\frac{Nc^2(u)}{2u}\right\}\left(\frac{N}{2\pi u}\right)^{\frac{1}{2}} h(u)du.$$

The behaviour of P depends on the position of $v = c(u)$ relative to the mean path $v = 0$. One intuitively expects that $P \sim 0$ when $c(1) \gg 0$ and $P \sim 1$ when $c(1) \ll 0$. In the first case, replacing $c(u)$ by a horizontal boundary of height $c(1)$ over $0 \leqq u \leqq U$ gives the crude upper bound

$$(6.2)\qquad P < 2\{1 - \Phi(c(1)\,[N/U]^{\frac{1}{2}})\} \sim \left(\frac{2U}{\pi N}\right)^{\frac{1}{2}} \exp\{-NC^2(1)/2U\}/c(1)$$

which could certainly be improved. In the second case, suppose that $c(u)$ crosses $v = 0$ at points u_1, u_2, such that $u_1 > a + \varepsilon$, $u_2 < a - \varepsilon$ where a and ε are as previously defined (see Figure 4). Then using (5.11) and (5.13) we find that

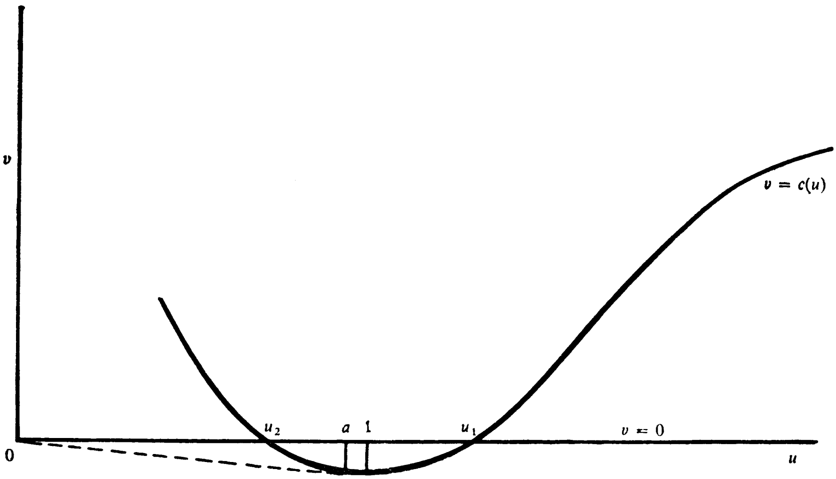


Figure 4.

$$(6.3)\qquad \begin{aligned} P &\sim \int_{-\infty}^{\infty} \exp\left\{-\frac{N}{2u}\{c'(u_2)\}^2(u-u_2)^2\right\} \left|c'(u_2)\right| \left(\frac{N}{2\pi u_2}\right)^{\frac{1}{2}} du + O(N^{-1}) \\ &= 1 + O(N^{-1}). \end{aligned}$$

It follows that the interesting case is when $c(1)$ is small. The only appreciable values of the integrand in (6.1) are near $u = 1$. On expanding $c(u)$ as far as the quadratic term, (6.1) becomes

$$(6.4)\qquad P \sim \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}N\{c(1) + \tfrac{1}{2}c''(1)\,(u-1)^2\}^2\}\left(\frac{N}{2\pi}\right)^{\frac{1}{2}} h(u)du,$$

the first neglected terms in the integrand involving $u - 1$, $N(u-1)^5$ and $Nc(1) \cdot (u-1)^3$. The important range of u is seen to be $u - 1 = O(N^{-\frac{1}{4}})$ and $c(1)$ is taken to be $O(N^{-\frac{1}{2}})$. The neglected terms are then $O(N^{-\frac{1}{4}})$. Notice also that

$$(6.5)\qquad 1 - a \sim c'(a)/c''(a) = c(a)/ac''(a) = O(N^{-\frac{1}{2}})$$

so that a can be replaced by 1 to the order considered.

We now have to approximate to $h(u)$ over a range of $u - 1$ which is $O(N^{-\frac{1}{4}})$. This is intermediate between $O(1)$ appropriate to (i) and $O(N^{-\frac{1}{3}})$ appropriate to (iii). Some care is therefore needed when discussing the order of magnitude of the various terms[2].

For values of $|u - 1|$ which are comparable with $N^{-\frac{1}{4}}$, and hence large compared with $N^{-\frac{1}{3}}$, the asymptotic form of $F(x)$ shows that $h(u)$ behaves like $c''(1)(1-u)^+$. For smaller values of $|u - 1|$ the effect of $F(x)$ is to smooth

[2] The discussion in Daniels (1945) is not adequate in this respect.

out the abrupt change of gradient of $(1-u)^{+}$ at $u=1$. (Notice that the argument leading to (5.18) is unaffected by the fact that $|c(1)|$ is small.) We therefore write

$$h(u) = c'(1)\,(1-u)^{+} + k(u) \tag{6.6}$$

and correspondingly

$$F(x) = (x)^{+} + G(x) \tag{6.7}$$

where $G(x) \sim \frac{1}{2}x^{-2}$, $x \gg 0$; $G(x) = o(|x|^{-8})$, $x \ll 0$. Then (6.4) becomes

$$P \sim 1 - \Phi\{N^{\frac{1}{2}}c(1)\} + Q + O(N^{-\frac{1}{4}}). \tag{6.8}$$

The contribution from the first term of (6.6) is just the simple approximation (4.4). The extra term is

$$\begin{aligned} Q &= \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}N\{c(1) + \tfrac{1}{2}c''(1)(u-1)^2\}^2\} \left(\frac{N}{2\pi}\right)^{\frac{1}{2}} k(u)\,du \\ &= \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}N\{c(1) + \tfrac{1}{2}N^{-\frac{2}{3}}[c''(1)]^{-\frac{1}{3}}x^2\}^2\} \frac{N^{-\frac{1}{6}}}{(2\pi)^{\frac{1}{2}}} \{c''(1)\}^{-\frac{1}{3}}G(x)dx \end{aligned} \tag{6.9}$$

$$= \exp\{-\tfrac{1}{2}N\{c(1)\}^2\}\frac{N^{-\frac{1}{6}}}{(2\pi)^{\frac{1}{2}}} \{c''(1)\}^{-\frac{1}{3}} \int_{-\infty}^{\infty} G(x)dx + R. \tag{6.10}$$

To estimate the remainder R we cannot expand the exponent in (6.9) since $G(x) \sim O(x^{-2})$ for large $x > 0$. Instead, assuming $G(x)$ is bounded and positive, as numerical calculations of $F(x)$ suggest, we choose $A > 0$ such that $G(x) + G(-x) < Ax^{-2}$ and observe that, with $\alpha = N^{\frac{1}{2}}c(1)$,

$$\begin{aligned} &\left(\frac{N^{-\frac{1}{3}}}{2\pi}\right)^{\frac{1}{2}} \{c''(1)\}^{-\frac{1}{3}} \int_0^{\infty} x^{-2}\{e^{-\frac{1}{2}\alpha^2} - \exp\{-\tfrac{1}{2}\{\alpha + \tfrac{1}{2}N^{-\frac{1}{6}}[c''(1)]^{-\frac{1}{3}}x^2\}^2\}\}\,dx \\ &= \frac{N^{-\frac{1}{4}}}{(2\pi)^{\frac{1}{2}}} \{c''(1)\}^{-\frac{1}{4}} \int_0^{\infty} z^{-2}\{e^{-\frac{1}{2}\alpha^2} - \exp\{-\tfrac{1}{2}(\alpha + z^2)^2\}\}dz. \end{aligned} \tag{6.11}$$

This implies that R is $O(N^{-\frac{1}{4}})$, and that

$$P \sim 1 - \Phi\{N^{\frac{1}{2}}c(1)\} + \lambda \frac{N^{-\frac{1}{6}}}{(2\pi)^{\frac{1}{2}}} \{c''(1)\}^{-\frac{1}{3}} \exp\{-\tfrac{1}{2}N\{c(1)\}^2\} + O(N^{-\frac{1}{4}}) \tag{6.12}$$

where

$$\lambda = \int_{-\infty}^{\infty} \{F(x) - (x)^{+}\}\,dx. \tag{6.13}$$

However, to the same order (6.12) can be replaced by the simpler form

$$P \sim 1 - \Phi\{N^{\frac{1}{2}}c(1) - \lambda N^{-\frac{1}{6}}[c''(1)]^{-\frac{1}{3}}\} + O(N^{-\frac{1}{4}}) \tag{6.14}$$

which may also be conveniently expressed on the original scale as

$$P \sim 1 - \Phi\{N^{\frac{1}{2}}[c_N - \lambda(c''_N)^{-\frac{1}{3}}]\} + O(N^{-\frac{1}{4}}) \tag{6.15}$$

where $c_N = Nc(1)$.

7.

Let us now return to the notation of the epidemic problem. The boundary $w = c_n$ was defined by (3.4) and (3.6), for which $c_N = y_0 - \hat{y}$ and $C''_N = (4\rho)^{-1}$. The approximation (6.15) is then

$$P \sim 1 - \Phi\{N^{-\frac{1}{2}}[y_0 - \hat{y} - \lambda(4\rho)^{\frac{1}{3}}]\} + O(N^{-\frac{1}{4}}) \tag{7.1}$$

where $\hat{y}$ is given by (2.4), λ by (6.13), and

$$N = \rho\{\log(\xi/\rho) + (\xi - \rho)/\xi\}. \tag{7.2}$$

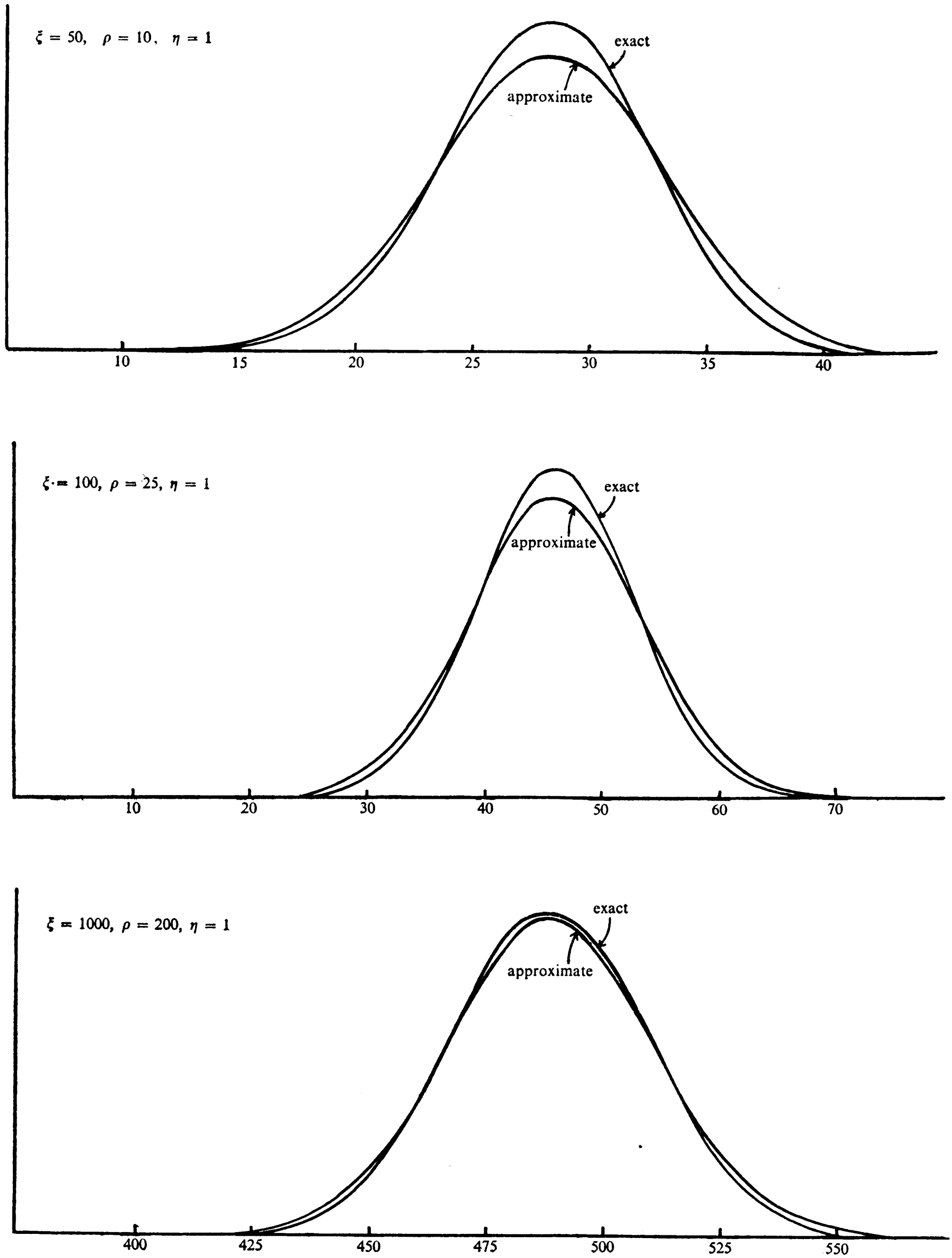
It was observed in Section 4 that the simple normal approximation (4.4) to the distribution of y_0 for epidemic paths involved an appreciable bias in the mean even for large values of ξ and ρ. This bias is now seen to be $\lambda(4\rho)^{\frac{1}{3}}$ and if it is ignored the remainder is increased to $O(N^{-\frac{1}{6}})$. The constant λ has to be either evaluated from the solution of the integral equation (5.19), or estimated by comparing $\hat{y}$ with the mean of the exact distribution computed numerically for various values of ξ and ρ. Preliminary computations on the integral equation suggest that $\lambda \sim 1{\cdot}0$, and this value has been used in the final column of Table 1 which gives $\hat{y} + (4\rho)^{\frac{1}{3}}$.

It will be seen from Table 1 that in all cases $(4\rho)^{\frac{1}{3}}$ almost completely removes the bias in $\hat{y}$ as an approximation to μ. The standard deviation σ is overestimated by $N^{\frac{1}{2}}$, but the effect becomes negligible when $\xi = 1000$, $\rho = 200$. The exact and approximate distributions for $\xi = 50$, $\rho = 10$; $\xi = 100$, $\rho = 25$; $\xi = 1000$, $\rho = 200$ are shown in Figure 5. The normal approximation is seen to fit the distribution well.

The agreement found is almost too good considering that the remainder is $O(N^{-\frac{1}{4}})$. It would be interesting to discover a reason for this, as a superficial examination of the first terms neglected shows no obvious tendency for them to cancel out.

Acknowledgments

I am grateful to Mr. A. D. Barbour for useful discussions, and to a referee for valuable comments on a previous draft of the paper.



$\xi = 50, \quad \rho = 10, \quad \eta = 1$
exact
approximate
10
15
20
25
30
35
40
$\xi = 100, \rho = 25, \eta = 1$
exact
approximate
10
20
30
40
50
60
70
$\xi = 1000, \rho = 200, \eta = 1$
exact
approximate
400
425
450
475
500
525
550

Figure 5.

References

Bartlett, M. S. (1966) *Stochastic Processes.* 2nd ed. Cambridge University Press.

Barbour, A. D. (1972) The principle of the diffusion of arbitrary constants. *J. Appl. Prob.* **9,** 519–541.

Borovkov, A. A. (1972) Limit theorems for random walks with boundaries. *Proc. Sixth Berkeley Symposium* **3,** 19–30.

Daniels, H. E. (1945) The statistical theory of the strength of bundles of threads. *Proc. Roy. Soc. London* A **183,** 405–435.

Daniels, H. E. (1963) The Poisson process with a curved absorbing boundary. *Bull. Inst. Internat. Statist.* 34*th Session* **40,** 994–1008.

Daniels, H. E. (1967) The distribution of the total size of an epidemic. *Proc. Fifth Berkeley Symposium* **4,** 281–293.

Nagaev, A. V. and Startsev, A. N. (1970) The asymptotic analysis of a stochastic model of an epidemic. *Theor. Probability Appl.* **15,** 98–107.