ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIII, 2 (1972)

R. BARTOSZYŃSKI (Warszawa)

ON A CERTAIN MODEL OF AN EPIDEMIC

1. Introduction and intuitive background. In this paper we shall consider a certain model of an epidemic. Roughly speaking, the situation which we shall be investigating is the following. Consider a population consisting of a large (assumed infinite in the model) number of families. Assume that if at a certain moment t an infection occurs within a given family, it originates an epidemic within this family, i.e., it originates a stochastic process $\xi_0, \xi_1, \ldots$, where ξ_n is interpreted as the number of individuals in this family who are infectious at moment $t+n+1$. At every moment t there is then a certain number of infectives in the population equal to the sum of the respective values of the stochastic processes in the families already infected. Each of these infectives infects a certain number of new families. We assume this number to have the Poisson distribution with mean λ, and we assume that the numbers of new families infected by different individuals are independent. We also assume independence of epidemics within different families. The whole description of the process depends then on the value of λ and on specific assumptions about probability distributions for the stochastic processes representing the epidemics within families.

We shall investigate conditions under which the epidemic considered expires with probability one, and we shall study two stochastic processes, defined, loosely speaking, as the number of freshly infected families at times $t = 0, 1, 2, \ldots$ and the number of infectives at times $t = 1, 2, \ldots$

We shall try to make such specific assumptions about probabilistic mechanisms of our process which can be expressed in terms of quantities estimable from empirical data. This will allow us to verify the assumptions of the model, and, if these tests do not lead to negative results, possibly to quantitative predictions.

The investigations of this paper have been inspired by Dr. Juan Angulo from the School of Medicine, Emory University, Atlanta, Georgia, who suggested to me working up a model for the epidemic of variola minor which occurred in 1963 in Sao Paolo. The reasonably complete data on

this epidemic make it possible to verify the assumptions of the model presented. The results of these tests, if positive, will be published elsewhere. The present paper is of purely theoretical character.

2. Formal definition of the model and outline of the contents of the paper. In this section we outline the formal construction of the sample space and probability distribution on it, and define the relevant events and random variables.

Let $\Theta = \{\theta_1, \theta_2, \ldots\}$ be a finite or countable set, and let π be a probability distribution on Θ. We write $\pi_j = \pi(\theta_j)$. Assume that to each θ_j with $\pi_j > 0$ there corresponds a sequence

$$\{p_{j,n}\}, \quad n = 0, 1, 2, \ldots, \tag{2.1}$$

where each $p_{j,n}$ is a probability distribution concentrated on the set of all $(n+1)$-tuples $(r_0, r_1, \ldots, r_n)$ of non-negative integers. Moreover, assume that the distributions (2.1) are consistent in the sense that

$$\sum_{r_n} p_{j,n}(r_0, \ldots, r_{n-1}, r_n) = p_{j,n-1}(r_0, \ldots, r_{n-1}) \tag{2.2}$$

for all j with $\pi_j > 0$, all $n = 1, 2, \ldots$ and all $r_0, \ldots, r_{n-1}$ for which the right-hand side of (2.2) is positive.

By the well-known theorem of Kolmogorov (see, for instance, [4]) there exist processes (integer valued), to be denoted generally by $\xi_0, \xi_1, \ldots$, with joint probability distributions given by (2.1); we interpret these processes as representing histories of epidemics within families, the parameter θ_j reserved for representing some observable characteristics of families such as their sizes, sanitary conditions, etc.

We denote by

$$p_n(r_0, \ldots, r_n) = \sum_j \pi_j p_{j,n}(r_0, \ldots, r_n) \tag{2.3}$$

the unconditional distribution for the stochastic process $\{\xi_n\}$ of epidemics within families. Thus, (2.3) represents the probability distribution of initial generations of an epidemic within a randomly selected family, where by *random selection* we mean the selection of parameter θ_j according to the probability distribution π.

Let us denote by Ω, our sample space, the set of all arrays ω of the form

$$\begin{array}{llll} x_0^{(0)}, & x_1^{(0)}, & x_2^{(0)}, & x_3^{(0)}, \quad \ldots, \\ & x_0^{(1,1)}, & x_1^{(1,1)}, & x_2^{(1,1)}, \quad \ldots, \\ & x_0^{(1,2)}, & x_1^{(1,2)}, & x_2^{(1,2)}, \quad \ldots, \\ & \ldots\ldots\ldots\ldots\ldots\ldots & & \end{array} \tag{2.4}$$

$$
\begin{array}{llll}
\dots & \dots & \dots & \dots \\
x_0^{(1,k_1)}, & x_1^{(1,k_1)}, & x_2^{(1,k_1)}, & \dots, \\
& x_0^{(2,1)}, & x_1^{(2,1)}, & \dots, \\
& \dots & \dots & \dots \\
& x_0^{(2,k_2)}, & x_1^{(2,k_2)}, & \dots, \\
& \dots & \dots & \dots
\end{array}
$$

where $k_1, k_2, \dots$ and all entries are non-negative integers.

Given an array ω of form (2.4), put $m_n(\omega)$ equal to the sum of all entries in the n-th column, i.e.,

$$
\begin{aligned}
m_0(\omega) &= x_0^{(0)}, \\
m_n(\omega) &= x_n^{(0)} + (x_{n-1}^{(1,1)} + \dots + x_{n-1}^{(1,k_1)}) + \dots + (x_0^{(n,1)} + \dots + x_0^{(n,k_n)}).
\end{aligned} \tag{2.5}
$$

We define the class $\mathcal{G}$ of measurable subsets of Ω in an usual way, as the smallest σ-field containing all cylinders with finite bases. To define the probability on $(\Omega, \mathcal{G})$ it suffices to determine its values on a special class of cylinders, namely on cylinders obtained by specifying numbers $k_1, \dots, k_n$ and the values in first $n+1$ columns of array (2,4). We set this value to be

$$
p_n(x_0^{(0)}, \dots, x_0^{(0)}) \prod_{t=1}^{n} \frac{(\lambda m_{t-1})^{k_t}}{k_t!} \exp[-\lambda m_{t-1}] \times \prod_{s=1}^{k_t} p_{n-t}(x_0^{(t,s)}, \dots, x_{n-t}^{(t,s)}), \tag{2.6}
$$

where λ is a positive constant and m_t is defined by (2.5). In intuitive terms, this assumption means that the rows of the array (2.4) represent independent realizations of stochastic processes with distributions (2.3), while the numbers k_n of new rows originating in the n-th column have the Poisson distribution with mean λm_{n-1}.

Let $\mathcal{E}$ be the event, described picturesquely as "extinction" of the epidemic, defined formally as the set of all arrays (2.4) which have only a finite number of rows, each of them containing only a finite number of non-zero terms. In section 3 we shall study conditions under which $P(\mathcal{E}) = 1$, i.e., conditions under which the epidemic considered is bound to expire with probability one.

Now, to each array ω of form (2.4) we may assign two numerical sequences

$$
1, k_1(\omega), k_2(\omega), \dots \tag{2.7}
$$

and

$$m_0(\omega), m_1(\omega), m_2(\omega), \ldots \tag{2.8}$$

Relations (2.7) and (2.8) define two stochastic processes, to be denoted by $U_0, U_1, \ldots$ and $V_0, V_1, \ldots$, respectively, where U_n is interpreted as the number of new families infected at time n, and V_n is interpreted as the total number of infectives at time $n+1$. In section 4 we derive formulas for probability generating functions for processes $\{U_n\}$ and $\{V_n\}$.

Since the behaviour of our epidemic depends on particular properties of the processes $\{\xi_n\}$ for various values of θ, in section 5 we study sets of assumptions implying properties of these processes relevant to the application of theorems of sections 3 and 4, in particular those implying the finiteness of the sum $\xi_0+\xi_1+\xi_2+\ldots$ and finiteness of the expectation of this sum.

Finally, in section 6 we give an outline of various schemes of interpretation of our process in terms of actual epidemics, and discuss briefly the problem of estimation of the relevant parameters.

To make the presentation simpler, we refer rather to the intuitive interpretation than to the formal definition of our process.

3. Theorems on extinction. We start from conditions under which our epidemic will expire with probability one, i.e., conditions implying $\mathrm{P}(\mathscr{E}) = 1$. Given a family, say F, call all families infected by infectives from family F the *direct descendants* of F. Clearly, this defines a simple Galton-Watson branching process (see, for instance, [3]) of multiplication of "particles", the latter in this case being families. The process considered will expire with probability one if and only if the expected number of direct descendants of a single family is one or less. Now, this expected number equals λh, where $h = \mathrm{E}(\xi_0+\xi_1+\ldots)$, finite or not. Finiteness of h implies that the sum $\xi_0+\xi_1+\ldots$ must be finite with probability one, i.e., that every row in the array ω contains only a finite number of non-zero terms. We may state, therefore,

THEOREM 3.1. *If* $h = \mathrm{E}(\xi_0+\xi_1+\ldots) < \infty$ *and* $\lambda \leqslant 1/h$, *then* $\mathrm{P}(\mathscr{E}) = 1$, *i.e., the epidemic is bound to expire with probability one.*

To derive a method of computing the value of $\mathrm{P}(\mathscr{E})$ when it is less than one, we may explore further our interpretation in terms of the simple Galton-Watson process. Let

$$q_j(r) = \mathrm{P}_j\{\xi_0+\xi_1+\ldots = r\}, \tag{3.1}$$

where P_j is the probability distribution for the process $\{\xi_n\}$ conditional upon the value θ_j. We have $\sum_r q_j(r) \leqslant 1$, the difference $1-\sum_r q_j(r)$ representing the probability, conditional upon θ_j, that the sum $\xi_0+\xi_1+\ldots$

will be infinite. Define

$$Q_j(s) = \sum_r s^r q_j(r) \tag{3.2}$$

and

$$Q(s) = \sum_j \pi_j Q_j(s). \tag{3.3}$$

We have $Q(1) = 1$ if and only if $Q_j(1) = 1$ for all j such that $\pi_j > 0$, the condition $Q_j(1) = 1$ being equivalent to

$$\sum_r q_j(r) = \mathrm{P}_j\{\xi_0 + \xi_1 + \ldots < \infty\} = 1.$$

Now, $h = \mathrm{E}(\xi_0 + \xi_1 + \ldots)$ is infinite if $Q(1) < 1$, and equals $Q'(1)$, finite or not, otherwise.

We now try to find the probability generating function, possibly defective, for the number of families infected by all infectives of a given family. Denote this random variable by Z, and its probability generating function by $z(s)$. We have then

$$z(s) = \sum_k s^k \mathrm{P}(Z = k) = \sum_k s^k \sum_j \pi_j \mathrm{P}_j(Z = k)$$

$$= \sum_k s^k \sum_j \pi_j \sum_r q_j(r) \mathrm{P}_j(Z = k \mid \xi_0 + \xi_1 + \ldots = r).$$

By definition, probability under the last sum equals $(\lambda r)^k e^{-\lambda r}/k!$, hence (interchanging the order of summation)

$$z(s) = \sum_j \pi_j \sum_r q_j(r) \sum_k \frac{(\lambda s r)^k}{k!} e^{-\lambda r} = \sum_j \pi_j \sum_r q_j(r)\, e^{-\lambda r(s-1)}$$

$$= \sum_j \pi_j Q_j(e^{\lambda(s-1)}) = Q(e^{\lambda(s-1)})$$

with $Q(s)$ given by (3.3).

Next, the basic theorem on the probability of extinction of the Galton-Watson branching process remains valid in the case where the probability generating function is defective, i.e., if it does not assume the value 1 at $s = 1$. We may state, therefore, a somewhat more general result than theorem 3.1, namely

THEOREM 3.2. *The epidemic considered is bound to expire with probability one if and only if* $Q(1) = 1$, $Q'(1) = h < \infty$ *and* $\lambda \leqslant 1/h$. *In every case, the probability of extinction equals to the least positive root of the equation* $x = Q(e^{\lambda(x-1)})$, *where* $Q(s)$ *is given by* (3.3).

4. Processes $\{U_n\}$ and $\{V_n\}$. We start with an investigation of the process $\{U_n\}$ defined in section 2, where U_n is interpreted as the number of freshly infected families at time n. Put

$$f_n(s) = \sum_k s^k \mathrm{P}(U_n = k) \tag{4.1}$$

and write

$$K_n(s_0, \ldots, s_n) = \sum_j \pi_j K_{j,n}(s_0, \ldots, s_n), \tag{4.2}$$

where

$$K_{j,n}(s_0, \ldots, s_n) = \sum_{r_0, \ldots, r_n} s_0^{r_0} \ldots s_n^{r_n} p_{j,n}(r_0, \ldots, r_n) \tag{4.3}$$

is the joint probability generating function of the distribution (2.1).

THEOREM 4.1. *The probability generating functions $f_0(s), f_1(s), \ldots$ are given by the recursive formulas*

$$f_0(s) = s, \tag{4.4}$$

$$f_n(s) = K_{n-1}\{\exp[\lambda(f_{n-1}(s)-1)], \ldots, \exp[\lambda(f_0(s)-1)]\} \tag{4.5}$$

for $n = 1, 2, \ldots$

Proof. Relation (4.4) follows directly from the fact that we have $U_0 = 1$ by definition (2.7). Let now $n > 0$ and let us write the expression for $f_n(s)$ by conditioning it upon the history in the original process (to be denoted by $\xi_0^{(0)}, \xi_1^{(0)}, \ldots$). If $\xi_0^{(0)} = r_0$, $\xi_1^{(0)} = r_1, \ldots, \xi_{n-1}^{(0)} = r_{n-1}$, then at time k (where $k = 1, 2, \ldots, n$) we have r_{k-1} infectives in the original family, and they originate i_{k-1} epidemics within new families; the contribution to U_n from all these families has then the probability generating function $f_{n-k}^{i_{k-1}}(s)$. Using the fact that i_k has the Poisson distribution with mean λr_k, we may write

$$\begin{aligned}
f_n(s) &= \sum_j \pi_j \sum_{r_0, \ldots, r_{n-1}} p_{j,n-1}(r_0, \ldots, r_{n-1}) \times \\
&\quad \times \sum_{i_0, \ldots, i_{n-1}} \frac{(\lambda r_0)^{i_0}}{i_0!} \ldots \frac{(\lambda r_{n-1})^{i_{n-1}}}{i_{n-1}!} \exp[-\lambda(r_0 + \ldots + r_{n-1})] \times \\
&\quad \times f_{n-1}^{i_0}(s) \ldots f_0^{i_{n-1}}(s) \\
&= \sum_j \pi_j \sum_{r_0, \ldots, r_{n-1}} p_{j,n-1}(r_0, \ldots, r_{n-1}) \times \\
&\quad \times \exp[\lambda r_0(f_{n-1}(s)-1)] \ldots \exp[\lambda r_{n-1}(f_0(s)-1)] \\
&= K_{n-1}\{\exp[\lambda(f_{n-1}(s)-1)], \ldots, \exp[\lambda(f_0(s)-1)]\}, \qquad \text{q.e.d.}
\end{aligned}$$

Write now

$$h_k = \mathrm{E}\,\xi_k = \sum_j \pi_j \sum_{r_0,\ldots,r_k} r_k p_{j,k}(r_0, \ldots, r_k), \tag{4.6}$$

finite or not. By formal differentiation at $s = 1$ we obtain from theorem 4.1 the following theorem giving recursive formulas for expectations $a_n = \mathrm{E}(U_n) = f_n'(1)$, finite or not:

THEOREM 4.2. *The expectations $a_0, a_1, \ldots$ are given by the recursive formulas*

$$a_0 = 1,$$
$$a_n = \lambda(a_{n-1}h_0 + \ldots + a_0 h_{n-1})$$

with h_k defined by (4.6).

We now prove an analogous result for the process $\{V_n\}$, where V_n is the total number of infectives at time $n+1$. We have

THEOREM 4.3. *The probability generating functions $g_n(s)$ of the random variables V_n are given by the recursive formulas*

$$g_0(s) = K_0(s), \tag{4.7}$$

$$g_n(s) = K_n\{\exp[\lambda(g_{n-1}(s)-1)], \ldots, \exp[\lambda(g_0(s)-1)], s\} \tag{4.8}$$

for $n = 1, 2, \ldots$

Proof. The proof is similar to that of theorem 4.1. Relation (4.7) follows directly from the definition. Let $n > 0$ and suppose that $\xi_0^{(0)} = r_0, \ldots, \xi_n^{(0)} = r_n$. The value of V_n is then equal to the sum of r_n and the contributions from families infected at earlier moments. If at time k ($k = 1, 2, \ldots, n-1$) there were r_{k-1} infectives in the original family, and they infected altogether i_{k-1} new families, then the contribution to V_n from these families has probability generating function $g_{n-k}^{i_{k-1}}(s)$. We may write, therefore,

$$\begin{aligned} g_n(s) &= \sum_j \pi_j \sum_{r_0,\ldots,r_n} p_{j,n}(r_0, \ldots, r_n) \times \\ &\quad \times \sum_{i_0,\ldots,i_{n-1}} \frac{(\lambda r_0)^{i_0}}{i_0!} \cdots \frac{(\lambda r_{n-1})^{i_{n-1}}}{i_{n-1}!} \exp[-\lambda(r_0 + \ldots + r_{n-1})] \times \\ &\quad \times s^{r_n} g_{n-1}^{i_0}(s) \ldots g_0^{i_{n-1}}(s) \\ &= \sum_j \pi_j \sum_{r_0,\ldots,r_n} p_{j,n}(r_0, \ldots, r_n) \times \\ &\quad \times \exp[\lambda r_0(g_{n-1}(s)-1)] \ldots \exp[\lambda r_{n-1}(g_0(s)-1)] s^{r_n} \\ &= K_n\{\exp[\lambda(g_{n-1}(s)-1)], \ldots, \exp[\lambda(g_0(s)-1)], s\}, \qquad \text{q.e.d.} \end{aligned}$$

Differentiating formally at $s = 1$, we obtain for the expected values $b_n = \mathrm{E}(V_n) = g_n'(1)$, finite or not, the following

THEOREM 4.4. *The expectations $b_0, b_1, \ldots$ are given by the recursive formulas*

$$b_0 = K_0'(1),$$

$$b_n = h_n + \lambda(h_{n-1}b_0 + \ldots + h_0 b_{n-1})$$

for $n = 1, 2, \ldots$, where h_n is defined by (4.6).

5. Epidemics within families. The theorems of sections 3 and 4 show that the behaviour of our epidemic depends on the properties of the process $\{\xi_n\}$, in particular on the almost sure finiteness of the sum $\xi_0 + \xi_1 + \ldots$ and on the finiteness of the expectation of this sum. We now try to impose such conditions on the process $\{\xi_n\}$ which imply the above-mentioned two properties, and which might be reasonably expected to be satisfied in the practical situations.

Firstly, it appears reasonable to assume that whatever the value θ_j, once a sufficiently long stretch of consecutive zeros occurs in the process $\{\xi_n\}$, all subsequent values must be equal to zero. Formally, assume that

(i) there exists an integer $d \geqslant 1$ such that for all j and n

$$\mathrm{P}_j(\xi_{n+d} = 0 \mid \xi_n = \xi_{n+1} = \ldots = \xi_{n+d-1} = 0) = 1. \tag{5.1}$$

Denote by B_n the event under the condition in (5.1), i.e., let B_n be the event $\xi_n = \xi_{n+1} = \ldots = \xi_{n+d-1} = 0$, where d is the integer appearing in (i).

Next, it appears reasonable to assume that the conditional probability of B_n, given the values $r_0, \ldots, r_{n-1}$, should be, in some sense, bounded from below by a positive number. More precisely, we assume that

(ii) for every j there exists a positive number c_j such that for all n

$$\mathrm{P}_j(B_n \mid \xi_{n-1} = r_{n-1}, \ldots, \xi_{n-d} = r_{n-d}, \xi_{n-d-1} = r_{n-d-1}, \ldots, \xi_0 = r_0) \geqslant c_j^{\max(r_{n-1}, \ldots, r_{n-d})},$$

whatever the values $r_{n-d-1}, \ldots, r_0$.

We may now prove

THEOREM 5.1. *If the process $\{\xi_n\}$ satisfies conditions* (i) *and* (ii), *then for every j we have*

$$\mathrm{P}_j(\limsup \xi_n = \infty \text{ or } \lim \xi_n = 0) = 1.$$

The proof is analogous to that of theorem 5.1 in [2], and will be omitted. Intuitively, the idea of the proof is the following: whatever the value $R > 0$, the occurrence of a stretch $r_{n-d}, r_{n-d+1}, \ldots, r_{n-1}$ with

$\max(r_{n-d}, \ldots, r_{n-1}) = R$ involves a positive "risk" of the event B_n, and, consequently, of the event $\xi_n = \xi_{n+1} = \ldots = 0$, hence $\lim \xi_n = 0$. Thus, if the process does not expire, no stretch of length d with a bounded maximum can occur infinitely often, hence the sequence of maxima extended over stretches of the length d must grow indefinitely. Note that if $d = 1$, we may replace lim sup by lim in theorem 5.1.

We may now impose another natural condition, corresponding to the requirement that at no time the value of stochastic process $\{\xi_n\}$ can exceed the size of the family. Formally, assume

(iii) to each j there corresponds a natural number $N(j)$ such that

$$\mathrm{P}_j(\xi_n \leqslant N(j)) = 1 \quad \text{for all } n = 0, 1, 2, \ldots$$

Clearly, (iii) rules out the possibility of $\limsup \xi_n = \infty$, and we may state

THEOREM 5.2. *Under conditions* (i)-(iii), *we have*

$$\mathrm{P}_j(\xi_0 + \xi_1 + \ldots < \infty) = 1 \quad \textit{for all } j.$$

Assume now that conditions (i)-(iii) hold, and write $v_j = c_j^{N(j)} > 0$. Let us fix θ_j and consider consecutive stretches of length d, i.e., $(\xi_0, \ldots, \xi_{d-1})$, $(\xi_d, \ldots, \xi_{2d-1}), \ldots$ Conditions (ii) and (iii) imply that the probability of a given stretch (except possibly the initial one) consisting of zeros only is at least v_j, whatever the past history of the process in the preceding stretches. Condition (i) implies that once a stretch will consist of zeros only, all subsequent stretches will have the same property. Thus, the probability that the process will assume positive values in at least $k+1$ stretches is at most $(1-v_j)^k$. Now, the sum of values in each stretch is, by (iii), bounded from above by $dN(j)$. Summing up the corresponding geometric series, we obtain

THEOREM 5.3. *Under conditions* (i)-(iii), *the expectation* $\mathrm{E}_j(\xi_0 + \xi_1 + \ldots)$ *is finite for every* j, *and is bounded from above by* $dN(j)/v_j^2$.

We may now formulate a sufficient condition for the finiteness of $h = \mathrm{E}(\xi_0 + \xi_1 + \ldots)$:

THEOREM 5.4. *Under conditions* (i)-(iii), *if the series*

$$\sum_j \pi_j N(j)/v_j^2$$

converges, then $h = \mathrm{E}(\xi_0 + \xi_1 + \ldots)$ *is finite.*

6. Interpretations, discussion and some particular cases. The model considered in the preceding section admits several interpretations. Firstly, it ought to be pointed out that the term "family" as used above does not have to be treated literally: in practical applications we may identify

"families" as considered in the model with, say, inhabitants of one house, whether related by blood or marriage or not, or groups of children who habitually play together (a class in school, say), etc.

Next, parameter θ is to be identified with such observable characteristics of families which have, or are thought to have, influence on the course of an epidemic within a family. The most obvious and easily identifiable such parameter is the size of the family. If the considered population is sufficiently "homogeneous", it may be taken as the only parameter, i.e. we may put $\theta_j = j$ for families of size j. In this case π_j will simply be equal to the proportion of families of size j in the considered population, an easily accessible quantity.

In general, one could incorporate into θ some other characteristics such as an appropriate index indicating sanitary conditions, etc. Then θ would be a vector-valued parameter, one coordinate showing the size of the family, and the remaining ones showing some other characteristics (not necessarily numerical). One of the problems would then be to evaluate π_j, i.e., the proportion of families in the considered population with the particular value θ_j of the parameter.

The other, much more serious, problem concerns the question of estimation of probability distribution for epidemics within families for given value of θ, i.e., probabilities $p_{j,n}(r_0, \ldots, r_n)$ for the vector $(r_0, \ldots, r_n)$ given $\theta = \theta_j$.

There exist at least two ways of interpreting the process $\{\xi_n\}$ representing the epidemics within the families. Thus, we may select an appropriate unit of time, say, one day, and interpret ξ_n as the number of individuals in the family who are infectious on day $n+1$ after the initial infection. Under this interpretation (to be referred to as scheme I), a typical history of the process $\{\xi_n\}$ would start with an initial stretch of zeros (the period of incubation of the initial infective), followed by a stretch of positive terms, then again a stretch of zeros, and so on. The characteristic feature of the process $\{\xi_n\}$ under this scheme of interpretation is that the sum $\xi_0 + \xi_1 + \ldots$ may exceed the size of the family, as each infective is counted once for every day when he is infectious.

The difficulty in applying this scheme of interpretation in practice lies mainly in the difficulty of estimating the joint probability distribution for the process $\{\xi_n\}$ from the empirical data.

A more promising interpretation of the process $\{\xi_n\}$ is the following: if the disease under consideration has a relatively constant latency period, and a relatively short period of infectiousness, we may idealize it by assuming that the latency period is constant and the period of infectiousness infinitely short. If we measure time in units equal to the latency period, we may interpret ξ_n as the number of members of the family who became

infected n latency periods after the initial infection (they become then infectious still one more latency period later, consistently with the general interpretation of stochastic process $\{\xi_n\}$).

We refer to this interpretation as to scheme II (note that under this scheme it appears reasonable to put $d = 1$ in conditions (i) and (ii) of section 5).

The main advantage of scheme II over scheme I is that one can easily obtain (at least crude) estimate of λ. In fact, if at some time, K cases of a given disease occurred and about one period latency later new cases were reported in L families previously unaffected, then the ratio L/K may serve as a crude estimate of the between-families infection rate λ.

The second advantage of scheme II lies in the fact that one can design several plausible models of epidemics within a family. We briefly describe these models in the sequel; in all of these models we treat the size j of the family as an initial parameter, denoting (whenever applicable) the remaining components of the parameter θ as θ^*.

By far the simplest model is obtained if we assume that the disease is "absolutely contagious". Under this assumption, given the size of the family to be j and given that $0 < \xi_0 = k < j$, we have $\xi_1 = j-k$, $\xi_2 = \xi_3 = \ldots = 0$. A reasonable approximation to reality might be $\xi_0 = 1$ whatever the value of j, that is to say, every epidemic within a family always starts from one infective, and one latency period later all the remaining members of the family become infected.

To obtain a somewhat more realistic model we may assume, in addition, that some members of the family may be immune (due, say, to vaccination), and that the probability of an individual being immune equals α independently of the immunity of others. In this case, in families of size j we have

$$\xi_0 = 1, \quad P_j(\xi_1 = k) = \binom{j-1}{k} \alpha^k (1-\alpha)^{j-1-k}$$

$$\text{for } k = 0, 1, \ldots, j-1 \text{ and } \xi_2 = \xi_3 = \ldots = 0.$$

In the second model we could make α dependent on some parameter θ^*. In these both models we can easily write down explicit expressions for functions and quantities relevant for the application of theorems of sections 3 and 4. Thus, in the first case we have $Q_j(s) = s^j$, hence

$$Q(s) = \sum_j \pi_j s^j,$$

which makes it possible to apply theorems 3.1 and 3.2. Next, we have

$$p_{j,0}(1) = 1, \quad p_{j,1}(1, j-1) = 1, \quad p_{j,n}(1, j-1, r_2, \ldots, r_n) = 0$$

unless $r_2 = \ldots = r_n = 0$. Thus,

$$K_0(s_0) = s_0, \qquad K_1(s_0, s_1) = \sum_j \pi_j s_0 s_1^{j-1}, \qquad K_n(s_0, s_1, \ldots, s_n) = K_1(s_0, s_1)$$

for all $n > 1$,

which enables us to apply theorems 4.1 and 4.3. Finally, we have $h_0 = 1$, $h_1 = D-1$, $h_2 = h_3 = \ldots = 0$, where D is the expected size of the family; the last formulas enable us to apply theorems 4.2 and 4.4.

In case of the possibility of immunity, the above-mentioned expressions take the form

$$Q(s) = \sum_j \pi_j s(1-a+as)^{j-1}, \qquad K_n(s_0, s_1, \ldots, s_n) = K_1(s_0, s_1)$$

for all $n > 1$, $h_0 = 1$, $h_1 = 1 + a(D-1)$ and $h_2 = h_3 = \ldots = 0$.

Next, we may design still more realistic model if we assume that, given the family size to be j, given that $\xi_n = r > 0$, and $j - \xi_0 - \xi_1 - \ldots - \xi_n = k > 0$, the random variable ξ_{n+1} has the binomial distribution $b\big(k;\ p(r)\big)$; if $\xi_n = 0$ or $\xi_0 + \ldots + \xi_n = j$, then $\xi_{n+1} = 0$. The plausible assumptions about $p(r)$ are $p(r) = p = \text{const}$ for all $r > 0$, and $p(r) = 1-(1-p)^r$ for some $0 < p < 1$. The first condition corresponds to the assumption that whatever the number of infectives in the family, provided only that this number is positive, all non-infected members of the family have the same probability p of getting infected, and the infections are independent for different susceptibles.

The second assumption, presumably most realistic among the cases considered above, is that if there are r infectives and k susceptibles, each of the latter makes, independently of others, one "contact" with each of the infectives, the results of contacts being independent and leading to an infection with probability p. The assumptions of the last model give $p(r) \approx rp$ for small p and r, hence the infection rate is roughly proportional to both the number of susceptibles and the number of infectives, a standard assumption in the theory of epidemics.

In the last two models, known in the literature as *chain binomial models* (see, for instance, [1]), the formulas for joint probability distributions of epidemics within families are somewhat involved, but one can easily obtain a system of recursive equations for calculating the expectation $h = \mathrm{E}(\xi_0 + \xi_1 + \ldots)$.

Indeed, let $\mathrm{E}(r, k)$ be the expectation of $\xi_{n+1} + \xi_{n+2} + \ldots$ given that $\xi_n = r$ and $j - \xi_0 - \xi_1 - \ldots - \xi_n = k$, where j is the size of the family. We have then

$$\mathrm{E}(r, k) = \sum_i \binom{k}{i} p^i(r)\big(1-p(r)\big)^{k-i}[i + \mathrm{E}(i, k-i)],$$

with $E(0, k) = E(r, 0) = 0$. We have then

$$h = \sum_j \pi_j E(1, j-1).$$

This enables us to apply theorem 3.1. Finally, initial values of h_k can be computed directly, and one can apply theorems 4.2 and 4.4 for predicting the behaviour of few generations of the epidemic. The value of p in either of the last two models (possibly depending on some parameter θ^* characterizing social or sanitary conditions) is estimable from the data on histories of epidemics within families.

References

[1] N. T. J. Bailey, *Elements of stochastic processes with applications to the natural sciences*, Wiley, New York 1963.
[2] R. Bartoszyński, *Branching processes and models of epidemics*, Dissertationes Mathematicae 61 (1969).
[3] T. Harris, *Theory of branching processes*, Springer, Berlin 1963.
[4] M. Loève, *Probability theory*, Van Nostrand, New York 1963.

MATHEMATICAL INSTITUTE
POLISH ACADEMY OF SCIENCES
WARSAW

Received on 5. 2. 1970

R. BARTOSZYŃSKI (Warszawa)

O PEWNYM MODELU EPIDEMII

STRESZCZENIE

W pracy rozpatrywany jest następujący model epidemii: rozważa się populację złożoną z nieskończonej liczby rodzin. Jeżeli w danej rodzinie w momencie t nastąpi zarażenie, zapoczątkowuje to epidemię wewnątrz tej rodziny, interpretowaną jako pewien proces stochastyczny $\xi_0, \xi_1, \ldots$, gdzie ξ_n jest liczbą osób tej rodziny zakaźnych w chwili $t+n+1$. W każdym momencie czasu w populacji znajduje się wobec tego pewna liczba osób zakaźnych. Każda z nich zaraża pewną liczbę osób z nowych rodzin. Te ostatnie zarażenia są niezależne i liczby osób zarażonych mają rozkład Poissona.

W pracy badane są warunki dla wygaśnięcia z prawdopodobieństwem jeden tak określonej epidemii oraz analizowane są dwa procesy stochastyczne – opisujące liczby rodzin zarażonych oraz łączne liczby osób zakaźnych w kolejnych momentach czasu.