ON THE STOCHASTIC THEORY OF EPIDEMICS

A. T. BHARUCHA-REID COLUMBIA UNIVERSITY

1. Introduction

Early work in the mathematical theory of epidemics was mainly concerned with the development of deterministic models for the spread of disease through a population. An excellent review of the deterministic approach has been given by Serfling [15]. In this approach a functional equation (differential or integral equation, etc.) for n(t), the number of infected individuals in the population at time t, is derived on the basis of certain assumptions concerning the mechanism by which the disease is to be transmitted among members of the population. This equation, together with some initial condition (the number of infected individuals at time zero), is then solved to obtain n(t). In assuming a deterministic causal mechanism for the spread of an epidemic the number of infected individuals at some time t > 0 will always be the same if the initial conditions are identical. Because of the large number of random or chance factors which determine the manner in which an epidemic develops it became clear to workers in epidemic theory that probabilistic or stochastic models would have to be used to supplement or replace the existing deterministic ones.

The development of the theory of stochastic processes has given the mathematical epidemiologist the proper theoretical framework within which his mathematical models can be constructed. Of particular interest are stochastic processes of the branching or multiplicative type. These processes can be described as mathematical models for the development of systems whose components can reproduce, be transformed, and die; the development being governed by probability laws [9]. A discussion of some stochastic models in epidemic theory has been given by Taylor [16], and a detailed discussion of stochastic epidemic theory will be given in a monograph by the author [4]; hence, in this paper we will not give a review of previous work in this area. The purpose of the present paper is to consider the possible application of the Bellman-Harris theory of age-dependent branching processes [2] to epidemics, and to discuss some statistical problems associated with stochastic epidemics.

2. Age-dependent branching processes and epidemics

2.1. Introduction. In the Bellman-Harris theory the incubation period (defined as the length of time an individual is infected before infecting someone else) is a random variable, say τ , with general distribution $G(\tau)$, $0 < \tau < \infty$. At the end of this

This study was supported by funds provided under Contract AF 18(600)-939 with the USAF School of Aviation Medicine, Randolph Field, Texas.

period the infected individual can infect $n(n=0, 1, \cdots)$ other individuals with probabilities q_n , where the q_n sum to unity, and each newly infected individual has the same distribution $G(\tau)$ for the time that will elapse before he infects someone else. The Bellman-Harris process can be formulated as follows: let X(t) be an integer-valued random variable representing the number of infected individuals in the population at time t. Put $p(x,t) = Pr\{X(t) = x\}$, $x \ge 0$, and let

(2.1)
$$\pi(s, t) = \sum_{x=0}^{\infty} p(x, t) s^{x}, \quad |s| < 1$$

be the generating function for the probabilities p(x, t) when X(0) = 1. If X(0) = n > 1 the generating function for the probabilities is given by $\pi^n(s, t)$. In treating both cases it is necessary to assume that the infected individuals do not interact with one another. The generating function (2.1) has been shown to satisfy the non-linear Stieltjes functional equation

(2.2)
$$\pi(s, t) = s[1 - G(t)] + \int_0^t h[\pi(s, t - \tau)] dG(\tau)$$

where

$$h(s) = \sum_{n=0}^{\infty} q_n s^n ,$$

that is, h(s) is the generating function for the infection probabilities q_n . The equation for the generating function can be derived as follows. By definition

(2.4)
$$p(x, t) = Pr\{X(t) = x\} = \int_0^t Pr\{X(t) = x | \tau\} dG(\tau)$$

where $Pr\{X(t) = x | \tau\}$ is the probability of having x infected individuals at time t from a single infected individual at time zero who is known to have infected others at $t = \tau$. Now

(2.5)
$$Pr\{X(t) = x | \tau\} = \sum_{n=0}^{\infty} q_n \left\{ \sum_{i_1 + i_2 \cdots i_n = x} \prod_{k=1}^n p(i_k, t - \tau) \right\},$$

where the term in braces is the coefficient of s^x in the expansion of

(2.6)
$$\left[\sum_{n=0}^{\infty} p(x, t - \tau) s^{x}\right]^{n} = \pi^{n}(s, t - \tau) .$$

(The reasoning employed here is the same as that used in the theory of compound probability distributions.) Multiplying p(x, t) by s^x , summing over x, and adding the term for p(1, t) = 1 - G(t), we obtain (2.2). If G(t) has a density function of bounded total variation, we can write (2.2) as

(2.7)
$$\pi(s, t) = s[1 - G(t)] + \int_0^t h[\pi(s, t - \tau)]g(\tau)d\tau.$$

Differentiation of (2.7) with respect to s yields the moments of X(t) as integral equations of the renewal type, the properties of which are well known. For example, the expected value of X(t) is

(2.8)
$$E[X(t)] = m(t) = 1 - G(t) + K \int_0^t m(t - \tau)g(\tau)d\tau$$

where K is the sum of $nq_n = (dh/ds)_{s=1}$.

- 2.2. Some models. In this section we will use the age-dependent approach to show how several models for the spread of an epidemic can be constructed. Some of these models are well known and therefore will not be discussed in detail. In all cases we consider an infinite population of susceptible individuals; and the random variable X(t) will represent the size of the infected population at time t. We also assume X(0) = 1.
- (i) A deterministic model (Galton-Watson process). Let $h(s) = s^2$, and let $G(\tau)$ be the step-function

(2.9)
$$G(\tau) = \begin{cases} 1, & \tau \ge 1/\lambda \\ 0, & \tau < 1/\lambda \end{cases}$$

where, as before, τ is the length of the incubation period, and λ is the infection rate. With h(s) and $G(\tau)$ thus defined, the equation for the generating function becomes

(2.10)
$$\pi\left(s, \frac{k+1}{\lambda}\right) = \pi^2\left(s, \frac{k}{\lambda}\right),$$

with initial condition $\pi(0, s) = s$. In this process the size of the infected population doubles at intervals of time equal to $1/\lambda$.

(ii) The simple birth process (Yule-Furry process). To obtain this well-known process we put $h(s) = s^2$ and $G(t) = 1 - \exp(-\lambda t)$ with $\lambda > 0$. The "state of the system" at time t then depends only on the size of the infected population at that time and is independent of its previous history; hence, the process is of the Markov type. The infection rate λ has the following interpretation: the probability that X(t) will experience a positive unit jump in the interval $(t, t + \delta t)$ is $\lambda \delta t + o(\delta t)$.

The generating function in this case becomes

(2.11)
$$\pi(s, t) = se^{-\lambda \tau} + \int_0^t \pi^2(s, t - \tau) \lambda e^{-\lambda \tau} d\tau,$$

which can be written as the differential equation

$$\frac{\partial \pi}{\partial t} = \lambda \pi (\pi - 1) .$$

The solution to (2.12) and the associated probability distribution are given and discussed by Kendall [11]

(iii) The simple birth-and-death process (Feller, Kolmogorov, Arley, etc.) We have now $G(t) = 1 - \exp[-(\lambda + \mu)t]$, where λ is defined as in (ii), and $\mu \delta t + o(\delta t)$ is the probability that X(t) will experience a negative unit jump in the interval $(t, t + \delta t)$. The generating function satisfies the equation

(2.13)
$$\pi(s, t) = se^{-(\lambda+\mu)t} + \int_0^t [\lambda \pi^2(s, t-\tau) + \mu] e^{-(\lambda+\mu)\tau} d\tau,$$

which can be written as the differential equation

(2.14)
$$\frac{\partial \pi}{\partial t} = \lambda \pi^2 - (\lambda + \mu)\pi + \mu .$$

A detailed discussion of this process is also given in Kendall [11].

(iv) A modified birth-and-death process. This case, which we treat in some detail, assumes $G(t) = 1 - \exp(-\lambda t)$ with $\lambda > 0$ and $h(s) = q_0 + q_1 s + q_2 s^2$. This definition of h(s) states that an individual infected for a period of length τ has probability q_0 of dying and being removed from the infected population, probability q_1 of not infecting anyone else and remaining infected himself, and probability q_2 of infecting one susceptible and remaining infected himself. The q_2 can also be interpreted as the probability that the infected individual will infect two susceptibles, and then be removed from the population. In either case, the size of the infected population increases by one.

The generating function in this case satisfies the equation

(2.15)
$$\pi(s, t) = se^{-\lambda t} + \int_0^t \sum_{n=0}^2 [q_n \pi^n(s, t - \tau)] \lambda e^{-\lambda t} d\tau ,$$

which, as in the previous cases, can be reduced to the differential equation

(2.16)
$$\frac{1}{\lambda} \frac{\partial \pi}{\partial t} = q_2 \pi^2 - (q_0 + q_2) \pi + q_0.$$

The solution to (2.16) with initial condition $\pi(0, s) = s$ is

(2.17)
$$\pi(s, t) = \frac{q_0 - q_2[(q_2s - q_0)/(q_2s - q_2)]e^{\lambda(q_0 - q_2)t}}{q_2\{1 - [(q_2s - q_0)/(q_2s - q_2)]e^{\lambda(q_0 - q_2)t}\}}.$$

Expanding (2.17) as a power series in s, and using (2.1) we obtain

(2.18)
$$p(x, t) = \left[\frac{A}{C} + \frac{B}{D}\right] \left[\frac{D}{C}\right]^{x}, \qquad x \ge 1$$

where

$$A = q_0 q_2 e^{\lambda(q_0 - q_2)t} - q_0 q_2$$

$$B = q_0 q_2 - q_2^2 e^{\lambda(q_0 - q_2)t}$$

$$C = q_0 q_2 e^{\lambda(q_0 - q_2)t} - q_2^2$$

$$D = q_2^2 e^{\lambda(q_0 - q_2)t} - q_2^2$$

Using (2.8), the integral equation for the expected value of X(t) is

(2.19)
$$E[X(t)] = m(t) = e^{-\lambda t} + (1 - q_0 + q_2) \int_0^t m(t - \tau) \lambda e^{-\lambda \tau} d\tau .$$

This reduces to the differential equation

$$\frac{dm(t)}{dt} = \lambda(q_2 - q_0)m(t)$$

whose solution when X(0) = 1 is

(2.21)
$$m(t) = e^{\lambda (q_2 - q_0) t} .$$

As t approaches infinity we have

(2.22)
$$\lim_{t \to \infty} E[X(t)] = \begin{cases} 0, & q_0 > q_2 \\ 1, & q_0 = q_2 \\ \infty, & q_0 < q_2 \end{cases}$$

It is of interest to determine the probability p(0, t) that at time t all of the infected individuals will be removed from the population. From (2.1) and (2.17) we see that this probability is given by

(2.23)
$$p(0, t) = \frac{A}{C} = \frac{q_0 e^{\lambda (q_0 - q_2) t} - q_0}{q_0 e^{\lambda (q_0 - q_2) t} - q_2}.$$

The probability that all of the infected individuals will eventually be removed is given by

(2.24)
$$\lim_{t\to\infty} p(0, t) = \begin{cases} 1, & q_0 > q_2 \\ q_0/q_2, & q_0 < q_2 \end{cases}$$

Should the epidemic start with n > 1 infected individuals at t = 0, the expected number of infected individuals in the population can be obtained by multiplying (2.21) by n. An explicit expression for p(x, t) can be obtained by expanding $\pi^n(s, t)$ as a power series in s and proceeding as before. We have

(2.25)
$$\pi^n(s, t) = \left(\frac{A + Bs}{C}\right)^n \sum_{x=0}^{\infty} {n \choose x} \left[-\frac{D}{C}\right]^x s^x.$$

Hence

$$(2.26) p(x, t) = \sum_{i=0}^{\infty} {n \choose i} \left[\frac{A}{C} \right]^{n-i} \left[\frac{B}{D} \right]^{i} {n \choose x} \left[-\frac{D}{C} \right]^{x-i}, x \ge n.$$

This distribution is of the negative binomial type. The asymptotic behavior of the mean is the same as before, except that now E[X(t)] = n when $q_0 = q_2$. In addition, we have

$$(2.27) p(0, t) = \left(\frac{A}{C}\right)^n,$$

and

(2.28)
$$\lim_{t \to \infty} p(0, t) = \begin{cases} 1, & q_0 > q_2 \\ (q_0/q_2)^n, & q_0 < q_2. \end{cases}$$

2.3. Remarks on the use of age-dependent theory for more complex situations. In section 2.2 we have given some models which can perhaps be used to discuss some very simple epidemic situations. We should like, however, to treat more realistic situations which would be of greater interest to the epidemiologist. In comparing the classical differential-difference equation approach to the age-dependent approach in the study of stochastic branching processes, we find that the latter has several features (variable incubation period τ , and infection probabilities q_n) which make

it especially suitable for many applications in epidemic theory. On the other hand, we find that the difficulties encountered in solving the Bellman-Harris functional equation (2.2) are very great in all but the relatively simple cases.

One approach to this problem which this author is now considering is based on the relationship between branching processes, random walk on the real line, and diffusion theory. This relationship has been considered by several workers in various problems. The interpretation of an age-dependent branching process as a random walk on the integers $0, 1, 2, \cdots$ is as follows: at t = 0 the particle (the size of the infected population) is in position $k, 0 < k < \infty$. At $t = \tau$ the particle moves n < 0 steps n < 0, n < 0

Associated with every discrete random walk problem is a continuous diffusion problem, which involves the study of the Fokker-Planck equation (the forward equation of diffusion theory). To obtain the Fokker-Planck equation it is necessary to determine the infinitesimal mean displacement and the infinitesimal variance of X(t). In our case these coefficients will obviously depend on the functions h(s) and G(t). Having obtained the diffusion equation, we can then utilize the recent results of Feller [7] to study its properties when various assumptions are made concerning h(s) and G(t).

3. Statistical problems

3.1. Introduction. Following the formulation and study of stochastic models for the spread of epidemics the next step is to consider the problems of statistical inference associated with these models. Since most of the stochastic epidemic models can be formulated as branching processes, the problem facing us is the development of statistical methods for general branching processes and their subsequent application to epidemic situations. It is only recently that research workers have considered the problems of statistical inference arising in the theory of stochastic processes and only a few studies have been devoted to problems associated with branching processes. Interesting results for certain branching processes have been obtained by Anscombe [1], Immel [10], Kendall [11], [12], and Moran [13], [14].

Recently we have been considering the problems of testing hypotheses and estimation for branching processes within the framework of the Wald theory of sequential decision functions [3]. This work can be considered as an extension of some of the results of Dvoretzky, Kiefer, and Wolfowitz [5], [6] to nonstationary stochastic processes. Applications of our results to stochastic epidemics will be given in [4]; however, we will illustrate some of these methods by considering a problem mentioned by Taylor [16] concerning the comparison of two stochastic epidemics.

3.2. Comparison of two stochastic epidemics. Let the random variables $X_1(t)$ and $X_2(t)$ represent the number of infected individuals at time t in two independent populations. Assume the probability law $p_{\lambda_i}(x_i, t) = Pr\{X_{\lambda_i}(t) = x_i\}, i = 1, 2, \text{ is known (or that we have a good model) except for the value of the parameter <math>\lambda_i$, the infection rate. Suppose now that the epidemiologist observes continuously the de-

velopment of the two epidemics, and wishes, on the basis of the observed sample functions $x_1(t)$ and $x_2(t)$, to test the hypothesis H_1 that $\lambda_1 < \lambda_2$ against the alternative H_2 that $\lambda_1 > \lambda_2$. This type of problem has been treated by Girshick [8] (see also Wald [17]) for some discrete distributions. We now consider the application of the Girshick method to two cases: (i) the spread of the epidemic can be described by a simple birth process with unknown infection rate λ ; and (ii) the epidemic can be described by a simple birth-and-death process with unknown infection rate λ and known parameter μ , which we shall interpret to be the rate at which members of the population arrive at the hospital to report infection.

(i) Simple birth process. To test the hypothesis H_1 , select two positive values L and M, where L < M. Denote by \bar{H}_1 the hypothesis that the joint distribution of x_1 and x_2 is given by $p_L(x_1, t)p_M(x_2, t)$, and denote by \bar{H}_2 the hypothesis that the joint distribution is given by $p_M(x_1, t)p_L(x_2, t)$. We can now set up the Wald sequential probability ratio test for testing the simple hypothesis \bar{H}_1 against the simple alternative \bar{H}_2 . The original hypothesis H_1 will be accepted or rejected depending on whether the Wald test leads to the acceptance or rejection of \bar{H}_1 .

The decision function (given by the probability ratio) is defined as

(3.1)
$$d(t) = \log \frac{p_M(x_1, t)p_L(x_2, t)}{p_L(x_1, t)p_M(x_2, t)}.$$

For the simple birth process

(3.2)
$$p_{\lambda}(x, t) = e^{-\lambda t} (1 - e^{-\lambda t})^{x-1}, \quad t \ge 0, \quad x \ge 1, \quad X(0) = 1;$$

hence

(3.3)
$$d(t) = [x_2(t) - x_1(t)] \log \frac{1 - e^{-Lt}}{1 - e^{Mt}}.$$

We now select two constants $A = \log [(1 - \alpha_2)/\alpha_1]$, and $B = \log [\alpha_2/(1 - \alpha_1)]$, where α_i is the probability of accepting \bar{H}_i when it is false. The sequential test is performed as follows: The two epidemics are observed continuously, and if at any time t = T we observe $d(T) \leq B$ we conclude that $\lambda_1 > \lambda_2$. If $d(T) \geq A$ we conclude $\lambda_1 < \lambda_2$. If neither holds we continue to observe.

It has been shown by Girshick [8] that when the distribution being studied admits a sufficient statistic for the unknown parameter there exists a function $\gamma = (\lambda_1, \lambda_2)$, say, such that the probability that the sequential probability ratio test will terminate with the acceptance of H_1 depends only on the value of γ . The function γ satisfies the conditions (1) $\gamma(\lambda_1, \lambda_2) = 0$ when $\lambda_1 = \lambda_2$, (2) $\gamma(\lambda_1, \lambda_2) < 0$ when $\lambda_2 > \lambda_1$, (3) $\gamma(\lambda_1, \lambda_2) = -\gamma(\lambda_2, \lambda_1)$. By the Neyman-Fisher factorization theorem x(t) is sufficient for λ ; and following Girshick we find that in this case

(3.4)
$$\gamma(\lambda_1, \lambda_2) = \log \frac{1 - e^{-\lambda_1 t}}{1 - e^{-\lambda_2 t}}.$$

We can now define the decision boundaries as

(3.5)
$$A^*(t) = \log \frac{1 - \alpha_2}{\alpha_1} / \log \frac{1 - e^{-Lt}}{1 - e^{-Mt}},$$

(3.6)
$$B^*(t) = \log \frac{\alpha_2}{1 - \alpha_1} / \log \frac{1 - e^{-Lt}}{1 - e^{-Mt}}.$$

The corresponding decision function is simply

$$d^*(t) = x_2(t) - x_1(t) .$$

This sequential test can be carried out graphically in the usual manner.

(ii) Simple birth-and-death process. We now consider the case where the two epidemics can be described by the simple birth-and-death process, that is,

(3.8)
$$P_{\lambda,\mu}(x, t) = \psi_{\lambda,\mu}(t)[1 - \Phi_{\lambda,\mu}(t)][\Phi_{\lambda,\mu}(t)]^{x-1}, \qquad x \ge 1$$

where

(3.9)
$$\psi_{\lambda,\mu}(t) = \frac{\mu e^{(\lambda-\mu)t} - \mu}{\lambda e^{(\lambda-\mu)t} - \mu}, \qquad \Phi_{\lambda,\mu}(t) = \frac{\lambda e^{(\lambda-\mu)t} - \lambda}{\lambda e^{(\lambda-\mu)t} - \mu}.$$

The infection rate λ is the unknown parameter and μ is the known parameter, which we interpret as the rate at which individuals arrive at the hospital to report infections. We should like to decide on the basis of the observed sample functions whether $\lambda_1 < \lambda_2$ or $\lambda_1 > \lambda_2$.

The procedure is the same as before. For constants A, B, and known parameters μ_1 and μ_2 we compute, using, (3.7), the decision function

(3.10)
$$d(t) = [x_2(t) - x_1(t)] \log \frac{\Phi_L(t)}{\Phi_M(t)};$$

and if for any t = T, $d(T) \leq B$ we conclude $\lambda_1 > \lambda_2$. If $d(T) \geq A$ we conclude $\lambda_1 < \lambda_2$. If neither holds, we continue to observe. If we assume $\mu_1 = \mu_2$ the function

(3.11)
$$\gamma(\lambda_1, \lambda_2) = \log \frac{\Phi_{\lambda_1}(t)}{\Phi_{\lambda_2}(t)}$$

satisfies the required conditions; and we obtain the new decision boundaries

$$(3.12) A^*(t) = \log \frac{1-\alpha_2}{\alpha_1} / \log \frac{\Phi_L(t)}{\Phi_M(t)},$$

(3.13)
$$B^{*}(t) = \log \frac{\alpha_{2}}{1 - \alpha_{1}} / \log \frac{\Phi_{L}(t)}{\Phi_{M}(t)},$$

together with the new decision function

(3.14)
$$d^*(t) = x_2(t) - x_1(t).$$

We remark in closing that these methods go through for the death process, Pólya process, and many other branching processes which are of interest in applications.

REFERENCES

- F. J. Anscombe, "Sequential estimation," Jour. Roy. Stat. Soc., Ser. B, Vol. 15 (1953), pp. 1-29.
- [2] R. Bellman and T. E. Harris, "On age-dependent binary branching processes, Annals of Math., Vol. 55 (1952), pp. 280-295.
- [3] A. T. Bharucha-Reid, "Sequential decision problems for branching stochastic processes," to be published.
- [4] ——, "The Stochastic Theory of Epidemics and Related Statistical Problems," to be published by the USAF School of Aviation Medicine, Randolph Field, Texas.
- [5] A. DVORETZKY, J. KIEFER, and J. WOLFOWITZ, "Sequential decision problems for processes with continuous time parameter. Testing hypotheses," *Annals of Math. Stat.*, Vol. 24 (1953), pp. 254–264.
- [6] ——, "Sequential decision problems for processes with continuous time parameter. Problems of estimation," Annals of Math. Stat., Vol. 24 (1953), pp. 403-405.
- [7] W. Feller, "The parabolic differential equations and the associated semi-groups of transformations," *Annals of Math.*, Vol. 55 (1952), pp. 468-519.
- [8] M. A. GIRSHICK, "Contributions to the theory of sequential analysis: I," Annals of Math. Stat., Vol. 17 (1946), pp. 123-143.
- [9] T. E. Harris, "Some mathematical models for branching processes," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, 1951, pp. 305-328.
- [10] E. R. Immel, "Problems of estimation and hypothesis testing in connection with birth-and-death stochastic processes," (abstract), Annals of Math. Stat., Vol. 22 (1951), p. 485.
- [11] D. G. KENDALL, "Stochastic processes and population growth," Jour. Roy. Stat. Soc., Ser. B, Vol. 11 (1949), pp. 230-264.
- [12] ——, "Les processus stochastiques de croissance en biologie," Ann. de l'Inst. Henri Poincaré, Vol. 13 (1952), pp. 43-108.
- [13] P. A. P. Moran, "Estimation methods for evolutive processes," Jour. Roy. Stat. Soc., Ser. B, Vol. 13 (1951), pp. 141-146.
- [14] ______, "The estimation of the parameters of a birth and death process," Jour. Roy. Stat. Soc., Ser. B, Vol. 15 (1953), pp. 241-245.
- [15] R. E. Serfling, "Historical review of epidemic theory," Human Biology, Vol. 24 (1952), pp. 145-166.
- [16] W. F. TAYLOR, "Problems in contagion," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4, Berkeley and Los Angeles, University of California Press, 1956, pp. 167-179.
- [17] A. Wald, Sequential Analysis, New York, John Wiley and Sons, 1947.