Chapter 4

MODELS IN EPIDEMICS

4.1 Generalities

One of the important areas of real-life applications of stochastic processes is in epidemiology, more specifically, in analyzing the spread of epidemics. Roughly speaking, the situation that we want to look at is as follows. There is a group of individuals, all mixing homogeneously together. Due to some reason, one or more of them contracted an infectious disease. They are the individuals *initially infected.* After a certain period of time, called the *latent period*, the infected become *infectious.* This means that they are now capable of passing on the infection to other individuals in the population to be called *susceptibles.* This leads to new infections. Thus, some of the susceptibles move to the infected group and this continues. Simultaneously, as time passes there is also what is called *removal* of infectious from circulation. Such removals in reality may take place by way of death or by way of detection and quarantine. Of course, removal may not always mean that the concerned individuals are actually taken out of the population. For example, an infectious person may have been cured and has become immune. Thus, as far as the epidemic is concerned, they are as good as removed.

In the next few sections, we shall model such phenomena mathematically. In the light of the different models, we would like to investigate how the epidemic progresses in terms of the changes in the number of infected and susceptibles. A quantity of vital interest is the total size of the epidemic, to be explained later. The rate at which the infection spreads is called the *rate of infection*, whereas the rate at which the infectious individuals get removed is called the *rate of removal.* If the infection rate is too small compared to the removal rate, then one intuitively feels that the epidemic should not build up. An important class of theorems in epidemiology known as the *Threshold Theorems* are aimed at justifying this mathematically. One distinctive feature of this chapter is that unlike the previous chapters where the models were discrete in time, here we have an evolution taking place in continuous time. As a result, in Sections 4.2

A. Goswami and B. V. Rao, *A Course in Applied Stochastic Processes*

and 4.3, we will use continuous time Markov chains.

Needless to say that the description given above conforms to the real world meaning of epidemics. However the same picture obtains if we wish to study the spread of a disease across the cells of a single individual.

4.2 Simple Epidemic

For the sake of simplicity, we first consider the case when there is no latent period and also there are no removals. The first assumption means that an individual becomes infectious as soon as he receives the infection. In the absence of removals, an infectious individual remains in circulation forever. In this case, it is intuitively clear that infection would continue to spread until all are infected.

4.2.1 Deterministic Model

We start with a deterministic model first. We consider a population of $n+1$ individuals in which initially, that is, at time $t=0$, there are n susceptibles and 1 infectious. We denote by $x(t)$ and $y(t)$, the number of susceptibles and the number of infected individuals respectively at time t. Of course, it is clear that for every t, $x(t)+y(t)=n+1$. Also $x(0)=n$ and $y(0)=1$. In view of the fact that $y(t)=n+1-x(t)$, it suffices to describe $x(t)$. The central step in modelling the process $x(t)$ involves deciding on the mechanism governing the evolution of $x(t)$. More precisely, suppose that at some time instant, say t, we have $x(t)=a$. This means that at time t there are a susceptibles and $n+1-a$ infected in the population. The question is how $x(t)$ should change in a small time interval, say Δt. This, of course, means how many new infections take place during the time period $(t, t+\Delta t)$. This is where we bring in our modelling assumptions. First of all, the number of new infections should be proportional to the duration of the interval, namely, Δt. Indeed, one does feel that the number of new infections in a given time interval should be large or small depending on whether the interval is large or small. Secondly, the number of new infections should be proportional to the possible number of contacts between the infected and the susceptibles. Since at time t we have a susceptibles and $n+1-a$ infected the possible number of contacts (that is, pairings) between these two groups is $a(n+1-a)$. Our modelling assumption, therefore, reduces to speculating that $x(t+\Delta t)-x(t) \sim -\beta x(t)[n+1-x(t)]\Delta t$. Here β is a positive constant, frequently referred to as the infection rate. Dividing by Δt and taking the limit as $\Delta t \to 0$ the above amounts to

$$x'(t) = -\beta x(t)[n+1-x(t)]\,. \tag{1}$$

This is our precise mathematical assumption regarding the evolution of $x(t)$. An initiated reader would, of course, raise objections at this point. The fact is that $x(t)$ denotes the number of susceptibles at time t and is hence integer

valued. Let alone being differentiable, $x(t)$ cannot even be continuous unless it is a constant function of t. Here is one way to make sense out of Equation (1). Assume that the population size $(n+1)$ is large and consider the proportion $\bar{x}(t) = x(t)/(n+1)$ of susceptibles rather than $x(t)$ itself. In that case $\bar{x}(t)$ can be regarded as taking values in the continuum $[0,1]$ (at least approximately). Our modelling assumptions can now be summarized in a genuine differential equation for $\bar{x}(t)$, obtained as follows. As seen earlier, for small Δt,

$$\frac{x(t+\Delta t) - x(t)}{\Delta t} \sim -\beta x(t)(n+1-x(t)),$$

that is,

$$\frac{\bar{x}(t+\Delta t) - \bar{x}(t)}{\Delta t} \sim -\beta(n+1)\bar{x}(t)(1-\bar{x}(t),$$

which on taking limit as $\Delta t \downarrow 0$ yields

$$\bar{x}'(t) = -\beta(n+1)\bar{x}(t)[1-\bar{x}(t)]\,.$$

Of course, all this is just a matter of mathematical precision and should be viewed as a way of rationalizing Equation (1). In any case you should remember that Equation (1) itself reflects only an approximation for the actual state of affairs.

Returning now to Equation (1), it can be written as

$$\frac{x'(t)}{x(t)} + \frac{x'(t)}{n+1-x(t)} = -\beta(n+1)\,.$$

By integrating and using the initial condition $x(0) = n$, one obtains

$$x(t) = \frac{n(n+1)}{n + e^{\beta(n+1)t}}\,.$$

It follows that

$$y(t) = \frac{(n+1)e^{\beta(n+1)t}}{n + e^{\beta(n+1)t}}\,.$$

The rate at which the infections accrue is given by

$$\begin{aligned}\omega(t) &= y'(t) = -x'(t)\\ &= \beta x(t)[n+1-x(t)]\\ &= \beta n(n+1)^2 \frac{e^{\beta(n+1)t}}{[n+e^{\beta(n+1)t}]^2}\,.\end{aligned}$$

This is of considerable interest in epidemiology and the graph of $\omega(t)$ is called the *epidemic curve*. The above epidemic curve starts at $\omega(0) = \beta n$, increases until it reaches a peak at time $t_0 = \log n/[\beta(n+1)]$ and then gradually dies down. The time point t_0 is usually of some interest. In our model, the number of susceptibles and the number of infected become almost equal at time t_0.

The fact that $\lim_{t\to\infty} x(t) = 0$ is only natural because in this model the entire population is clearly going to be infected eventually. This completes our discussion of the deterministic model. After all, the main purpose of this model is to motivate a more realistic model, namely, a stochastic one.

4.2.2 Simple Stochastic Epidemic

In the stochastic model we do not say that in the small time-interval t to $t + \Delta t$, a certain number of new infections is sure to take place. Instead, we introduce a chance mechanism for the number of new infections. As before, let us imagine a population of size $n + 1$ with n susceptibles and 1 infected initially. Let X_t and Y_t denote the number of susceptibles and the number of infected, respectively, at time t. Here X_t and Y_t are going to be random variables. Of course $X_t + Y_t = n + 1$, for all t ; $X_0 = n$ and $Y_0 = 1$. Now we come to the main assumption. Given that at time t, $X_t = a$ and $Y_t = n+1-a$, we assume that during the time interval t to $t+\Delta t$, the probability of exactly one new infection is $\beta a(n + 1 - a)\Delta t + o(\Delta t)$ and that of no new infection is $1 - \beta a(n+1-a)\Delta t + o(\Delta t)$. This of course implies that the probability of two or more new infections during the period t to $t + \Delta t$ is $o(\Delta t)$. Thus $(X_t)_{t\geq 0}$ is a continuous time pure death chain starting at $X_0 = n$ (see the concluding paragraph of Section 0.9.1). The death rates are given by

$$\mu_i = \beta i(n + 1 - i), \qquad i = 0, 1, \ldots, n\,.$$

Denoting, $p_r(t) = P[X_t = r]$, the usual Kolmogorov equations are

$$p_r'(t) = -\beta r(n + 1 - r)p_r(t) + \beta(r + 1)(n - r)p_{r+1}(t) \quad \text{for } 0 \leq r < n,$$
$$\text{and} \quad p_n'(t) = -\beta n p_n(t)\,. \tag{2}$$

Of course $p_n(0) = 1$ where as $p_r(0) = 0$ for $0 \leq r < n$.

As we shall see later, it is possible to solve the above equations successively for $p_n, p_{n-1}, \ldots, p_0$. But the formulae are too complicated to give an insight into the phenomenon. Of course, since $(X_t)_{t\geq 0}$ is a pure death chain with death rates μ_i given in (1), the general theory tells us that the first new infection takes place after a random time distributed as $\mathcal{E}xp(\mu_n)$, the next infection occurs after a further random time with distribution $\mathcal{E}xp(\mu_{n-1})$, and so on. Finally, the process comes to a halt when all are infected. This happens in a finite time with probability one. Denoting by T the total duration, it is clear that T is the sum of n independent exponential random variables with parameters $\mu_n, \ldots, \mu_1$. One can find the exact distribution of T. However, we shall be

content with noting that

$$\begin{aligned} E(T) &= \frac{1}{\beta}\sum_{i=1}^{n}\frac{1}{i(n+1-i)} \\ &= \frac{1}{\beta(n+1)}\sum_{i=1}^{n}\left[\frac{1}{i}+\frac{1}{n+1-i}\right] \\ &= \frac{2}{\beta(n+1)}\sum_{i=1}^{n}\frac{1}{i}\,. \end{aligned}$$

But,

$$\sum_{i=1}^{n}\frac{1}{i} = \sum_{i=1}^{n}\frac{1}{i}\int_{i}^{i+1}dx \geq \sum_{i=1}^{n}\int_{i}^{i+1}\frac{1}{x}\,dx = \log(n+1)$$

and,

$$\sum_{i=1}^{n}\frac{1}{i} = 1+\sum_{i=2}^{n}\frac{1}{i}\int_{i-1}^{i}dx \leq 1+\sum_{i=2}^{n}\int_{i-1}^{i}\frac{1}{x}\,dx = 1+\log n\,.$$

Thus,

$$\frac{2}{\beta}\frac{\log(n+1)}{n+1} \leq E(T) \leq \frac{2}{\beta}\frac{1+\log n}{n+1}\,.$$

That is, $E(T) = O(\log n/n)$. Incidentally, this also shows that $E(T) \downarrow 0$ as $n \to \infty$. This seems to contradict one's first intuition that large populations should take longer to reach complete infection. However, one should not forget that larger population implies increased death rate also.

Returning to the Kolmogorov Equations (2), we now indicate how one can go about solving them and also what the nature of the expressions for $p_r(t)$ is. Firstly, it is immediate that

$$p_n(t) = e^{-\beta n t}\,.$$

This can also be directly seen from the fact that $p_n(t)$ is the probability that upto and including time t, no new infection has taken place and that the time till the first new infection is exponentially distributed with parameter βn. Next, the Equations (2) can be rewritten as

$$p_r'(t) + \mu_r p_r(t) = \mu_{r+1}p_{r+1}(t) \quad \text{for } 0 \leq r < n\,.$$

Multiplying both sides by $e^{\mu_r t}$ one obtains

$$\frac{d}{dt}\left[p_r(t)e^{\mu_r t}\right] = \mu_{r+1}e^{\mu_r t}p_{r+1}(t)\,.$$

Solving this one obtains the recurrence relation

$$p_r(t) = e^{-\mu_r t}\int_0^t \mu_{r+1}e^{\mu_r s}p_{r+1}(s)\,ds\,. \tag{3}$$

Thus explicit expressions for the $p_r(t)$, $0 \leq r < n$ can be obtained. Since the actual formulae are quite complicated and do not seem to reveal much, we shall be content with giving the following partial result.

Proposition 4.1: *For each* $r > \frac{n}{2}$, $p_r(t)$ *is a linear combination of the functions* $e^{-\mu_i t}$, $r \leq i \leq n$.

Proof: Clearly the assertion is true for $r = n$. Let $n > r > \frac{n}{2}$ and $p_{r+1}(t) = \sum_{i=r+1}^{n} C_{r+1,i} e^{-\mu_i t}$. We shall show that $p_r(t) = \sum_{i=r}^{n} C_{r,i} e^{-\mu_i t}$. This will complete the proof. Using (3) and our hypothesis concerning $p_{r+1}(t)$,

$$p_r(t) = e^{-\mu_r t} \int_0^t \sum_{i=r+1}^{n} \mu_{r+1} C_{r+1,i} \quad e^{(\mu_r - \mu_i)s} \, ds \, .$$

Observe that $\mu_i = \mu_r$ if and only if $i = r$ or $i = n + 1 - r$. It follows that if $r > n/2$, then $\mu_i \neq \mu_r$ for every $i > r$. As a consequence,

$$\begin{aligned} p_r(t) &= \sum_{i=r+1}^{n} \mu_{r+1} \frac{C_{r+1,i}}{\mu_r - \mu_i} [e^{-\mu_i t} - e^{-\mu_r t}] \\ &= \sum_{i=r}^{n} C_{r,i} e^{-\mu_i t} \, , \end{aligned}$$

where

$$C_{r,r} = - \sum_{i=r+1}^{n} \frac{\mu_{r+1} C_{r+1,i}}{\mu_r - \mu_i}$$

and for $i > r$,

$$C_{r,i} = \frac{\mu_{r+1} C_{r+1,i}}{\mu_r - \mu_i} \, .$$

∎

Remark: The condition that $r > n/2$ was crucially used in the above proof. For $r \leq n/2$, $p_r(t)$ is not a linear combination of the above type. Extra complications crop up due to the fact that for any $r \leq n/2$ there is indeed an $i > r$, namely $i = n + 1 - r$, such that $\mu_i = \mu_r$. This gives rise to terms involving $te^{-\mu_r t}$ also, thus making explicit expression for $p_r(t)$ more complicated.

4.3 General Epidemic

We now consider a slight generalization of the earlier model in that, we allow removals. Of course, there is still no latent period. Thus a person infected becomes infectious instantly and remains so until he is removed from the population. This phenomenon is referred to as *General Epidemic*. As in the case of simple epidemic, here also we first consider a deterministic model and then a stochastic one.

4.3.1 Deterministic Model

This deterministic model was proposed in 1927 by W. O. Kermack and A.G. McKendrick. Here is the model. We have a population of n individuals and initially a certain number of them are infected; the rest are naturally susceptibles. As time passes, new infections take place and also some infected individuals are removed. Thus at any point of time, the population consists of three groups of individuals — susceptibles, infected and removed. Let $x(t), y(t)$ and $z(t)$ denote the number of individuals in these three groups respectively at time t. Clearly $x(t) + y(t) + z(t) = n$, for all t. We assume that $x(0) = x_0 > 0, y(0) = y_0 > 0$ and $z(0) = 0$. We assume that the number of new infections in time interval $(t, t + \Delta t)$ depends only on the number of susceptibles and the number of infected at time t, but not on the number of individuals removed till time t. This stands to reason because new infections arise out of contacts between the susceptibles and the infected.

As in the simple epidemic model, we postulate that the actual number of new infections during $(t, t + \Delta t)$ is $\beta x(t)y(t)\Delta t$. The rationale behind this postulate has already been explained in Section 4.2.1. Regarding removals, we assume that the number of individuals removed during $(t, t+\Delta t)$ depends only on the number of infected at time t and neither on the number of susceptibles at time t nor on the number of individuals removed till time t. This last assumption may be a little unrealistic in some situations. For example, one can very well have a situation where the health authorities have limited resources and try to put a check on the new removals. However, for the sake of simplicity, we rule out such a possibility. Following the same idea as in the case of new infections, we postulate that the number of individuals removed during $(t, t + \Delta t)$ is $\gamma y(t)\Delta t$ where γ is again a positive constant like β. The constant β is still called the *infection rate*, while γ is called the *removal rate*. The quantity $\rho = \gamma/\beta$ will play an important role in our analysis and is usually referred to as the *relative removal rate*.

Our postulates above lead to the following differential equations :

$$\begin{aligned} x'(t) &= -\beta x(t)y(t), \\ y'(t) &= \beta x(t)y(t) - \gamma y(t), \\ z'(t) &= \gamma y(t), \end{aligned} \tag{4}$$

with the initial conditions $x(0) = x_0, y(0) = y_0, z(0) = 0$.

The Equations (4) are known as the *Kermack-McKendrick Equations*, or, simply as *KK Equations*. The Equations (4) reveal that the functions $x(t), y(t)$ and $z(t)$ have derivatives upto any order. For example, the differentiability of $x(t)$ and $y(t)$ implies the differentiability of x' in view of the first equation in (4).

The first equation of (4) implies that $x'(t) \leq 0$ for all t, so that $x(t)$ is a non-increasing function. Similarly, from the third equation, it follows that $z(t)$ is a non-decreasing function. We shall now solve for x in terms of z. Using the

third equation of (4) in the first, we get

$$x'(t) = -\frac{1}{\rho}\, z'(t)x(t)\,.$$

Let $t_0 = \sup\{t : x(t) > 0\}$. From the continuity of $x(t)$ and the fact that $x(0) = x_0 > 0$, we conclude that $t_0 > 0$. It could however be infinite. We shall now proceed to argue that t_0 is indeed infinite. First note that, by monotonicity of x, we have $x(t) > 0$ for all $t < t_0$. Therefore on the interval $(0, t_0)$

$$\frac{x'(t)}{x(t)} = -\frac{1}{\rho} z'(t), \quad \text{that is,} \quad \frac{d}{dt} \log x(t) = -\frac{1}{\rho} z'(t)\,.$$

This yields the simple solution

$$x(t) = x_0 e^{-z(t)/\rho} \quad \text{for} \quad t \in (0, t_0)\,.$$

If t_0 were finite then the continuity of x and z would imply that

$$x(t_0) = x_0 e^{-z(t_0)/\rho} > x_0 e^{-n/\rho} > 0\,.$$

But by definition of t_0, we must have $x(t_0) = 0$, if t_0 is finite. This contradiction shows that t_0 is indeed infinite. Thus we have

$$x(t) = x_0 e^{-z(t)/\rho} \quad \text{for all} \quad t \geq 0\,. \tag{5}$$

First observe that, x and z being monotone, both the limits $x_\infty = \lim_{t\to\infty} x(t)$ and $z_\infty = \lim_{t\to\infty} z(t)$ exist. Moreover, from (5), we get that $x_\infty = x_0 e^{-z_\infty/\rho}$. Clearly, $z_\infty \leq n$, so that, $x_\infty \geq x_0 e^{-n/\rho} > 0$. Since $x(t) + y(t) + z(t) = n$ for each t, it follows that $y_\infty = \lim_{t\to\infty} y(t)$ also exists.

We shall now show that $y_\infty = 0$. These have the following epidemiological interpretation. After a sufficiently long time has elapsed and a stable state is reached no infected individual remains in circulation and the population still retains a positive number of uninfected people. This, of course, is a consequence of the dynamics embodied in Equations (4). Reality may not always follow Equations (4). Turning to the second Equation in (4) let us rewrite it as

$$y'(t) = \beta y(t)\,[\,x(t) - \rho\,]\,.$$

If $x(t) \geq \rho$ for all t, then $y(t)$ would be non-decreasing throughout. In particular $y(t) \geq y_0 > 0$ for all t, so that, $z'(t) \geq \gamma y_0$ for all t. But this would mean that $z(t) \geq \gamma y_0 t$ for all t, contradicting the fact that $z(t)$ is bounded by n. Thus there exists a finite time point $t_1 \geq 0$, such that on $[t_1, \infty)$, $x(t) \leq \rho$. As a consequence, y is non-increasing on $[t_1, \infty)$ and, in particular, $y(t) \geq y_\infty$. From the third Equation in (4), it follows that if $t > t_1$, then $z(t) \geq z(t_1) + \gamma y_\infty (t - t_1)$. Now $z(t) \leq n$ for all t would force y_∞ to be zero.

Recalling the definition of t_1 in the above paragraph, it is clear that $t_1 = 0$ or $t_1 > 0$ according as $x_0 \leq \rho$ or $x_0 > \rho$. We now bring to the fore the fact that

these two cases are indeed different in terms of their epidemiological manifestations. Let us first discuss the case $x_0 \leq \rho$. In this case, as observed above, $y(t)$ is a non-increasing function throughout. This means that the number of infected in circulation keeps on decreasing as time passes. In epidemiological terms, one says that the epidemic never really builds up. This should not however be construed as saying that there are no new infections. This only means that the removal rate is sufficiently high compared to the infection rate, so as to keep the number of infected individuals in circulation going down.

The more interesting case is when $x_0 > \rho$. In this case, as observed above, $y(t)$ increases upto a positive time instant t_1 and then decreases. That is, initially the epidemic does build up and reaches a peak at time instant t_1, after which it gradually subsides. In this case, it is interesting to get an idea of the size of the epidemic. A good measure of the size could be $(x_0 - x_\infty)$, which is precisely the total number of individuals who got infected during the course of the epidemic. Of course, we know that

$$x_\infty = x_0\, e^{-z_\infty/\rho}\,.$$

Using the fact that $y_\infty = 0$, so that $z_\infty = n - x_\infty$, we obtain

$$x_\infty = x_0\, e^{-(n-x_\infty)/\rho},$$

or equivalently,

$$n - x_\infty = n - x_0\, e^{-(n-x_\infty)/\rho}\,.$$

Denoting $n - x_\infty$ by u and setting $g(u) = n - x_0 e^{-u/\rho}$, we get

$$u = g(u)\,. \tag{6}$$

Given n, x_0 and ρ, the above is an equation in u, whose solution would give us $n - x_\infty$ or equivalently x_∞. Let us first point out that the equation (6) has a unique positive solution. Indeed, setting $h(u) = g(u) - u$, we observe that

$$h'(u) = g'(u) - 1 = \frac{x_0}{\rho} e^{-u/\rho} - 1\,.$$

Let $u_0 = \rho \log \dfrac{x_0}{\rho}$. Since $x_0 > \rho$, we have $u_0 > 0$. Further, it is easy to see that $h'(u) > 0$ on $[0, u_0)$ and $h'(u) < 0$ on (u_0, ∞). Consequently, h is strictly increasing on $[0, u_0)$ and is strictly decreasing on (u_0, ∞). Noting that $h(0) = n - x_0 = y_0 > 0$ and $\lim\limits_{u\to\infty} h(u) = -\infty$, it can be easily concluded that $h(u) = 0$ has a unique positive solution or equivalently that (6) has a unique positive solution. However, computing the exact value of the solution is difficult, perhaps impossible. There is no standard method of solving equations of the form $u = g(u)$, where g is an exponential function. So the next best thing is to replace g by an approximating polynomial and solve the resulting equation to get an approximate solution. The classical finite Taylor expansion of g would be an ideal choice for such an approximation. Following Kermack

and McKendrick, we use the second order Taylor polynomial. More precisely, we replace $g(u) = n - x_0 e^{-(u/\rho)}$ by

$$n - x_0 \left[1 - \frac{u}{\rho} + \frac{u^2}{2\rho^2}\right] = (n - x_0) + x_0 \frac{u}{\rho} - x_0 \frac{u^2}{2\rho^2} ,$$

so that Equation (6) takes the form

$$u = (n - x_0) + x_0 \frac{u}{\rho} - \frac{x_0}{2\rho^2} u^2 .$$

or equivalently,

$$\frac{x_0}{2\rho^2} u^2 + \left(1 - \frac{x_0}{\rho}\right) u - y_0 = 0 . \tag{7}$$

In view of $x_0 > \rho$, this quadratic equation is easily seen to have only one positive solution.
However, if we also assume that y_0 is small enough and can be neglected from (7), we get a simple formula for this unique positive solution, namely,

$$u^* = \frac{2\rho}{x_0}(x_0 - \rho) .$$

We can utilize this simple form of u^* to get a quantitative idea of the spread of the epidemic. Noting that u^* was obtained as an approximation for $n - x_\infty$ and that $n = x_0 + y_0$ we have

$$\begin{aligned} x_\infty &\sim n - \frac{2\rho}{x_0}(x_0 - \rho) \\ &= x_0 - 2(x_0 - \rho)\frac{\rho}{x_0} + y_0 \\ &\geq x_0 - 2(x_0 - \rho) \qquad [\text{ since } x_0 > \rho, \quad y_0 > 0] \\ &= \rho - (x_0 - \rho), \end{aligned} \tag{8}$$

that is, approximately, $\rho - x_\infty \leq x_0 - \rho$.

Thus, we are lead to the following conclusion. If the initial number of susceptibles exceeds ρ, then the epidemic certainly builds up. However, after the epidemic has died out, the final number of susceptibles can go only as far below ρ as the initial number was above ρ. Noting that $x_0 - x_\infty$ gives the total number of new infections during the course of the epidemic, the above observation really says that this number is approximately no more than $2(x_0 - \rho)$. We summarize our observations in the following theorem:

Theorem 4.2:
(a) *We always have,*

$$\lim_{t\to\infty} y(t) = 0 \text{ and } \lim_{t\to\infty} x(t) \geq x_0\, e^{-(n/\rho)} > 0 .$$

(b) *If $x_0 \leq \rho$, then $y(t)$ continuously decreases in t. Thus, as long as $x_0 \leq \rho$,*

the epidemic does not build up.
(c) *If $x_0 > \rho$, then $y(t)$ initially increases with t, reaches a peak and then gradually decreases. Thus the epidemic does build up. If it is further assumed that y_0 is negligible, then $2(x_0 - \rho)$ is an approximate upper bound for the number of people infected in the course of the epidemic.*

Thus, ρ acts as a threshold value for the initial number of susceptibles in order for the epidemic to build up or not. For this reason, parts (b) and (c) of Theorem 4.2 are referred to as the *Kermack-McKendrick Threshold Theorem.* Going back to (8), it is clear that if y_0 is negligibly small and x_0 is only marginally above the threshold value ρ, so that $\frac{x_0}{\rho} \sim 1$, then one can safely say that $x_\infty \sim \rho - (x_0 - \rho)$. Very often the existing literature states this approximate equality as part of the threshold theorem rather than the approximate inequality we stated in part (c). This is alright as long as the assumptions $y_0 \sim 0, \frac{x_0}{\rho} \sim 1$ are kept in mind.

We now turn to the assumption that y_0 is small. This amounts to saying that initially there is only a trace of the infection in the population. This is not altogether unjustified — and in fact quite natural — for the following reason. In the study of continuous time epidemic model, it is only natural to take the time origin as the time point when the infection first surfaced in the population. Granted that, the assumption of y_0 being small is only logical because most infections start by traces.

Before ending the section, we would like to take up a curious point. One of the key steps in getting an approximate solution of the equation $u = g(u)$ was to replace $g(u)$ by an appropriate Taylor polynomial. In deriving the Threshold Theorem, the second order polynomial was used. The natural question is: why not start with the first order polynomial? Here is an argument. Using the first order polynomial would lead to the equation $u = y_0 + \frac{x_0}{\rho}u$. In case $y_0 \sim 0$ as we have been assuming throughout, this equation almost reduces to $u = \frac{x_0}{\rho}u$. This is of course no good. If $\frac{x_0}{\rho} \not\sim 1$, this equation admits no solution other than zero, whereas if $\frac{x_0}{\rho} \sim 1$, we end up with too many solutions ! Going in the other direction, it may be worthwhile to try and see what one obtains by approximating $g(u)$ by a third order polynomial.

4.3.2 General Stochastic Epidemic

We start with a population consisting initially of a susceptibles and b infected persons. For any time instant t, X_t will denote the number of susceptibles at time t, Y_t the number of infected in circulation at time t and Z_t the number of persons removed till time t. We shall assume now that the spread of the epidemic is governed by a chance mechanism, so that X_t, Y_t and Z_t are random variables. Our object of study is the evolution of the process $(X_t, Y_t, Z_t)_{t \geq 0}$.

The initial conditions, as stated already, are $X_0 = a, Y_0 = b$ and $Z_0 = 0$. It is clear that, for all t, $X_t + Y_t + Z_t = a + b$. Thus, studying the two-dimensional process $(X_t, Y_t)_{t\geq 0}$ suffices. We now describe the chance mechanism. The idea is the same as in the case of simple stochastic epidemic. Given $X_t = x$ and $Y_t = y$, we assume that during a small time interval $(t, t + \Delta t)$, there will be one new infection with probability $\beta xy\Delta t + o(\Delta t)$ and no new infection with probability $1 - \beta xy\Delta t + o(\Delta t)$. This, of course, means that the probability of two or more new infections during $(t, t+\Delta t)$ is $o(\Delta t)$. Regarding removals, our assumption is that during the same time interval, there will be one removal with probability $\gamma y\Delta t + o(\Delta t)$ and no removals with probability $1 - \gamma y\Delta t + o(\Delta t)$. Further, the two events, namely, that of infection and that of a removal during such small time intervals, are assumed to be independent. This description clearly entails that $(X_t, Y_t)_{t\geq 0}$ is a bivariate continuous time Markov chain with state space

$$S = \{(r, s) : r, s \text{ non-negative integers}, r \leq a, r + s \leq a + b\}.$$

It is also clear, by considering the embedded discrete chain, that this is an absorbing chain with the states $\{(r, 0) : 0 \leq r \leq a\}$ being the absorbing states and all others transient. In particular, $Y_\infty = \lim_{t\to\infty} Y_t = 0$ with probability 1, and $X_\infty = \lim_{t\to\infty} X_t$ exists. In line with the deterministic case, the random variable $(X_0 - X_\infty)$ would denote the size of the epidemic. For $(r, s) \in S$, let

$$p_{r,s}(t) = P(X_t = r \text{ and } Y_t = s).$$

It is convenient to have $p_{r,s}$ defined for $(r, s) \notin S$ also, by simply adopting the convention that, for $(r, s) \notin S$, $p_{r,s}(t) = 0$ for all t. We then have

$$\begin{aligned} p_{r,s}(t + \Delta t) &= p_{r+1,s-1}(t)\beta(r + 1)(s - 1)\Delta t[1 - \gamma(s - 1)\Delta t] \\ &\quad + p_{r,s+1}(t)\gamma(s + 1)\Delta t[1 - \beta r(s + 1)\Delta t] \\ &\quad + p_{r,s}(t)[1 - \beta rs\Delta t][1 - \gamma s\Delta t] \\ &\quad + p_{r+1,s}(t)[\beta(r + 1)s\Delta t][\gamma s\Delta t] + o(\Delta t). \end{aligned}$$

This gives us the Kolmogorov equations

$$\frac{dp_{r,s}(t)}{dt} = \beta(r+1)(s-1)p_{r+1,s-1}(t) + \gamma(s+1)p_{r,s+1}(t) - s(\beta r + \gamma)p_{r,s}(t). \quad (9)$$

In particular,

$$\frac{dp_{a,b}(t)}{dt} = -b(\beta a + \gamma)p_{a,b}(t).$$

Using the initial condition $p_{a,b}(0) = 1$, we get

$$p_{a,b}(t) = e^{-b(\beta a+\gamma)t}.$$

For $r = a, s = b - 1$, the Equation (9) becomes

$$\frac{dp_{a,b-1}(t)}{dt} = \gamma b p_{a,b}(t) - (b - 1)(\beta a + \gamma)p_{a,b-1}(t),$$

or equivalently,

$$\frac{dp_{a,b-1}(t)}{dt} + (b-1)(\beta a + \gamma)p_{a,b-1}(t) = \gamma b e^{-b(\beta a+\gamma)t}.$$

Using the initial condition $p_{a,b-1}(0) = 0$, the solution can easily be seen to be

$$p_{a,b-1}(t) = \frac{\gamma b}{\beta a + \gamma} e^{-(b-1)(\beta a+\gamma)t}[1 - e^{-(\beta a+\gamma)t}].$$

Thus the Kolmogorov equations (9) can be successively solved, using the initial condition $p_{r,s}(0) = 0$ for $(r,s) \neq (a,b)$, to get $p_{r,s}(t)$ for all r, s. However, explicit formulae turn out to be extremely complicated and fail to give any insight into the state of affairs. Nevertheless, several people had attempted to get $p_{r,s}(t)$ explicitly by different methods. We briefly illustrate two such attempts here. For further details the reader may consult the book of Bailey.

The first one, due to Siskind, converts the system of differential equations (9) to a single partial differential equation. The idea is to look at the joint p.g.f. of (X_t, Y_t) defined as

$$F(t,u,v) = \sum_{r,s} p_{r,s}(t) u^r v^s .$$

The equations (9) lead to the following partial differential equation for F:

$$\frac{\partial F}{dt} = \beta(v^2 - uv)\frac{\partial^2 F}{\partial u \partial v} + \gamma(1-v)\frac{\partial F}{\partial v},$$

with the initial condition

$$F(0,u,v) = u^a v^b .$$

Siskind solved this explicitly and derived formulae for the functions $p_{r,s}(t)$.

The second one, due to Gani, looks at the Laplace transforms of the functions $p_{r,s}(t)$. Recall that for any bounded continuous function $p(t)$ on $[0,\infty)$, its *Laplace Transform* is the function $q(\lambda)$ on $(0,\infty)$, defined by

$$q(\lambda) = \int_0^\infty e^{-\lambda t} p(t)\, dt$$

Recall further that q determines p uniquely. The idea of Gani in considering the Laplace transforms $q_{r,s}(\lambda)$ of $p_{r,s}(t)$ was to convert the system of equations (9) into a system of recurrence relations for the functions $q_{r,s}$ given by,

$$(\lambda + \beta rs + \gamma s)q_{r,s} = \beta(r+1)(s-1)q_{r+1,s-1} + \gamma(s+1)q_{r,s+1}, \qquad (10)$$

for $(r,s) \neq (a,b)$, with the initial condition

$$(\lambda + \beta ab + \gamma b)q_{a,b} \equiv 1 .$$

These recurrence relations may not be difficult to solve. But in order to get back to the functions $p_{r,s}$, one still faces the problem of inverting the Laplace transforms $q_{r,s}$, which is a difficult one. However, there is one probabilistic question that can be answered without having to go for inversion. For example, what is the probability that the total size of the epidemic, not including the initial number of infected, is k for a given $0 < k < a$? In other words, we are interested in the quantity $P_k = \lim_{t\to\infty} p_{a-k,0}(t)$. Under suitable conditions, it is easy to verify that, if $q(\lambda)$ is the Laplace transform of $p(t)$, then

$$\lim_{t\to\infty} p(t) = \lim_{\lambda\to 0} \lambda q(\lambda)\,.$$

Thus, $P_k = \lim_{\lambda\to 0} \lambda\, q_{a-k,0}(\lambda)$. Using the recurrence relations (10), one can reduce this to

$$P_k = \gamma \lim_{\lambda\to 0} q_{a-k,1}(\lambda)\,.$$

Thus, knowledge of $q_{a-k,1}$ would give us P_k for each k.

4.3.3 A Closer Analysis: Threshold Theorems

We now turn to what are regarded as two fundamental theorems in Markov models for epidemics — the so called *"Threshold Theorems"*. Two quantities that are of interest in understanding the extent of the epidemic are its duration and size. To make the definitions of these quantities precise, let us turn to the Markov process $(X_t, Y_t)_{t\geq 0}$ and observe the following salient features. Recall that the process has a finite state space given by

$$S = \{(r,s) : r, s \text{ non-negative integers }\,;\; r \leq a;\; r+s \leq a+b\}\,.$$

As mentioned earlier, the states $\{(r,0) : r \leq a\}$ are precisely the absorbing states for the process. Also, the set of states $\{(0,s) : s \leq a+b\}$ forms a closed set and once the process hits this set, it then evolves like a death chain in the second coordinate, getting ultimately absorbed at $(0,0)$. Let us denote by F the union of the above two sets of states. Clearly, with probability one, the chain hits the set F in a finite amount of time. Once the set F is entered, no new infections are possible and therefore the epidemic can be thought of as having ended for all practical purposes. It is natural therefore, to regard the time τ needed to enter F as the duration of the epidemic. One would like to draw conclusions about the distribution, and in particular, the expected value of τ. To the best of our knowledge, the existing literature does not contain any non-trivial information on this. On the contrary, much emphasis has been given to what may be called the *size of the epidemic.* From the definition of τ, it is clear that $X(\tau)$ gives the eventual number of susceptibles left in the population. In other words, $X(\tau) = \lim_{t\to\infty} X(t)$. The size of the epidemic is clearly given by the random variable

$$W = X(0) - \lim_{t\to\infty} X(t) = a - X(\tau)\,.$$

One would like to obtain the distribution of the random variable W. Following the general theory of finite state Markov processes, the process (X_t, Y_t) evolves as follows. Given that at time instant t the process is in a non-absorbing state, it waits there for an exponential time and then makes a jump. From a non-absorbing state (r, s), jumps are possible to the states $(r-1, s+1)$ and $(r, s-1)$ with probabilities $p_r = \beta rs/(\beta rs + \gamma s) = r/(r+\rho)$ and $q_r = \rho/(r+\rho)$ respectively. Here, ρ is as defined in the deterministic case, namely, $\rho = \gamma/\beta$. A moment's reflection shows that while τ is the sum of all these waiting times starting from the beginning till the chain hits F, the random variable W has nothing to do with the waiting times. In order to get the distribution of W, it is therefore sufficient to keep track of only the states visited by the chain at its successive jumps. In other words, the distribution of W depends on (X_t, Y_t) only through the embedded Markov chain, as discussed in Section 0.9.1. Let us note that the embedded chain here can be described as follows.

Let $\tau_0 \equiv 0$ and $(\tau_n, n \geq 1)$ be the successive jump times of the process (X_t, Y_t). For $n \geq 0$, let $U_n = X_{\tau_n}$ and $V_n = Y_{\tau_n}$. Then $(U_n, V_n)_{n \geq 0}$ is the embedded chain with state space S. The transition probabilities are as follows. From a state (r, s) with $s \neq 0$, transition can take place to $(r-1, s+1)$ with probability $p_r = r/(r+\rho)$ and to $(r, s-1)$ with probability $q_r = \rho/(r+\rho)$. The states $(r, 0)$ are absorbing. Now note that, $\lim_n U_n = \lim_{t \to \infty} X_t$, so that, $W = a - \lim_n U_n$. This is precisely what we meant, when we said earlier that W depends only on the embedded chain. This was observed by Foster [1955] and was beautifully exploited by him and later, by Rajarshi [1981] to get the exact distribution of the random variable W.

We start with some notation. For a non-negative integer w let A_w denote the set of all sequences $\widetilde{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_w)$ of length $w+1$, where the α_i are non-negative integers satisfying

i) $\alpha_w > 1$,

ii) for $j < w$, $\alpha_0 + \alpha_1 + \ldots + \alpha_j < b + j$,

iii) $\alpha_0 + \alpha_1 + \ldots + \alpha_w = b + w$.

Theorem 4.3 (Foster): *For* $0 \leq w \leq a$,

$$P(W = w) = \prod_{l=0}^{w-1} p_{a-l} \sum_{\widetilde{\alpha} \in A_w} \prod_{j=0}^{w} q_{a-j}^{\alpha_j} .$$

Proof: We can view the state space S of the chain $(U_n, V_n)_{n \geq 0}$ as the set of lattice points (r, s) — that is, points with integer coordinates — in the xy plane. The evolution of the chain can then be regarded as the motion of a particle through these lattice points. From a point (r, s) with $s > 0$ the particle moves either one step vertically downwards to $(r, s-1)$ or one step diagonally northwest (that is, up and left) to $(r-1, s+1)$. The probabilities of these two types of transitions are q_r and p_r respectively. Once the particle hits the x-axis it halts. On the other hand once it hits the y-axis, then only the vertically

downward transitions are allowed until it reaches the origin. Viewed this way, the event $(W = w)$ means that the particle starting from (a, b) hits the x-axis precisely at the point $(a - w, 0)$.
Observe that for this to happen, the particle has to make exactly w many northwest transitions, with the first coordinate reducing by 1 after each such transition, until it finally becomes $a - w$. Note that, each of these transitions would result in an increase in the second coordinate by 1, so that the particle has to make $b + w$ many vertically downward transitions in order to reach the x-axis. Starting from (a, b), denote by α_0, the number of downward transitions before the first diagonal transition. In general, for $1 \leq j < w$, let α_j be the number of downward transitions between the j-th and $(j+1)$th diagonal steps. Finally α_w is the number of downward transitions after the w-th diagonal step. Clearly α_j, for $0 \leq j \leq w$, are non-negative integers. Moreover, their sum is $b + w$. The fact that the particle does not hit x-axis before making the w-th diagonal transition implies that, for each $j < w$, $\alpha_0 + \alpha_1 + \cdots + \alpha_j < b + j$. In particular, $\alpha_0 + \alpha_1 + \cdots + \alpha_{w-1} < b + w - 1$, implying that $\alpha_w > 1$. Thus the sequence $\widetilde{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_w) \in A_w$. Conversely, any $\widetilde{\alpha} \in A_w$ is a possible choice for the number of vertical motions in between the successive diagonal ones, so that the required event $(W = w)$ occurs. For any $\widetilde{\alpha} \in A_w$, the probability of making the transitions as prescribed by $\widetilde{\alpha}$ is $\prod_{l=0}^{w-1} p_{a-l} \prod_{j=0}^{w} q_{a-j}^{\alpha_j}$. Here the fact that the transition probabilities from any state (r, s) depend only on the first coordinate r is important. The proof is now complete. ∎

For an estimate of the above probability later, we need the following lemma, as in Rajarshi [1981].

Lemma 4.4: *The number of elements in the set A_w is* $\dfrac{b}{b+2w}\binom{b+2w}{b+w}$.

Proof: From Chapter 0.8.1, $\dfrac{b}{b+2w}\binom{b+2w}{b+w}$ is precisely the number of paths of a random walk starting at $(0, 0)$ and ending at $(b + 2w, b)$, which lie strictly above the horizontal axis. The proof will be completed by establishing a one-one correspondence between such paths and elements of A_w. Here is the correspondence. Consider such a path of random walk. First of all, the path would have w many downward steps and $b + w$ many upward steps. Let α_0 be the number of upward steps after the w-th, that is, the final downward transition. Let α_1 be the number of upward steps between the $(w - 1)$th and w-th downward motions. In general, α_j will denote the number of upward steps between the $(w - j)$th and $(w - j + 1)$th downward motions. Finally, α_w is the number of upward steps before the first downward transition. This defines a sequence $\widetilde{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_w)$ of non-negative integers of length $w + 1$. We now show that $\widetilde{\alpha} \in A_w$. Since $\alpha_0 + \alpha_1 + \cdots + \alpha_w$ gives the total number of upward movements, it is clear that this sum is $b + w$. Now let $0 \leq j < w$. To show that $\alpha_0 + \alpha_1 + \cdots + \alpha_j < b + j$, it suffices to prove that $\alpha_{j+1} + \cdots + \alpha_w > w - j$. From the definition of the α_j, it should be clear that $\alpha_{j+1} + \cdots + \alpha_w$ is pre-

cisely the total number of upward motions before the $(w-j)$th downward motion. Since the path lies strictly above the horizontal axis we must have $\alpha_{j+1} + \cdots + \alpha_w > w - j$. In particular, $j = w - 1$ gives $\alpha_w > 1$. In fact, this is also directly obvious because α_w is the number of upward steps before the first downward step. Conversely, given $\widetilde{\alpha} \in A_w$, consider the path which makes α_w many upward transitions starting from $(0,0)$, then makes a downward transition followed by α_{w-1} many upward transitions, and so on. It is easy to see that this gives a path of the required type for the random walk. ∎

The above two results lead to the following simple estimate of the distribution of W, for large values of a. For stating this we need the following notation. For two functions $h(a)$ and $g(a)$, we will write $h(a) \sim g(a)$ as $a \to \infty$ to mean that $\lim_{a\to\infty}[h(a)/g(a)] = 1$. It is easy to see that, $h_1(a) \sim g_1(a)$ and $h_2(a) \sim g_2(a)$ as $a \to \infty$ imply that $h_1(a) + h_2(a) \sim g_1(a) + g_2(a)$ as $a \to \infty$.

Lemma 4.5: *For each* $w \geq 0$,

$$P(W = w) \sim \frac{b}{2w+b}\binom{2w+b}{w+b} p_a^w q_a^{b+w} \qquad as \quad a \to \infty\,.$$

Proof: For each $l = 0, 1, \ldots, w-1$,

$$\frac{p_{a-l}}{p_a} = \frac{a-l}{a-l+\rho}.\frac{a+\rho}{a} \to 1 \quad \text{as } a \to \infty\,.$$

For each $j = 0, 1, \ldots, w$,

$$\frac{q_{a-j}}{q_a} = \frac{\rho}{a-j+\rho}.\frac{a+\rho}{\rho} \to 1 \quad \text{as } a \to \infty\,.$$

It follows that for a fixed w and an $\widetilde{\alpha} \in A_w$ we have,

$$\prod_{l=0}^{w-1} p_{a-l} \prod_{j=0}^{w} q_{a-j}^{\alpha_j} \sim p_a^w q_a^{b+w}\,.$$

By summing over $\widetilde{\alpha} \in A_w$ and using Theorem 4.3, we get

$$P(W = w) \sim |\, A_w \,|\, p_a^w q_a^{b+w} \quad \text{as} \quad a \to \infty,$$

whence the assertion follows by using Lemma 4.4. ∎

We now present two theorems due to Williams (1971) and Whittle (1955), known as the *Threshold Theorems* for the General Stochastic Epidemic. The common theme of both the theorems is to identify ρ as a threshold quantity to determine whether the epidemic builds up or not. The following lemma will play a crucial role in the proofs of the threshold theorems. This is due to Williams. However, his proof is non-probabilistic and uses certain power series expansion. We give a probabilistic proof that uses transience of random walk.

Lemma 4.6: *For* $0 \leq p < 1$ *and* $q = 1 - p$,

$$\sum_{w=0}^{\infty} \frac{b}{2w+b}\binom{2w+b}{w+b} q^w p^{b+w} = \min\left\{\left(\frac{p}{q}\right)^b, 1\right\}.$$

Proof: It suffices to show that if $p \geq q$, then

$$\sum_{w=0}^{\infty} \frac{b}{2w+b}\binom{2w+b}{w+b} q^w p^{b+w} = 1.$$

First recall that $\frac{b}{2w+b}\binom{2w+b}{w+b}$ is precisely the number of paths of a random walk starting at the origin and reaching the state b in $(2w+b)$ many steps without hitting the x-axis. But by reversing the motion, this would be the same as the number of paths starting at state b and reaching the origin for the first time in $(2w+b)$ many steps, or equivalently (by shifting the x-axis), starting at the origin and reaching state $-b$ for the first time in $(2w+b)$ many steps. Thus if we consider a random walk with probability q (p respectively) of upward (downward respectively) transitions, then the summand is just the probability that such a random walk starting from the origin reaches $-b$ for the first time in $(2w+b)$ many steps. Since for different w, the above events are disjoint, summing over w gives us the probability that such a random walk starting from the origin ever reaches $-b$. By Exercise 2(vi) of Section 0.8.1, this probability is one whenever $q \leq p$ (Do not forget that here q is the probability of upward transition). ∎

Let us now present the two threshold theorems mentioned above in the way that we understand them. Both the threshold theorems are statements concerning the extent of the epidemic for large values of the initial number of susceptibles a. To be precise they both talk about the limiting probabilities as $a \to \infty$. Since we are varying a, it makes sense to allow ρ also to possibly vary with a. This dependence is going to be made explicit by using the notation ρ_a. In what follows we consider limits of certain probabilites as the pair (a, ρ_a) varies in such a way that $a \to \infty$ and ρ_a/a converges to a limit, say, δ. Since a is varying, the probabilities associated to the corresponding epidemic model will also vary. We will make it explicit by writing P_a for such probabilities. Note that we are keeping b, the initial number of infected individuals, fixed.

Theorem 4.7 (Williams' Threshold Theorem): *If* (a, ρ_a) *vary in such a way that* $a \to \infty$ *and* $\frac{a}{\rho_a} \to \delta$, *then*

$$\lim_{M\to\infty} \lim_{a\to\infty} P_a(W \leq M) \quad = \quad \min(\delta^b, 1).$$

Proof: By Theorem 4.5, for each w,

$$\lim_{a\to\infty} P_a(W = w) = \frac{b}{2w+b}\binom{2w+b}{w+b}\left(\frac{1}{1+\delta}\right)^w \left(\frac{\delta}{1+\delta}\right)^{b+w},$$

so that for any M,

$$\lim_{a\to\infty} P_a(W \le M) = \sum_{w=0}^{M} \frac{b}{2w+b}\binom{2w+b}{w+b}\left(\frac{1}{1+\delta}\right)^w\left(\frac{\delta}{1+\delta}\right)^{b+w}.$$

Lemma 4.6 now completes the proof. ∎

The quantity $\lim_{a\to\infty} P_a(W \le M)$ can be interpreted as the probability of having an epidemic of size at most M, for large values of a. Taking now the limit of that probability as $M \to \infty$ could therefore have the interpretation of being the probability of a finite epidemic for large values of a. Thus, Theorem 4.7 says that, if $\delta \ge 1$ then the epidemic is surely of finite size, while for $\delta < 1$ the probability of a finite epidemic is δ^b, which is strictly smaller than one. Indeed, this is how the Threshold Theorem is stated in the literature. There is one little subtlety, namely, instead of stating the result in terms of $\delta = \lim_a(\rho_a/a)$, the standard practice is to simply say that for large values of a, the probability of a finite epidemic equals one if $\rho_a \ge a$, while it equals $(\rho_a/a)^b$ if $\rho_a < a$. This is what Theorem 4.7 may be argued to say, provided the inequalities and equalities are all interpreted properly. For example, $\rho_a < a$ should be interpreted as $\lim_a(\rho_a/a) < 1$ (or more generally, for the present purpose, $\limsup_a(\rho_a/a) < 1$).

We shall now proceed to Whittle's Threshold Theorem. This deals with the probability, for large values of a, of the epidemic not exceeding a certain proportion of the initial number of susceptibles a. More specifically, for fixed x, $0 < x < 1$, we consider the probability $P_a(W \le xa)$. Whittle's Threshold Theorem attempts to get two-sided bounds for these probabilities, at least for large values of a.

Getting an asymptotic lower bound is not difficult. Let us assume as before that the parameters (a, ρ_a) vary in such a way that $a \to \infty$ and $(\rho_a/a) \to \delta$. Denoting $\pi_x^a = P_a(W \le xa)$, we show that

$$\lim_{a\to\infty} \pi_x^a \quad \ge \quad \min\left\{\delta^b, 1\right\}. \tag{11}$$

To see this, fix any n and observe that for large a, we have $xa > n$, so that

$$\pi_x^a \quad \ge \quad \sum_{w=0}^{n} P_a(W = w).$$

Using Lemma 4.5,

$$\lim_{a\to\infty} \pi_x^a \quad \ge \quad \sum_{w=0}^{n} \frac{b}{2w+b}\binom{2w+b}{w+b}\left(\frac{1}{1+\delta}\right)^w\left(\frac{\delta}{1+\delta}\right)^{b+w}.$$

The above inequality being true for all n, Lemma 4.6 yields the inequality (11).

Inequality (11) constitutes only one half of Whittle's Threshold Theorem and is often stated as

$$P(W \leq xa) \quad \geq \quad \min\left\{\left(\frac{\rho}{a}\right)^b, 1\right\} \quad \text{for large values of } a\,.$$

This is true as long as it is properly interpreted as discussed after the statement of Theorem 4.7 above.

The other half of Whittle's Theorem, which seeks an upper bound, is based on the use of a *comparison technique* which is interesting in its own right and is described below.

Our epidemic process (X_t, Y_t) is a Markov process starting from (a, b) and having transition mechanism determined by the parameters p_r. Consider now another chain evolving in the same manner, but with a different transition mechanism determined by parameters p'_r. To avoid complication, we use the same notation (X_t, Y_t) for this new process also. The difference in transition mechanism is indicated by using P' for probabilities of the new chain. If $p_r \geq p'_r$ for every r, then it is natural (why?) to expect that the random variable X_τ is stochastically larger under P' than under P, that is,

$$P'(X_\tau \geq k) \quad \geq \quad P(X_\tau \geq k) \quad \text{for each } k\,. \tag{12}$$

We shall show that this indeed is the case. But, for the present, we assume this and proceed to complete the remaining half of Whittle's theorem.
Recall that the parameters for our epidemic process are defined as

$$p_r \quad = \quad \frac{r}{r+\rho_a} \quad \text{for} \quad 0 \leq r \leq a\,.$$

Let us define p'_r for $0 \leq r \leq a$, as

$$\begin{aligned} p'_r \quad &= \quad p_r \qquad\qquad\quad \text{for} \quad r < (1-x)a \\ &= \quad \frac{(1-x)a}{(1-x)a+\rho_a} \qquad \text{for} \quad r \geq (1-x)a\,. \end{aligned}$$

Clearly, for every r, $p_r \geq p'_r$, so that by (12)

$$P(X_\tau \geq (1-x)a) \quad \leq \quad P'(X_\tau \geq (1-x)a)\,.$$

In view of the fact that $W = a - X_\tau$, the above inequality is the same as

$$P(W \leq xa) \quad \leq \quad P'(W \leq xa)\,.$$

Invoking the arguments of Theorem 4.3 and using Lemma 4.4, the right hand side can easily be seen to equal

$$\sum_{w=0}^{xa} \frac{b}{2w+b}\binom{2w+b}{w+b}\left(\frac{(1-x)a}{(1-x)a+\rho_a}\right)^w\left(\frac{\rho_a}{(1-x)a+\rho_a}\right)^{b+w},$$

which by Lemma 4.6 is clearly bounded above by $\min\left\{\left(\frac{\rho_a}{(1-x)a}\right)^b, 1\right\}$. We conclude that

$$P(W \leq xa) \quad \leq \quad \min\left\{\left(\frac{\rho_a}{(1-x)a}\right)^b, 1\right\}.$$

We have thus proved

Theorem 4.8 (Whittle's Threshold Theorem): *For any x, $0 < x < 1$, and for large values of a,*

$$\min\left\{\left(\frac{\rho_a}{a}\right)^b, 1\right\} \quad \leq P(W \leq xa) \leq \quad \min\left\{\left(\frac{\rho_a}{(1-x)a}\right)^b, 1\right\}.$$

Although we have stated the result in the way it is usually done, the reader should note that, the right hand side inequality is actually valid for all a, whereas the left hand side is valid only in the limit, that is, in the sense discussed earlier. It may be noted that a comparison technique, similar to the one used above, can be used also to get a lower bound valid for all a. Indeed, one can show (left as an exercise) that

$$P(W \leq xa) \quad \geq \quad \sum_{w=0}^{xa} \frac{b}{2w+b}\binom{2w+b}{w+b}\left(\frac{a}{a+\rho_a}\right)^w\left(\frac{\rho_a}{a+\rho_a}\right)^{b+w}$$

for all a and all x, $0 < x < 1$. It is tempting to claim that the right hand side of the above inequality is approximately $\min\left\{\left(\frac{\rho_a}{a}\right)^b, 1\right\}$, for large a, in view of Lemma 4.6. One may then erraneously claim that $\min\left\{\left(\frac{\rho_a}{a}\right)^b, 1\right\}$ is an actual lower bound for $P(W \leq xa)$ for all large a. In fact, the standard literature seems to make that claim. We wish we could justify this, thus avoiding interpretation through limits.

We now get back to our claim (12). Since comparison technique is an important and useful technique in the context of Markov chains, we will prove a slightly more general result. First, let us introduce some notation.

For any pair of integers $a \geq 1$, $b \geq 1$, and any a-tuple $\theta = (\theta_1, \ldots, \theta_a)$, with $0 \leq \theta_r \leq 1$ for all r, let P_θ denote the probability law of the Markov chain starting from (a, b) having state space

$$S = \{(r, s) : r, s \text{ non-negative integers}; \quad r \leq a; \quad r + s \leq a + b\}$$

and evolving in the following manner. State (r, s) is absorbing unless both r and s are strictly positive. From a non-absorbing state (r, s), the chain moves to $(r-1, s+1)$ with probability θ_r and moves to $(r, s-1)$ with probability $1 - \theta_r$. We will denote this process by $(U_n^{a,b}, V_n^{a,b})$. It is left as an exercise for

the reader to verify that we have a Markov chain on a finite state space for which every non-absorbing state is transient and hence it is an absorbing chain. Let T be the time till absorption. Thus $P_\theta(T < \infty) = 1$. Our objective would be to get a stochastic comparison of the random variable $U_T^{a,b}$, to be denoted by $Z^{a,b}$, for various a-tuples θ. The relevance of this in our context stems from the fact that, with $\theta_r = r/(r+\rho)$ for $1 \leq r \leq a$, the chain (U_n, V_n) is just the embedded chain associated with our epidemic process (X_t, Y_t) stopped at time τ. In particular the random variable X_τ and $Z^{a,b}$ are identical. We want to prove

Theorem 4.9: *If θ and θ' are two a-tuples with $\theta'_r \geq \theta_r$ for all r, then $Z^{a,b}$ is stochastically larger under P_θ than under $P_{\theta'}$, that is, for all k,*

$$P_\theta(Z^{a,b} \geq k) \quad \geq \quad P_{\theta'}(Z^{a,b} \geq k)\,.$$

To prove the theorem, we need a series of lemmas.

Lemma 4.10: *Let $w_1, \ldots, w_n$ and $v_1, \ldots, v_n$ be non-negative numbers such that for $1 \leq j \leq n$, $\sum_{i \leq j} w_i \leq \sum_{i \leq j} v_i$. Then, for any sequence of numbers $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$, one has $\sum_{i \leq n} c_i w_i \leq \sum_{i \leq n} c_i v_i$.*

Proof: Note that the hypothesis implies that for each $j = 1, \ldots, n-1$, the inequality

$$(c_j - c_{j+1}) \sum_{i \leq j} w_i \leq (c_j - c_{j+1}) \sum_{i \leq j} v_i$$

holds. Putting $c_{n+1} = 0$, the same inequality is seen to hold for $j = n$ also. Adding these n inequalities yields the desired result. ∎

Lemma 4.11: *For any θ, the probability $P_\theta(Z^{a,b} \leq k)$ is non-decreasing in b.*

Proof: We shall show that $P_\theta(Z^{a,b+1} \leq k) \geq P_\theta(Z^{a,b} \leq k)$. Suppose that $s \leq k$ and α is a path from (a, b) hitting the x-axis for the first time at $(s, 0)$. Let α^* be the path obtained by adding one to the second co-ordinate of all points of the path α. Clearly α^* is a path from $(a, b+1)$ and hitting the horizontal line $y = 1$ for the first time at the point $(s, 1)$. Let η be the hitting time of the line $y = 1$. The correspondence $\alpha \longleftrightarrow \alpha^*$ and the fact that the two paths α, α^* have the same probabilities (because the transition probabilities from any state depend only on the first coordinate of the state and we have not disturbed the first coordinates of points in α to get α^*) can be put together to deduce that $P_\theta(U_T^{a,b} \leq k) = P_\theta(U_\eta^{a,b+1} \leq k)$. However from the dynamics of the process it is clear that the event $(U_\eta^{a,b+1} \leq k)$ implies $(U_T^{a,b+1} \leq k)$. It now follows that $P_\theta(U_T^{a,b+1} \leq k) \geq P_\theta(U_T^{a,b} \leq k)$, as was to be shown. ∎

Lemma 4.12: *Let θ and θ' be two a-tuples such that $\theta'_a \geq \theta_a$, while $\theta'_r = \theta_r$, for all $r < a$. Then $Z^{a,b}$ is stochastically larger under P_θ than under $P_{\theta'}$.*

Proof: Let $k \leq a - 1$. We prove

$$P_\theta(Z^{a,b} \leq k) \quad \leq \quad P_{\theta'}(Z^{a,b} \leq k)\,.$$

Let η be the hitting time of the vertical line $x = a - 1$. Note that the event $(Z^{a,b} \leq k)$ implies that $\eta < \infty$. Indeed, $\eta \leq b$ and hence

$$P_\theta(Z^{a,b} \leq k) \quad = \quad \sum_{i=1}^{b} P_\theta(Z^{a,b} \leq k \,|\, \eta = i) P_\theta(\eta = i)\,.$$

Using the Markov property, the conditional probability $P_\theta(Z^{a,b} \leq k \,|\, \eta = i)$ is the same as the probability $P_\theta(Z^{a-1,b-i+2} \leq k)$, so that

$$P_\theta(Z^{a,b} \leq k) = \sum_i P_\theta(Z^{a-1,b-i+2} \leq k) P_\theta(\eta = i)\,.$$

Analogously,

$$P_{\theta'}(Z^{a,b} \leq k) = \sum_i P_{\theta'}(Z^{a-1,b-i+2} \leq k) P_{\theta'}(\eta = i)\,.$$

Since $\theta'_r = \theta_r$, for $r \leq a - 1$, it is clear that for every i,

$$P_\theta(Z^{a-1,b-i+2} \leq k) = P_{\theta'}(Z^{a-1,b-i+2} \leq k) = c_i, \quad \text{say.}$$

Lemma 4.11 gives that c_i is non-increasing in i. Putting $w_i = P_\theta(\eta = i)$ and $v_i = P_{\theta'}(\eta = i)$ for $i \leq b$, we complete the proof simply by showing that the hypothesis of Lemma 4.10 holds. Observe that $w_i = (1 - \theta_a)^{i-1}\theta_a$ and $v_i = (1 - \theta'_a)^{i-1}\theta'_a$, so that for any j, $\sum_{i \leq j} w_i = 1 - (1 - \theta_a)^{j+1}$ and $\sum_{i \leq j} v_i = 1 - (1 - \theta'_a)^{j+1}$. From the hypothesis that $\theta'_a \geq \theta_a$, it follows that $\sum_{i \leq j} w_i \leq \sum_{i \leq j} v_i$ holds for all j. ∎

Lemma 4.13: *Let $1 \leq m \leq a$. Suppose θ and θ' are such that $\theta'_m \geq \theta_m$ while $\theta'_r = \theta_r$ for all $r \neq m$. Then $Z^{a,b}$ is stochastically larger under P_θ than under $P_{\theta'}$.*

Proof: In view of Lemma 4.12, we need only consider $m < a$. Observe that for $k \geq m$, the hitting time η of the vertical line $x = k$ has the same distribution under both P_θ and $P_{\theta'}$. In view of

$$P_\theta(Z^{a,b} \leq k) \quad = \quad P_\theta(\eta \leq b - a + k - 1)$$

and similar equality under $P_{\theta'}$, it follows that

$$P_\theta(Z^{a,b} \leq k) \quad = \quad P_{\theta'}(Z^{a,b} \leq k) \quad \text{for all} \quad k \geq m.$$

We now consider $k \leq m - 1$ and show

$$P_\theta(Z^{a,b} \leq k) \quad \leq \quad P_{\theta'}(Z^{a,b} \leq k)\,.$$

Let us now denote η to be the hitting time of the vertical line $x = m - 1$. By the same argument as used in Lemma 4.12, one sees that

$$P_\theta(Z^{a,b} \leq k) \quad = \quad \sum_{i=1}^{b+2(a-m)} P_\theta(Z^{m-1,b-i+2a-2m+2} \leq k) P_\theta(\eta = i)$$

and

$$P_{\theta'}(Z^{ab} \leq k) \quad = \quad \sum_{i=1}^{b+2(a-m)} P_{\theta'}(Z^{m-1,b-i+2a-2m+2} \leq k) P_{\theta'}(\eta = i)\,.$$

Let $w_i = P_\theta(\eta = i)$ and $v_i = P_{\theta'}(\eta = i)$ for $i \leq b + 2(a-m)$. As in the proof of Lemma 4.12, we get the desired result once we show that for every j, $\sum_{i \leq j} w_i \leq \sum_{i \leq j} v_i$, that is, $P_\theta(\eta \leq j) \leq P_{\theta'}(\eta \leq j)$. This can perhaps be seen directly but here is a trite method.
Let $\bar{\eta}$ be the hitting time of the vertical line $x = m$. Noting that $\bar{\eta}$ has the same distribution under P_θ and $P_{\theta'}$, it suffices to show that for every $l \leq j-1$, $P_\theta(\eta \leq j \,|\, \bar{\eta} = l) \leq P_{\theta'}(\eta \leq j \,|\, \bar{\eta} = l)$. Using the Markov property, one sees that $P_\theta(\eta \leq j \,|\, \bar{\eta} = l) = 1 - (1 - \theta_m)^{j-l}$, while $P_{\theta'}(\eta \leq j \,|\, \bar{\eta} = l) = 1 - (1 - \theta'_m)^{j-l}$, from which the required inequalities follow. The proof is now complete. ∎

Proof of Theorem 4.9: Define $a+1$ many a-tuples, $\theta^0, \theta^1, \ldots, \theta^a$ by

$$\begin{aligned} \theta_i^m &= \theta'_i \quad \text{for} \quad i \geq a - m + 1 \\ &= \theta_i \quad \text{for} \quad i \leq a - m\,. \end{aligned}$$

Note that for any $0 \leq m \leq a-1$, we have $\theta^m_{a-m} \leq \theta^{m+1}_{a-m}$ and $\theta^m_r = \theta^{m+1}_r$ for all $r \neq a - m$. It follows from Lemma 4.13 that

$$P_{\theta^m}(Z^{a,b} \leq k) \quad \leq \quad P_{\theta^{m+1}}(Z^{a,b} \leq k)$$

for all k and all m with $0 \leq m \leq a-1$. Noticing that $\theta^0 = \theta$ and $\theta^a = \theta'$ the proof is complete. ∎

4.4 Spread in Households: Chain Binomial Models

The models discussed so far study the spread of an epidemic in a community at large. In this section, we take up the question of how an infectious disease spreads in a particular household. We shall discuss two stochastic models to describe this phenomenon — one is due to M. Greenwood and the other due to J. Reed and W.H. Frost.

Suppose that in a household, some individuals got infected by a contagious disease. This puts the other members of the household at the risk of catching the disease. Of course, in reality there is a fixed period of incubation and it is only after that period, that the infected individuals become infectious. The disease now spreads through contacts between the infected and uninfected individuals. However, not every such contact is likely to result in a new infection. Thus, there is a chance factor arising out of both the possibility of contact as well as a contact resulting in an infection. Specification of this chance factor is what would constitute a stochastic model. Before going into the details of the models, we describe the common setup.

We assume that there are K individuals in a household and initially s_0 of them are infected. We denote by r_0 the initial number of uninfected individuals, that is, $r_0 = K - s_0$. We assume that the incubation period is one time unit. To simplify matters, we also assume that the infected individuals remain infectious only for an instant of time at the end of the incubation period. This is indeed a simplifying assumption. However, in reality the period of infectiousness may often be very small, for example, they may perhaps be quarantined or even be cured and become immune. Let s_1 denote the number of new infections at time 1. The number of uninfected in circulation now is $r_1 = r_0 - s_1$. In general, let s_n be the number of persons who got infected at time n and $r_n = r_{n-1} - s_n$ be the resulting number of uninfected in circulation. It is to be noted that, at time n, the persons who can pass on the infection are precisely those who became newly infected at time $n-1$. Also, at any point of time the persons who are susceptible are only those who have not been infected so far. Clearly, as soon as $s_n = 0$, there will be no more new infections and the epidemic will come to a halt. Of course $r_{n+1} = 0$ would also guarantee this (perhaps not in a desirable way).

4.4.1 Greenwood Model

According to the model proposed by M. Greenwood, the probability of a susceptible coming in contact with the group of infectious persons and getting himself infected is assumed to be a constant p, $0 < p < 1$. Moreover the fates of different susceptibles are assumed to be stochastically independent. Clearly, these assumptions lead to a binomial distribution for the number of new infections at time n. More precisely, if at time $(n-1)$, there are s_{n-1} newly-infected persons (with $s_{n-1} > 0$) and if r_{n-1} denotes the number of susceptibles, then the probability of s_n new infections at time n is

$$\binom{r_{n-1}}{s_n} p^{s_n} (1-p)^{r_{n-1}-s_n}, \qquad \text{for } s_n = 0, 1, \ldots . r_{n-1} .$$

In case $s_{n-1} = 0$, then $s_n = 0$ and hence $r_n = r_{n-1}$. Note that in case $s_{n-1} > 0$, its actual value has no relevance in the distribution of the number of new infections at time n. This is one of the important features of this model.

Denote by S_n and R_n, the number of new infections and the number of susceptibles respectively at time n. Thus S_n and R_n are random variables. Also $(R_n, S_n)_{n \geq 0}$ is a Markov chain. The state space of this process is

$$\{(i,j) : i \geq 0,\ j \geq 0,\ i + j \leq K\} .$$

This is an absorbing chain and the absorbing states are precisely the states $\{(i,0) : i \leq K\}$. The transition probabilities are given by

$$\begin{aligned} P_{(i,j)(i',j')} &= \binom{i}{j'} p^{j'} (1-p)^{i'} \quad \text{for } 0 \leq j' \leq i \text{ and } i' = i - j' \quad \text{if} \quad j > 0 \\ &= \delta_{(i,j)(i',j')} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if} \quad j = 0 . \end{aligned}$$

Note that, for $j > 0$, $P_{(i,j),(i',j')}$ does not depend on j. This enables us to replace the original bivariate chain by an appropriately stopped univariate chain as follows.

Consider the Markov chain $(X_n)_{n\geq 0}$ with state space $\{0, 1, \ldots, K\}$, initial state $X_0 = r_0$, and transition probabilities

$$P(X_{n+1} = j \mid X_n = i) = \binom{i}{j} p^{i-j}(1-p)^j \quad \text{for} \quad j = 0, 1, \ldots, i\,.$$

Let T be the stopping time defined by

$$T = \min\{n : X_n = X_{n-1}\},$$

that is, T is the first time the chain did not move from its previous state. Let (Y_n) be the process (X_n) stopped at T, that is,

$$\begin{aligned} Y_n &= X_n \quad \text{if} \quad n \leq T \\ &= X_T \quad \text{if} \quad n > T\,. \end{aligned}$$

It is to be noted that $(Y_n)_{n\geq 0}$ is no longer a Markov chain, as we are going to see. A moment's reflection shows that the process $(Y_n)_{n\geq 0}$ is precisely $(R_n)_{n\geq 0}$ of the Greenwood Model; just recall that $S_n = 0$ is same as saying that $R_n = R_{n-1}$. And of course, $(R_n)_{n\geq 0}$ itself is not a Markov chain. In the new formulation, the random variable T is clearly seen to represent the duration of the epidemic. The rest of this section is devoted to finding the distribution of T.

More generally, let $(X_n)_{n\geq 0}$ be a Markov chain with state space $\{0, 1, \ldots, K\}$ and an arbitrary transition matrix $P = ((p_{ij}))$. We only assume that the diagonal entries of P are positive. For this Markov chain, we want to find the distribution of the stopping time T defined as

$$T = \min\{n : X_n = X_{n-1}\}\,.$$

The analysis that follows is due to J. Gani and M. Jerwood. Let Q denote the diagonal matrix with diagonal entries same as those of P and let $R = P - Q$. Clearly, R has all its off-diagonal elements same as those of P, while all its diagonal entries are zero. Using this notation it is now easy to see that

$$P(T = n, X_n = j \mid X_0 = i) = R_{ij}^{n-1} p_{jj} \quad \text{for each } j\,, \tag{13}$$

so that

$$P(T = n \mid X_0 = i) = e_i' R^{n-1} Q e, \tag{14}$$

where e is the column vector with all entries one and e_i' is the row vector with i-th entry one and all other entries zero. Recall that R is a matrix with non-negative entries and having each row sum strictly less than one. Elementary matrix theory shows that $(I - R)$ is invertible and $(I - R)^{-1} = \sum_{n=0}^{\infty} R^n$. The facts that $Q = P - R$ and $Pe = e$ can now be used to deduce that

$\sum_{n=1}^{\infty} e_i' R^{n-1} Qe = 1$. In view of (14), we have thus proved that T is finite with probability one. We now go on to find the p.g.f. of T. Throughout, we assume that we are starting from a fixed initial state i and denote by $g_i(\theta)$, the corresponding p.g.f. of T. Thus

$$g_i(\theta) = \sum_{n=1}^{\infty} \theta^n e_i' R^{n-1} Qe = e_i'(I - \theta R)^{-1}(\theta Q)e\,.$$

In view of its similarity with the p.g.f. of the standard geometric distribution, the distribution of T has been called a *Markov Geometric Distribution* by Gani and Jerwood. The moments of T — in particular its expectation and variance — can now be easily obtained by successive differentiation of the p.g.f. For example, it turns out that

$$E(T \,|\, X_0 = i) = e_i'(I - R)^{-1}e\,.$$

Turning back to (13), it can also be written as

$$P(T = n \text{ and } X_T = j \,|\, X_0 = i) = R_{ij}^{n-1} p_{jj}\,.$$

Thus we actually have the joint distribution of (T, X_T). One can use this to find the marginal distribution of X_T, in particular the expected value and variance of X_T.

Let us now return to the Greenwood Model. This is a special case where,

$$P = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ p & q & 0 & \dots & . \\ p^2 & 2pq & q^2 & \dots & . \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ p^K & \binom{K}{1} p^{K-1} q & \binom{K}{2} p^{K-2} q^2 & \dots & q^K \end{pmatrix}.$$

Note that in this case R^K is the zero matrix, implying that $T \leq K$ with probability one. This is understandable, because the longest possible duration of the epidemic happens when one new person gets infected each day. However this does not make explicit computations all that easy (compare R^2 and R^3!).

4.4.2 Reed-Frost Model

The model proposed by J. Reed and W.H. Frost differs from the earlier one in that here the probability of one particular susceptible coming in contact with one particular infectious person and getting infected is denoted by p. Of course the usual assumption of independence of interaction between different individuals is retained. It follows that, with the same notations r_n, s_n as in the Greenwood Model, the probability of any one of the r_{n-1} susceptibles getting infected at time n is $1 - (1-p)^{s_{n-1}}$, so that the probability of s_n many new infections at time n is given by

$$\binom{r_{n-1}}{s_n} [1 - (1-p)^{s_{n-1}}]^{s_n} (1-p)^{s_{n-1}(r_{n-1} - s_n)} \quad \text{for} \quad s_n = 0, 1, \dots, r_{n-1}\,.$$

Note that in case $s_{n-1} = 0$, the above formula automatically implies that $s_n = 0$. For detailed analysis of the Reed-Frost Model, interested reader can consult Von Bahr & Martin Lof (1980) and F. Ball (1983).

4.5 Spatial Spread: Neyman-Scott Model

In the earlier models, the geographic location of the epidemic was fixed and the temporal spread was under study. In this section, we describe a model proposed by J. Neyman and E. L. Scott for the spread of epidemic over a geographical area.

The geographical area under consideration is called the *habitat*, denoted by H. Mathematically, H could be any subset of the Euclidean plane R^2, preferably open. However, to simplify matters, we take our habitat to be all of R^2. As before, the incubation period is assumed to be one time unit and the period of infectiousness is contracted to a single point. It is reasonable to assume that an infectious person at a particular location in the habitat infects only susceptibles at that location. However, it is equally reasonable that an infectious person at a crowded location is likely to infect more people than at a desolate location. This dependence of infectivity on the location should be captured in the model. Another feature of the proposed model is that it tries to capture the mobility factor also. In other words, it takes into account the fact that an individual infected at a particular location may move to another location by the time he becomes infectious. Indeed, that is how the infection spreads over the habitat. Mathematically, this will involve introducing two parameters, one for the spread of infection and the other for the mobility of the infected individual. This is done as follows.

For every $u \in H$, we have a probability distribution γ_u on non-negative integers, representing the distribution of the number of individuals infected by one infectious person at the location u. We denote the p.g.f. of γ_u by $g(\cdot \mid u)$. We emphasize the dependence of γ_u on u as mentioned earlier. To take care of the mobility factor we have, for every $u \in H$, a probability density function $f_u(\cdot)$ on H. This has the following interpretation. A person infected at u at time k moves to a region $S \subset H$ at time $(k+1)$ with probability $\int_S f_u(x)\,dx$. Our model assumes that different individuals act independently. That is, first of all, the number of individuals infected by different infectious persons are independent random variables, no matter where the infectious persons are located. Secondly, given any set of infected individuals at time k, no matter where they are located, the places where they move to at time $k+1$ are independent random variables. The dependence of $f_u(\cdot)$ on the location u has an easy explanation. A person infected at home is not very likely to move away from home by the time he becomes infectious; whereas, if one catches infection when he is on board a train, he is quite likely to move far away.

We shall now see how to describe mathematically the temporal spread of the epidemic over the entire habitat. To fix ideas, we start with one infectious

person at time 0 at location u. This infectious person will infect a certain number, say X_1^u , of individuals at location u. From what has been said earlier, X_1^u is a random variable with distribution γ_u and p.g.f. $g(\cdot\,|\,u)$. By the time these persons become infectious, they would have moved to various locations — each person, independently of the others, choosing a random location given by the probability density $f_u(\cdot)$. Each of them would now infect people in their respective new locations, who would in turn move to different locations by the time they become infectious, and so on. This is how the infection would spread over the habitat with time. Let X_n^u denote the number of infectious people in the entire habitat at time n. The dependence on u comes from the fact that we started initially with one infected person at the location u. We want to study the distribution of X_n^u for $n \geq 1$. Let us denote the p.g.f. of X_n^u by $G_n(\cdot\,|\,u)$. Thus clearly

$$G_1(\cdot\,|\,u) = g(\cdot\,|\,u)\,. \tag{15}$$

To get the p.g.f. of X_2^u, we argue as follows. Consider the i-th person infected by the initial infectious and let Y_i denote the number of individuals infected by him at time 1. It should be clear that Y_i has p.g.f. given by

$$H_1(t\,|\,u) = \int g(t\,|\,x) f_u(x)\,dx\,. \tag{16}$$

This is because, given that the i-th individual has moved to location x at time 1, the conditional p.g.f. of the number of individuals infected by him is $g(\cdot\,|\,x)$, so that the unconditional p.g.f. would indeed be H_1 as given. It should be noted here that to make sense of the integral in (16), some assumptions on g as a function of x are needed. [For example, assuming that g varies continuously with x would do. A reader familiar with the Lebesgue Integration Theory would quickly see that measurability of g in the x variable is all that is needed.]

Since the Y_i are i.i.d. and $X_2^u = \sum_{i=1}^{X_1^u} Y_i$, it follows that

$$G_2(\cdot\,|\,u) = g(H_1(\cdot\,|\,u)\,|\,u)\,.$$

Proceeding in an analogous manner we can deduce that, for every $n \geq 1$,

$$H_n(\cdot\,|\,u) = \int G_n(\cdot\,|\,x) f_u(x)\,dx, \tag{17}$$

$$G_{n+1}(\cdot\,|\,u) = g(H_n(\cdot\,|\,u)\,|\,u)\,. \tag{18}$$

Note that even if we are interested only in $G_n(\cdot\,|\,u)$ for the specified initial location u, we have to compute $G_n(\cdot|x)$ for all x in order to get $H_n(\cdot|u)$, which is required for the subsequent $G_{n+1}(\cdot\,|\,u)$. Having started with one infectious individual at time 0 at location u, we have obtained the p.g.f. of the number of infectious individuals in the habitat at time n to be $G_n(\cdot\,|\,u)$. Along the way, we came across another sequence of functions which are also p.g.f.s, namely the $H_n(\cdot|u)$. The reader would naturally wonder as to which stochastic process they

correspond to. Well, instead of starting with one *infectious* person, suppose we had started with one *infected* person at time 0 at location u. Let us now consider the number of infected individuals — say Z_n^u, for $n \geq 1$ — in the entire habitat at successive time points. It is then clear that Z_n^u has p.g.f. $H_n(\cdot \mid u)$. Neyman and Scott describe X_n^u (Z_n^u respectively) as the n-th generation of an epidemic started by an infectious (infected, respectively) at location u at time 0. Let us denote the expected values of X_n^u and Z_n^u by α_n^u and β_n^u respectively. Using Equations (15)–(18) one gets

$$\alpha_1^u = m(u), \tag{19}$$

$$\beta_1^u = \int m(x) f_u(x)\, dx, \tag{20}$$

$$\alpha_{n+1}^u = m(u)\beta_n^u, \tag{21}$$

$$\beta_{n+1}^u = \int \alpha_{n+1}^x f_u(x)\, dx, \tag{22}$$

where $m(u)$ denotes the mean of the distribution γ_u. It is worth noting that, in order for the above formulae to be true, it is not necessary to assume that $m(x)$ is finite for each x.

We next discuss the problem of extinction of the epidemic from the habitat. We say that the epidemic, originating at u, is *heading for an extinction*, if X_n^u converges to zero in probability. Since the X_n^u are integer valued, this is equivalent to saying that $P[X_n^u = 0] \longrightarrow 1$ as $n \to \infty$. Here is a first positive result in this direction.

Theorem 4.14: *If* $\sup_x m(x) < 1$, *then for every* u, *the epidemic originating at* u *heads for extinction.*

Proof: Denote $\sup_x m(x)$ by c. Then clearly for any u, $\alpha_1^u \leq c$ by (19) and $\beta_1^u \leq c$ by (20). Using induction and the relations (21) and (22), one easily obtains that, for every u in H and every $n \geq 1$, $\alpha_n^u \leq c^n$ and $\beta_n^u \leq c^n$. In particular if $c < 1$, then $\alpha_n^u \to 0$. That is, $E(X_n^u) \to 0$ as $n \to \infty$. Note that

$$E(X_n^u) = \sum_{k=1}^{\infty} kP(X_n^u = k) \geq \sum_{k=1}^{\infty} P(X_n^u = k) = 1 - P(X_n^u = 0)\,.$$

It follows that $P(X_n^u = 0) \to 1$ as $n \to \infty$. ∎

We admit that the hypothesis that $\sup_x m(x) < 1$ in the above Theorem 4.14 is a strong one. However, it should be noted that, first of all, nothing is assumed about the mobility distributions. Secondly, the conclusion of the theorem is also very strong in the sense that the convergence is uniform over u, that is, $\sup_u P(X_n^u > 0) \to 0$. The next theorem on extinction has a slightly weaker hypothesis.

Theorem 4.15: *If* $\sup_x \beta_1^x < 1$, *then for every* u, *such that* $m(u) < \infty$, *an*

epidemic originating at u heads for extinction.

Proof: Denote $\sup_x \beta_1^x$ by c. It is easy to see by induction that, for every $n \geq 1$, $\alpha_n^u \leq m(u)c^{n-1}$. Therefore, if $c < 1$ and $m(u) < \infty$, then $\alpha_n^u \to 0$. The proof is now completed as earlier. ∎

The reader may note that under the hypothesis of Theorem 4.14, one surely has $\sup_x \beta_1^x < 1$ and, of course, for every u, $m(u) < 1$. After the above two theorems, which assert that under certain conditions the epidemic heads for extinction, we now go to a result describing when an epidemic does not.

Theorem 4.16: *If $H_1(0 \mid u) = 0$ for every u, then for every u such that $\gamma_u(\{0\}) < 1$, an epidemic originating at u does not head for extinction.*

Proof: We first show that $H_n(0 \mid u) = 0$, for every u and every n. Indeed, $G_2(0 \mid u) = g(0 \mid u)$, by Equation(18) and the hypothesis. But this, in turn, implies that $H_2(0 \mid u) = 0$ for all u, by Equation (17) and the hypothesis. Induction will now do the job. In particular, for every u and every n, one has $P(X_n^u = 0) = G_n(0 \mid u) = g(0 \mid u) = \gamma_u(\{0\})$, independent of n. It follows that if $\gamma_u(\{0\}) < 1$, then X_n^u does not converge to zero in probability. ∎

We shall discuss one more problem related to this model. Can it so happen that an epidemic originating at some locations will head for extinction, whereas an epidemic originating at others will not? We show that this cannot happen unless there are deserts or unless the mobility is curtailed. A desert means a location where an infectious person can not infect anybody else. More precisely, we say that a point $u \in H$ is a *desert*, if the distribution γ_u is concentrated on the singleton $\{0\}$. Also we say that there is *full mobility* in the habitat if for every u, the density $f_u(\cdot)$ is strictly positive everywhere. This means that an infected person from any location can move to any other location with positive probability by the time he becomes infectious. We are now ready to state the main result.

Theorem 4.17: *Assume that there are no deserts and that there is full mobility. Then, an epidemic originating at u will head for extinction, either for all $u \in H$ or for no $u \in H$.*

For the proof, we need a little lemma on integrals. We simply state it without proof, just because the proof needs Lebesgue integration theory, something that we are not assuming from the reader. However, those who are familiar with this theory will quickly agree that the result is indeed elementary. For those who are not, here is a motivation: suppose $\{a_n\}$ is a non-negative sequence and you are told that for some strictly positive sequence $\{p_n\}$, the series $\sum a_n p_n = 0$. It trivially follows that a_n must equal zero for all n. Analogously, suppose that $a(x)$ is a non-negative function of a real variable and you are told that the integral $\int a(x)p(x)\,dx = 0$ for some strictly positive function p. Now of course we cannot say that $a(x)$ must equal zero for all x. However what the lemma asserts is that $a(x)$ is almost zero.

Lemma 4.18: *Let $a(x)$ be a non-negative function on the real line such that*

for some strictly positive function $p(x)$, $\int a(x)p(x)\,dx = 0$. *Then for every function* $q(x)$, $\int a(x)q(x)\,dx = 0$.

Proof of Theorem 4.17: We start by showing that $G_n(0 \mid u)$ increases with n, for every u. This does not require any of the hypotheses of the theorem. We prove by induction. First note that $G_1(0 \mid u) = g(0 \mid u)$ and $H_1(0 \mid u) = \int g(0 \mid x) f_u(x)\,dx \geq 0$, so that

$$G_2(0|u) = g(H_1(0|u)|u) \quad \geq \quad g(0|u) = G_1(0|u)\,.$$

The inequality is a consequence of the fact that $g(\cdot \mid u)$ is a p.g.f. and hence non-decreasing. Assuming now that $G_n(0|x) \geq G_{n-1}(0|x)$ for all x, Equations (17) and (18) can be used to show that $G_{n+1}(0|x) \geq G_n(0|x)$.
As a consequence, for every x, $\lim_n G_n(0|x)$ exists, to be denoted by $G_\infty(0|x)$. From Equation (17), it follows that for every x, $H_n(0|x)$ is also non-decreasing with n and hence has a limit, say $H_\infty(0|x)$. Further, the Equation (17) and the Monotone Convergence Theorem [see Exercise 4, Section 0.4] give

$$H_\infty(0|u) = \int G_\infty(0|x) f_u(x)\,dx\,. \tag{23}$$

Again Equation (18) and the continuity of g_u gives

$$G_\infty(0|u) = g(H_\infty(0|u)|u)\,. \tag{24}$$

To prove the theorem now, suppose that for some $u_0 \in H$, an epidemic starting at u_0 heads for extinction, that is $G_\infty(0|u_0) = 1$. We show that, under the hypotheses of the theorem, $G_\infty(0 \mid u) = 1$ for all u, that is, an epidemic starting at any u heads for extinction. First observe that the hypothesis that there are no deserts, implies in particular, that $\gamma_{u_0}(\{0\}) < 1$. This, in turn, implies that $g(\cdot|u_0)$ is strictly increasing and, hence $g(t|u_0) = 1$ if and only if $t = 1$. Therefore, Equation (24) implies that $H_\infty(0|u_0) = 1$. In view of Equation (23), this means that $\int [1 - G_\infty(0|x)] f_{u_0}(x)\,dx = 0$. Now invoke Lemma 4.18, with $a(x) = 1 - G_\infty(0|x)$ and $p(x) = f_{u_0}(x)$, to deduce that for any u, $\int [1 - G_\infty(0|x)] f_u(x)\,dx = 0$, that is, $H_\infty(0|u) = 1$. This implies, by Equation (24), that $G_\infty(0|u) = 1$, as was to be proved. ∎

So far, we have been considering the spread of the epidemic over the entire habitat. However, in practice one may be more interested in the spread of the epidemic over certain pockets of the habitat. More precisely, let $R_1, R_2, \ldots, R_k$ be k disjoint sub-regions of the habitat. For any $u \in H$, let X_{ni}^u, for $1 \leq i \leq k$, denote the number of infectious persons in the region R_i at time n, for an epidemic starting at u. Note that we do not demand that $\bigcup_{i=1}^k R_i = H$; also, we allow for the possibility that $u \notin \bigcup_{i=1}^k R_i$. We may be interested in the distribution of the vector process $\widetilde{X}_n^u = (X_{n1}^u, \ldots, X_{nk}^u)$, in particular its asymptotic properties. The methods of this section enable us to discuss these issues as well. We discuss some of them in the exercises.

4.6 Exercises

1. Consider the simple deterministic epidemic. Sometimes the following function is interpreted as the frequency function of the time of occurrence of a new infection.

$$\frac{\omega(t)}{n} = -\frac{1}{n}\frac{dx}{dt} = \beta(n+1)^2 \frac{e^{\beta(n+1)t}}{[n+e^{\beta(n+1)t}]^2}.$$

Show that this is indeed the probability density of a non-negative random variable with mean $\log(n+1)/\beta n$.

2. Consider the simple stochastic epidemic with $\beta = 1$. Let q_r be the Laplace transform of p_r, that is ,

$$q_r(\lambda) = \int_0^\infty e^{-\lambda t} p_r(t)\, dt \quad \text{for} \quad \lambda > 0.$$

Show that

$$q_r = \frac{(r+1)(n-r)}{\lambda + r(n+1-r)} q_{r+1} \quad \text{for} \quad 0 \le r < n,$$

$$q_n = 1/(\lambda + n).$$

Hence deduce that

$$q_r = \frac{n!(n-r)!}{r!} \prod_{j=r}^{n} \frac{1}{\lambda + j(n-j+1)}.$$

3. Consider the simple stochastic epidemic. Let $F(x,t)$ be the generating function of $p_r(t)$ for $0 \le r \le n$, that is,

$$F(x,t) = \sum_{r=0}^{n} p_r(t) x^r.$$

Show that $F(x,0) = x^n$ and

$$\frac{\partial F}{\partial t} = \beta(1-x)\left(n\frac{\partial F}{\partial x} - x\frac{\partial^2 F}{\partial x^2}\right).$$

4. In the Chain-Binomial models, let $P(s_0, s_1, \ldots, s_n)$ denote the probability $P(S_0 = s_0, S_1 = s_1, \ldots, S_n = s_n)$.
Show that in the Greenwood model

$$P(s_0, s_1, \ldots, s_n) = \frac{r_0!}{s_1! s_2! \cdots s_n! r_{n+1}!} p^{\sum_1^n s_i} q^{\sum_1^{n+1} r_j}$$

and in the Reed-Frost model

$$P(s_0, s_1, \ldots, s_n) = \frac{r_0!}{s_1! s_2! \cdots s_n! r_{n+1}!} \; q^{\sum_0^n s_j r_{j+1}} \prod_0^{n-1} (1 - q^{s_i})^{s_{i+1}} .$$

5. Let $P(n, j, a)$ be the probability that a household of size n will have a total of j cases when there are a initial cases. Show that for the Greenwood model

$$P(n, j, a) = \sum_{k=1}^{j-a} \binom{n-a}{k} p^k q^{n-a-k} P(n-a, j-a, k) .$$

Show that for the Reed-Frost model, $P(n, a, a) = q^{a(n-a)}$, and

$$P(n, j, a) = \sum_{k=1}^{j-a} \binom{n-a}{k} (1 - q^a)^k q^{a(n-a-k)} P(n-a, j-a, k) .$$

Hence deduce that in the Reed-Frost model,

$$P(n, j, a) = \binom{n-a}{j-a} q^{j(n-j)} P(j, j, a) .$$

6. The idea is to describe, following R. Bartoszynski, a branching process model of epidemics. Here is the set-up.
(i) Every infected individual passes through a period of illness of $X + Y$ days, that is, a period of incubation of X days followed by a period of infectiousness of Y days. (X, Y) has joint p.g.f. $F(s, t) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} p_{m,n} s^m t^n$. It is to be noted that Y is at least one. (ii) During the illness period of $X + Y$ days, a person may be detected and automatically isolated. The probability of getting detected on a day is $(1 - \alpha)$ during the incubation period and $(1 - \beta)$ during the infectious period. Here $0 < \alpha, \beta \leq 1$. (iii) During the Y days of infectiousness an undetected individual makes a certain number of contacts with the susceptibles. The number of contacts for different days are i.i.d with p.g.f. $R(t) = \sum_{k=0}^{\infty} r_k t^k$. (iv) Each contact of a susceptible with an infectious, independent of other contacts, leads to infection with probability γ where $0 < \gamma \leq 1$. (v) The events described above are independent for different individuals.
The interpretations of (i)–(v) are as follows. F describes the nature of the disease; R describes the social and environmental conditions like mobility etc.; α and β describe the efficiency of the health services in detecting the cases; γ measures the individual resistance via immunization programs of the health services.

(a) Put $q_{m,n} = \sum_{k=n+1}^{\infty} p_{m,k}$, $\quad Q(s, t) = \sum_{m,n=0}^{\infty} q_{m,n} s^m t^n$.
Show that $\sum_{m=0}^{\infty} q_{m,0} s^m = F(s, 1)$.
Show that for $|s| \leq 1$ and $|t| < 1$, $Q(s, t) = \dfrac{F(s, 1) - F(s, t)}{1 - t}$.

(b) Put ω_n = the probability that an infected individual remains undetected and infectious for exactly n days. Show that

$$\omega_n = \sum_{m=0}^{\infty} p_{m,n}\alpha^m\beta^n + (1-\beta)q_{m,n}\alpha^m\beta^n \quad \text{for} \quad n \geq 1,$$

$$\omega_0 = 1 - \sum_{1}^{\infty} \omega_n = 1 - \beta F(\alpha, 1)\,.$$

(c) Show that the p.g.f. of the number of persons infected by a single individual during one day of his infectiousness (when undetected) is given by $R(1-\gamma+\gamma t)$.

(d) Let $D(s) = \frac{\partial}{\partial t}F(s,t)\,|_{t=1} = \sum\sum np_{m,n}s^m$.
Show that $F(\alpha, 1)$ is the probability that an infective remains undetected during the whole incubation period.
Show that $D(\alpha)/F(\alpha, 1)$ is the expected length of infectious period for those who remain undetected during the incubation period.

(e) Let $G(t)$ be the p.g.f. of the number of individuals infected by a single infective. Show that

$$\begin{aligned} G(t) &= \sum_{n=0}^{\infty} \omega_n R^n(1-\gamma+\gamma t) \\ &= 1 - F(\alpha,1) + F(\alpha, \beta R(1-\gamma+\gamma t)) \\ &\quad +(1-\beta)\frac{F(\alpha,1) - F(\alpha, \beta R(1-\gamma+\gamma t))}{1-\beta R(1-\gamma+\gamma t)}\,. \end{aligned}$$

If $R'(1) = \gamma$ with $0 < \gamma < \infty$, then show that $G'(1) = \gamma r D(\alpha)$ in case $\beta = 1$, and, $= \gamma r \frac{\beta}{1-\beta}[F(\alpha,1) - F(\alpha,\beta)]$, in case $\beta < 1$.

(f) For $0 < x \leq 1$, put $h_1(x) = F(\alpha, x)$ and $h_2(x) = F(\alpha,1) - \frac{1}{\gamma r}\frac{1-x}{x}$.
Show that $h_1' > 0$, $h_2' > 0$, $h_1'' \geq 0$ and $h_2'' < 0$.
Show that $x = 1$ is a root of $h_1 = h_2$, and, if $h_1'(1) \leq h_2'(1)$, then this is the only root.
Show that if $h_1'(1) > h_2'(1)$ then there is one more root $x < 1$ of the equation $h_1 = h_2$.

(g) Consider the n-th generation of infected individuals as follows. $Z_0 = 1$ and for $n \geq 1$, Z_n = the number of persons infected by the Z_{n-1} persons of the $(n-1)$-th generation. Show that (Z_n) is a branching process with progeny generating function $G(t)$ as given above. Define $\beta^*(\alpha,\gamma)$ as the smallest positive root of

$$\gamma\, r\, x\, [F(\alpha, x) - F(\alpha, 1)] + 1 - x = 0\,.$$

Show that

$$P(\lim_n Z_n = 0) = 1 \quad \text{iff} \quad \beta \leq \beta^*(\alpha,\gamma)\,.$$

[There are generalizations to the case when there are N zones and there is mobility for people from one zone to another.]

7. The idea is to describe a simple mathematical model for muscle movements. This is due to S. W. Greenhouse.

Phenomenon: A muscle fibril consists of alternating thick and thin filaments. During shortening and stretching they slide along each other. While this happens, certain chemical interactions occur at the molecular level. There are sites on the thin filament. Concentrate on one site now. The site may be occupied by a molecule or may be vacant. There are positions on the thick filament which are alternately 'release' and 'load' positions. A release position can pick up a molecule and a load position can give a molecule. Suppose that 1,3,5,... are release positions and 2,4,6,... are load positions. Imagine a site now at 0 and sliding along the positions. Wish to know whether a site is filled or vacant at time n, and time is counted in units of positions crossed by the site. Note that if the site is filled and arrives at a load position, then nothing happens and it passes on to the next position. Similarly, if the site is vacant and arrives at a release position then nothing happens and it passes on to the next position.

In real life, positions on the thick filament are only release positions and moreover, a site may pickup a molecule at any point between two release positions. Further, during muscular contractions, the thin filament – and hence, the sites – move with varying velocity. The slower the speed, the greater the interaction and the parameters α and β given below change with n. This is a simplified treatment.

Mathematically, $X_0, X_1, \ldots$ is a sequence of random variables each taking values 0 (vacant site) and 1 (filled site). X_0 is the initial position and X_n its state after n interactions. The two matrices A and B given below represent the probabilities of transition from X_{2r} to X_{2r+1} and from X_{2r+1} to X_{2r+2} respectively.

$$A = \begin{pmatrix} 1 & 0 \\ \alpha & 1-\alpha \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1-\beta & \beta \\ 0 & 1 \end{pmatrix}.$$

Here $0 < \alpha < 1$ and $0 < \beta < 1$.
For $i = 0, 1$ and $n \geq 1$, let $p_n^i = P(X_n = 1 \,|\, X_0 = i)$.
Show that for even integers n,

$$p_n^0 = \beta \, \frac{1 - [(1-\alpha)(1-\beta)]^{n/2}}{1 - (1-\alpha)(1-\beta)},$$

$$p_n^1 = [(1-\alpha)(1-\beta)]^{n/2} + \beta \, \frac{1 - [(1-\alpha)(1-\beta)]^{n/2}}{1 - (1-\alpha)(1-\beta)},$$

and for odd integers n,

$$p_n^0 = \beta \, (1-\alpha) \, \frac{1 - [(1-\alpha)(1-\beta)]^{(n-1)/2}}{1 - (1-\alpha)(1-\beta)},$$

$$p_n^1 = (1-\alpha)[(1-\alpha)(1-\beta)]^{(n-1)/2} + \beta\,(1-\alpha)\,\frac{1-[(1-\alpha)(1-\beta)]^{(n-1)/2}}{1-(1-\alpha)(1-\beta)}.$$

8. The idea is to discuss a model for Leukemia proposed by I. A. Chow.
Phenomenon: The disease starts with anaemia and thrombocytopenia. This is followed by the appearance of immature leukemic cells which replace the normal mature leukocutes. As the disease advances the number of immature granulocytes increases while the number of normal cells (called polymorphonuclear granulocytes, abbreviated as PMNG) decreases. This is attributed to the fact that the abnormal immature cells have a long intravascular life and capacity for mitotic subdivision compared with normal cells. The incapability of the abnormal immature cells in phagocytosis makes the patient very susceptible to infection or haemmorhage leading to death.
Notation: Let $\lambda > 0$ and $\mu > 0$ be two numbers. m = maximum number of PMNG one can have. $X(t)$= Number of PMNG at time t. $X(0)$= initial number of PMNG at time 0 , say = n_0.
Modelling Assumptions: The probability that PMNG will decrease by 1 during $(t, t+\Delta t)$ given that there are n at time t is $n\mu\Delta t + o(\Delta t)$.
The probability that PMNG will increase by one during $(t, t+\Delta t)$ given that there are n at time t is $(m-n)\lambda\Delta t + o(\Delta t)$.
The probability that PMNG will not undergo any change during $(t, t+\Delta t)$ given that there are n at time t is $1-[n\mu+(m-n)\lambda]\Delta t + o(\Delta t)$.
Define $p_n(t) = P(X_t = n \,|\, X_0 = n_0)$ for $0 \le n \le m$, and $= 0$ for other values of n. Set $G(t,s) = \sum_{n=0}^{m} p_n(t)s^n$, the p.g.f. of X_t.

(a) Show that

$$\frac{dp_n(t)}{dt} = \lambda\,(m+n-1)\,p_{n-1} + \mu\,(n+1)\,p_{n+1} - [n\mu + (m-n)\lambda]\,p_n\,.$$

(b) Show that

$$\frac{\partial}{\partial t}G(t,s) = (1-s)\,(\mu+\lambda s)\,\frac{\partial}{\partial s}G(t,s) - (1-s)\,\lambda\,m\,G(t,s).$$

$$G(0,s) \quad = \quad s^{n_0}\,.$$

(c) Show that

$$G(t,s) \quad = \quad \left[1 - \frac{\lambda+\mu e^{-(\lambda+\mu)t}}{\lambda+\mu} + \frac{\lambda+\mu e^{-(\lambda+\mu)t}}{\lambda+\mu}s\right]^{n_0}$$

$$\times \left[1 - \lambda\frac{1-e^{-(\lambda+\mu)t}}{\lambda+\mu} + \lambda\frac{1-e^{-(\lambda+\mu)t}}{\lambda+\mu}s\right]^{m-n_0}.$$

(d) Show that X_t is the sum of two independent random variables, say, X_t^1 and X_t^2, where

$$X_t^1 \sim B(n_0, \alpha) \quad \text{with} \quad \alpha = \frac{\lambda+\mu e^{-(\lambda+\mu)t}}{\lambda+\mu},$$

$$X_t^2 \sim B(m - n_0, \beta) \quad \text{with} \quad \beta = \lambda \frac{1 - e^{-(\lambda+\mu)t}}{\lambda + \mu}.$$

This can be interpreted as follows. The PMNG at time t is made up of two kinds. First, there are those of the initial n_0 which are still surviving. Second, there are those that are liberated at some time $\tau < t$ and are still surviving at time t.

(e) Show that

$$E(X_t) = n_0 e^{-(\lambda+\mu)t} + m\lambda \frac{1 - e^{-(\lambda+\mu)t}}{\lambda + \mu}.$$

$$V(X_t) = n_0 \frac{\mu - \lambda}{\mu + \lambda} e^{-(\lambda+\mu)t} \left[1 - e^{-(\lambda+\mu)t}\right] + m\lambda \frac{1 - e^{-(\lambda+\mu)t}}{\lambda + \mu} \frac{\mu + \lambda e^{-(\lambda+\mu)t}}{\lambda + \mu}.$$

(f) Show that

$$p_0(t) = \left[\mu \frac{1 - e^{-(\lambda+\mu)t}}{\mu + \lambda e^{-(\lambda+\mu)t}}\right]^{n_0} \left[\frac{\mu + \lambda e^{-(\lambda+\mu)t}}{\lambda + \mu}\right]^m.$$

(g) Show that

$$p_0(\infty) = \lim_{t\to\infty} p_0(t) = \left(\frac{\mu}{\lambda + \mu}\right)^m.$$

(h) Assume that the volume of blood, say v units, is large and also the PMNG at time t is large. What is usually observed is Y_t, the density of PMNG, that is, the number of PMNG in unit volume of blood, at time t. Theoretically speaking, any of the X_t cells has a chance $1/v$ of appearing in the unit volume taken for the PMNG count. So it is believed that, given $X_t = n$, Y_t is Poisson with parameter n/v. In other words,

$$P(Y_t = n' \mid X_t = n) = \frac{(n/v)^{n'} e^{-(n/v)}}{n'!} \quad \text{for} \quad n' \geq 0.$$

Show that the conditional p.g.f. of Y_t given $X_t = n$ is $e^{-n(1-s)/v}$.

(i) If $H(t, s)$ is the unconditional p.g.f. of Y_t, then show that

$$H(t, s) = H_1(t, s) \cdot H_2(t, s),$$

where

$$H_1(t, s) = \left[1 - \frac{\lambda + \mu e^{-(\lambda+\mu)t}}{\lambda + \mu} + \frac{\lambda + \mu e^{-(\lambda+\mu)t}}{\lambda + \mu} e^{-(1-s)/v}\right]^{n_0},$$

$$H_2(t, s) = \left[1 - \lambda \frac{1 - e^{-(\lambda+\mu)t}}{\lambda + \mu} + \lambda \frac{1 - e^{-(\lambda+\mu)t}}{\lambda + \mu} e^{-(1-s)/v}\right]^{m-n_0}.$$

Conclude that Y, just like X, is the sum of two independent random variables.

(j) Show that

$$E(Y_t) = (n_0/v)\, e^{-(\lambda+\mu)t} + (m/v)\,\lambda\, \frac{1-e^{-(\lambda+\mu)t}}{\lambda+\mu},$$

$$\begin{aligned} V(Y_t) &= (n_0/v)\left[1 - (m/v)\,\frac{\lambda(1-e^{-(\lambda+\mu)t}}{\lambda+\mu}\right] \\ &\quad +(1/v)\,\frac{\mu-\lambda}{\mu+\lambda} e^{-(\lambda+\mu)t}\left[1-e^{-(\lambda+\mu)t}\right] \\ &\quad +(m/v)\,\lambda\,\frac{1-e^{-(\lambda+\mu)t}}{\lambda+\mu}\left[1 + (1/v)\frac{\mu+\lambda e^{-(\lambda+\mu)t}}{\lambda+\mu}\right]. \end{aligned}$$

(k) The parts (i) and (j) above are useful in estimating the parameters and making predictions. Chow considers these also.
In practice (n_0/v) and (m/v) are not observable and they are replaced by n_0', the initial PMNG density, and m', upper limit of the observed PMNG density.
If the patient is under treatment then the chances of a PMNG liberation from bone-marrow into the blood stream depends on the time instant t itself, apart from depending on the actual number at that time. In other words, λ is not a constant but a function of t. Similarly μ also is a function of t. These are denoted by $\lambda(t)$ and $\mu(t)$ respectively.
From now on this is what is assumed and m, X_0, X_t, n_0, p_n and $G(t,s)$ are as defined earlier.

(l) Argue that G satisfies a similar equation as earlier except that the numbers λ and μ are now functions of t.

(m) Show that

$$G(t,s) = [G_1(t,s)]^{n_o}\ [G_2(t,s)]^{m-n_0},$$

where

$$\begin{aligned} G_1(t,s) &= 1 - \left\{1 + \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau\right\} e^{-R(t)} \\ &\quad + \left\{1 + \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau\right\} e^{-R(t)}\, s, \end{aligned}$$

$$G_2(t,s) = 1 - \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau\, e^{-R(t)} + \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau\, e^{-R(t)}\, s$$

and

$$R(t) = \int_0^t [\lambda(\tau) + \mu(\tau)]\, d\tau\,.$$

(n) Show that

$$E(X_t) = n_0\, e^{-R(t)} + m\, e^{-R(t)} \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau$$

and

$$V(X_t) = n_0 e^{-R(t)} \left[1 - e^{-R(t)}\{1 + 2\int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau\}\right] + m\, e^{-R(t)} \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau \left[1 - e^{-R(t)} \int_0^t \lambda(\tau)\, e^{R(\tau)}\, d\tau\right].$$

4.7 References/Supplementary Readings

[1] von Bahr, B. and Martin-Löf, A. [1980] : Threshold limit theorems for some epidemic processes, *Adv. Appl. Prob.*, vol. 12, p. 319-349.

[2] Bailey, N. T. J [1975] : *The Mathematical Theory of Infectious Diseases and its Applications*, Charles Griffin and Co.

[3] Ball, F. [1983] : The threshold behaviour of epidemic models, *J. Appl. Prob.*, vol. 20, p. 227-241.

[4] Bartoszynski, R. [1969] : Branching Processes and Models of Epidemics, *Dissert. Math.*, vol. LXI.

[5] Bartoszynski, R. and Wyech-Los, M. [1965] : Contributions to the Theory of Epidemics, *Bernoulli-Bayes-Laplace Anniversary Volume* (ed. Neyman, J. and Lecam, L. M.) Springer-Verlag, p. 1-8.

[6] Bharucha-Reid, A. T. [1960] : *Elements of the Theory of Markov Processes and their Applications*, McGrawhill.

[7] Bourgin, D. G. [1974] : A simple model for virus-bacteria interaction, *Math. Biosc.*, vol. 20, p. 179.

[8] Chow, I. A. [1972] : Restricted immigration and death process and its application to .....Leukemia, *Math. Biosc.*, vol. 13, p. 253.

[9] Chow, I. A. [1960] : A stochastic approach to ..... Leukemia patients, *Math. Biosc.*, vol. 21, p. 103.

[10] Foster, F. G. [1955] : A note on Bailey's and Whittle's treatment of a general stochastic epidemic, *Biometrika*, vol. 42, p. 123.

[11] Gani, J. [1965] : On a partial differential equation of epidemic theory, *Biometrika*, vol. 52, p. 617.

[12] Gani, J. [1966] : On the General Stochastic Epidemic, *Proceedings of the Fifth Berkeley Symposium*, vol. IV, p. 271 - 279.

[13] Gani, J. and Jerwood, D. [1971] : Markov chain methods in chain binomial epidemic models, *Biometrics*, vol. 27, p. 591.

[14] Greenhouse, S. W. [1960] : A stochastic process arising in the study of

muscular contraction, *Proceedings of the Fourth Berkeley Symposium*, vol. 4, p. 257.

[15] Kendall, D. G. [1956] : Deterministic and stochastic epidemics in closed populations, *Proceedings of the Fourth Berkeley Symposium*, vol. 4, p. 149.

[16] Neyman, J. and Scott, E. L. [1963] : A stochastic model of epidemics, *Stochastic Models in Medicine and Biology*, ed. J. Gurland, p. 45.

[17] Rajarshi, M. B. [1981] : Simpler proofs of two threshold theorems for a general stochastic epidemic, *J. Appl. Prob.*, vol. 18, p. 721-724.

[18] Siskind, V. [1965] : A solution of the general stochastic epidemic, *Biometrika*, vol. 52, p. 613.

[19] Whittle, P. [1955] : The outcome of a stochastic epidemic – A note on Bailey's paper, *Biometrika*, vol. 42, p. 116.

[20] Williams, T. [1971] : An algebraic proof of the threshold theorem for the general stochastic epidemic (abstract), *Adv. Appl. Prob.*, vol. 3, p. 223.

[21] Yang, Grace [1968] : Contagion in stochastic models for epidemics, *Ann. Math. Stat.*, vol. 39, p. 1863.

[22] Yang, Grace [1972] : On the probability distribution of some stochastic epidemic models, *Theo.Pop.Bio.*, vol. 3, p. 448.