The contagious Poisson process is a Markov process which has been used to represent the reinforcement of events. Standard methods for estimating and testing this model are shown to be generally inefficient and usually inconsistent. New methods making use of events counts from two or more adjacent intervals of time are proposed and illustrated. The model is also generalized to allow for initial heterogeneity.

Estimation and Testing for a Markov Model of Reinforcement

PAUL D. ALLISON
Cornell University

A variety of social and psychological theories embody the idea that the occurrence of an event increases the probability that the event will occur again. Learning theory, for example, tells us that the emission of a behavior under a schedule of positive reinforcement increases the rate at which the behavior is emitted (Kimble, 1968). Labeling theory suggests that acts defined as deviant may be more likely to occur again as part of "secondary deviance" (Gibbs and Erickson, 1975). The theory of cumulative advantage asserts that individuals who are socially "successful" at some point in time are even more likely to be successful in the future (Cole and Cole, 1973).

For convenience, I will refer to all such theories as *reinforcement* theories. Although the term is borrowed from the theory of operant conditioning, it does not refer here to any specific psychological or sociological mechanism. In fact, this article will not deal with underlying mechanisms at all, but only with the general postulate that an event reinforces the tendency for the same event to recur. My aim is to examine the properties of a mathematical model of reinforcement which could be used to represent any of the theories just mentioned.

AUTHOR'S NOTE: *I am indebted to Barbara Reskin for use of the data reported in Table 1. Neil Henry, Tad Krauze, and Scott Long made many helpful suggestions.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 8 No. 4, May 1980 434-453

THE CONTAGIOUS POISSON PROCESS

The model was introduced to sociologists by Coleman (1964: 299), who referred to it as the contagious Poisson process. In the literature of stochastic processes it is more commonly known as "linear birth with immigration" (Boswell and Patil, 1970), but it has also been called a generalized Polya model (Bates and Neyman, 1952). It is a special case of the general continuous-time Markov process discussed by Hoel et al. (1972).

Although the contagious Poisson process is usually specified by a system of differential equations (e.g., Coleman, 1964: 299), a simpler approach is to specify the joint distribution of a sequence of random waiting times between events. Consider an individual (either a person or a collectivity) who is observed between time 0 and time T. Suppose that events occur to the individual at times $t_1, t_2 \ldots t_l$, where the t_i's are random variables distributed over the interval (0, T). The waiting times between events are defined by

$$\tau_1 = t_1,$$
$$\tau_i = t_i - t_{i-1} \qquad i = 2, 3. \ldots$$

In a *simple* Poisson process (Hoel et al., 1972: 94), these waiting times are independent, and each has an exponential distribution with the parameter λ. Thus, the probability distribution of τ_i is given by

$$\Pr(\tau_i < y) = 1 - e^{-\lambda y} \qquad i = 1, 2. \ldots$$

In the *contagious* Poisson process, the waiting times are still independent and exponentially distributed, but the parameter of the distribution increases by a constant β for each successive waiting time. Thus,

$$\Pr(\tau_i < y) = 1 - e^{-\lambda_i y} \qquad i = 1, 2 \ldots$$

where

$$\lambda_i = \alpha + (i - 1)\beta.$$

Since the expected value of τ_i is equal to $1/\lambda_i$, larger values of λ_i imply smaller expected waiting times and, hence, higher probabilities that an event will recur. The parameter α can be thought of as the initial propensity for the occurrence of events. Each time an event occurs, the propensity is incremented by an amount β which can be interpreted as the rate of reinforcement. Both α and β are scaled as the number of events per unit time.

The contagious Poisson process has been considered as a model for industrial accidents (Arbous and Kerrich, 1951), episodes of racial violence (Spilerman, 1970), the hospitalization of mental patients (Eaton, 1974), and the productivity of scientists (Allison and Krauze, 1977). If such a model is to be usefully applied to empirical data, it is essential that one be able to estimate the parameters (in this case α and β) and test the model against alternative hypotheses. For the most part, researchers have relied on the estimation and testing procedures given by Arbous and Kerrich (1951) and Coleman (1964). Unfortunately, these procedures have two severe limitations. First, since the parameter estimators are inefficient (Sichel, 1951), chi-square goodness-of-fit tests which depend on those estimators are not asymptotically valid. Hence, there has been a tendency to reject the model in cases where it might have been appropriate. Second, I will show that these methods rest on an assumption of initial homogeneity which is quite unlikely to be satisfied for most social science data. As a consequence, estimates of the reinforcement parameter β are usually severely biased upward.

In this article I will introduce new methods for estimation and testing which avoid both these problems. Throughout, I will assume that the data consist of a sample of N independent individuals, each emitting events according to a contagious Poisson process. Thus, the "contagion" is within individuals, not between them. In the next two sections, I will also assume that all individuals in the sample have the same parameters for the contagious Poisson process. In the final section, I will relax this assumption and allow for a certain degree of heterogeneity.

THE NUMBER OF EVENTS IN A SINGLE INTERVAL

Although waiting times provide a convenient way to specify the model, available data more typically consist of the number of events in some interval of time for each individual in the sample. Coleman (1964: 302), for example, considers the number of phonograph records bought by each of 891 high school girls in a one-month interval. He suggests that these data might be generated by a contagious Poisson process, since "a girl has only a small likelihood of buying a record, but having bought one she is more likely to buy a second . . . and so on." The problem is to estimate from such data the initial propensity α and the rate of reinforcement β.

Let X(t) be the number of events that happen to an individual between time 0 and time t. As Coleman and many others have shown, X(t) has a negative binomial distribution.[1] That is,

$$\text{pr}[X(t) = x] = \binom{k + x - 1}{x} \left(\frac{k}{\mu + k}\right)^k \left(\frac{\mu}{\mu + k}\right)^x, \qquad [1]$$

where μ is a scale parameter and k is a shape parameter. These parameters are determined, in turn, by the parameters of the contagious Poisson process:

$$k = \alpha/\beta, \qquad [2]$$

$$\mu = \frac{\alpha}{\beta}(e^{\beta t} - 1). \qquad [3]$$

Solving for α and β gives

$$\beta = \log[(\mu/k) + 1]/t, \qquad [4]$$

$$\alpha = \beta k, \qquad [5]$$

where log denotes the natural logarithm. Thus, if one can estimate μ and k, these estimates can be substituted into equations 4 and 5 to estimate α and β. The parameter μ is the expected value of X(t), and it is efficiently estimated by the sample mean of X. A variety

of methods have been proposed for estimating k (Johnson and Kotz, 1969: 131). The simplest is the method of moments, which in this case is

$$\tilde{k} = \frac{\bar{X}^2}{S^2 - \bar{X}}, \qquad [6]$$

where S^2 is the sample variance. (A tilde over the parameter will be used to denote a method-of-moments estimator. The carat will be reserved for maximum-likelihood estimators.) Substituting $\tilde{k}$ and mean X for k and μ in equations 4 and 5 yields method-of-moments estimators for α and β.

Although the method of moments is the most widely used procedure, it has been known for some time that it is not fully efficient (Sichel, 1951). In fact, for some fairly typical values of μ and k, the large-sample efficiency of $\tilde{k}$ falls below 50%. This loss of efficiency is especially serious when one is using parameter estimates to generate a goodness-of-fit test, as will be seen shortly.

Asymptotically efficient estimators of α and β can be obtained by the method of maximum likelihood. Again, the procedure is first to get maximum-likelihood estimates (MLEs) of μ and k, and then to substitute these estimates into equations 4 and 5.[2] The MLE of μ is just the sample mean of X (Bishop et al., 1975: 453). The MLE of k is a solution to the equation

$$\sum_{x=0}^{\max x} \left(\frac{A_x}{k + x}\right) - N \log \left(1 + \frac{\bar{X}}{k}\right) = 0, \qquad [7]$$

where A_x is the number of observations with more than x events, max x is the highest number of events observed in the sample, and N is the sample size (Fisher, 1953). Since there is no explicit solution to this equation, an iterative method is necessary. One such method is the Newton-Raphson algorithm (Grove, 1966). Denote the righthand side of equation 7 by U(k) and let $I(k) = \partial U(k)/\partial k$. It can be shown that

$$I(k) = \frac{N\bar{X}}{k^2 + k\bar{X}} - \sum_{x=0}^{\max x} \frac{A_x}{(k + x)^2}. \qquad [8]$$

The Newton-Raphson algorithm is then

$$k_{j+1} = k_j - U(k_j)/I(k_j). \quad [9]$$

If the starting value k_1 is a consistent estimator (e.g., the moment estimator $\tilde{k}$), the iteration will usually converge rapidly. The estimated asymptotic standard error of the final value $\hat{k}$ is given by $[-I(\hat{k})]^{-1/2}$ (Cox and Hinkley, 1974: 302). A FORTRAN program for this algorithm is available on request.

As an example, the method is applied to the data in Table 1. The events consisted of articles published by 237 chemists during the first six years after receipt of the doctorate. Since it has often been hypothesized that the publication of an article increases the probability that a scientist will publish again (Cole and Cole, 1973; Allison and Stewart, 1974; Price, 1976; Gaston, 1978), the contagious Poisson process would appear to be a reasonable model for the generation of article counts. Under this hypothesis, the parameters α and β were estimated both by the method of moments and by maximum likelihood.

Substituting the sample mean of 3.4599 and the sample variance of 14.3596 into equation 6, we get $\tilde{k} = 1.0983$. Then, substituting these estimates into equations 4 and 5, we get moment estimates of .2372 for β and .2605 for α. This means that at the beginning of the six-year period, each chemist had an estimated propensity to publish .2605 articles per year. Each time a chemist published an article, his or her propensity to publish went up by an estimated .2372 of an article per year. To get the MLE of k, the Newton-Raphson algorithm was used with the moment estimate of 1.0983 as the starting value. The results from each iteration are shown at the bottom of Table 1. Convergence occurred in the fourth pass through the algorithm, yielding $\hat{k} = 1.3051$. Substituting this value into equations 4 and 5 with $\hat{\mu}$ = mean X = 3.4599 gives $\hat{\beta} = .2158$ and $\hat{\alpha} = .2816$. While these estimates are quite close to the moment estimates, this will not generally be the case.[3]

After estimating the parameters, it is usually desirable to test whether the data really came from a negative binomial distribution. This can be done by the usual goodness-of-fit chi-square test (Hays, 1963: 580). The expected frequencies are easily obtained

by the following recurrence formulas (Bliss, 1953). Let E_x be the expected number in cases with x events. Then

$$\hat{E}_0 = N\left(1 + \frac{\bar{X}}{\hat{k}}\right)^{-\hat{k}}$$

$$\hat{E}_x = \left(\frac{\hat{k} + x - 1}{x}\right)\left(\frac{\bar{X}}{\hat{k} + \bar{X}}\right)\hat{E}_{x-1} \qquad x = 1, 2. \ldots \qquad [10]$$

These formulas give MLEs for the expected frequencies. If the moment estimator $\tilde{k}$ is used instead of $\hat{k}$, the resulting chi-square statistic will not have an asymptotic chi-square distribution. As Sichel (1951) has shown, this common error frequently results in the rejection of the negative binomial hypothesis when the correct test would have indicated an acceptable fit. In calculating the degrees of freedom for the chi-square statistic, two degrees of freedom must be subtracted for the two parameters estimated from the data.

Applying these methods to the data in Table 1, we find that the negative binomial distribution fits very well (results are shown at the bottom of the last column). If the data had led to rejection of the negative binomial distribution, this would also have implied rejection of the contagious Poisson process as a model for the data. Failure to reject does not add much support to the model, however, since there are many other models which also imply a negative binomial distribution (Boswell and Patil, 1970). One such model will be discussed later.

We now have efficient statistical methods for data consisting of the number of events in a single interval of time. However, as I suggested earlier, there are strong reasons for believing that the sorts of data usually available to social scientists (and others) cannot meet the minimal assumptions required for these methods. Consider Coleman's example of high school girls buying records. Apparently the girls were interviewed at some arbitrary point in time and asked how many records they had purchased in the preceding month. The problem arises from the fact that the estimation procedures just described are based on the assumption that the data consist of N realizations of the random variable X(t), the number of events that occur between time 0 and time t.

TABLE 1
Distribution of Publication Counts for 237 Chemists in the First Six Years After the Doctorate

Number of Publications	Number of Chemists	Expected Frequency	
0	37	43.7	
1	50	41.4	
2	37	34.7	
3	31	27.7	
4	24	21.7	
5	13	16.7	
6	10	12.7	
7	7	9.7	
8	7	7.3	
9	3	5.5	
10	4	4.1	a
11	1	3.1	
12	2	2.3	
13	2	1.7	
14	2	1.3	
15	1	.9	
16	2	.7	a
17	0	.5	
18	2	.4	
19	2	.3	
≥ 20	0	.7	

$\bar{X}$ = 3.4599
S^2 = 14.3596
$\tilde{k}$ = 1.0983

X^2 = 9.57
d.f. = 9
p = .39

Interation Results:

k_1 = 1.0983
k_2 = 1.2569
k_3 = 1.3022
k_4 = 1.3050
k_5 = 1.3051

$\hat{k}$ = 1.3051
S.E.($\hat{k}$) = .1725

a. Expected and observed frequencies for these cells were combined for the goodness-of-fit test.

However, time 0 is a very special time in the contagious Poisson process. At that time, the model presumes that every individual in the sample is identical with respect to the occurrence of events; i.e., at time 0 each individual has the same propensity α for events to occur. But as soon as events begin to occur, the sample begins to diverge. Individuals who experience many events then have greater propensities for future events, while individuals who have experienced no events still retain their initial propensity α. In the case of the high school girls, even if we assume that there was

some point in the past at which they were equally likely to buy records, there is no reason to believe that they were equally likely to do so at the beginning of the month about which they were questioned. The existence of a reinforcement process would insure that the formerly homogeneous group would now be heterogeneous. This problem occurs whenever the counts of events come from some arbitrary interval during a contagious Poisson process.

To examine the implications of the problem, let us define a random variable $Y_s(t)$, which is the number of events in some arbitrary interval of length s which begins at time t. Thus, $Y_s(t) = X(t + s) - X(t)$. The immediate objective is to find the distribution of $Y_s(t)$. The joint distribution of $Y_s(t)$ and $X(t)$ is given by Arbous and Kerrich (1951: 421):

$$\Pr[Y_s(t) = y,\ X(t) = x] \qquad [11]$$

$$= \frac{\Gamma(k + x + y)}{\Gamma(k)\, x!\, y!} \left(\frac{k}{\mu_x + \mu_y + k}\right)^k \left(\frac{\mu_y}{\mu_x + \mu_y + k}\right)^y \left(\frac{\mu_x}{\mu_x + \mu_y + k}\right)^x,$$

where

$$k = \alpha/\beta$$
$$\mu_x = \alpha(e^{\beta t} - 1) / \beta$$
$$\mu_y = \alpha(e^{\beta s} - 1)e^{\beta t} / \beta.$$

Equation 11 is known as a bivariate negative binomial distribution. Since this is a special case of the negative multinomial distribution, it immediately follows from results given by Bishop et al. (1975: 455) that the marginal distribution of $Y_s(t)$ is a negative binomial:

$$\Pr[Y_s(t) = y] = \binom{k + y - 1}{y} \left(\frac{k}{k + \mu_y}\right)^k \left(\frac{\mu_y}{k + \mu_y}\right)^y. \qquad [13]$$

Thus, the goodness-of-fit test described earlier is still appropriate for data drawn from an arbitrary interval of time. A comparison of equations 13 and 1 also shows that the distributions of $Y_s(t)$ and $X(t)$ are governed by the same shape parameter k. And since $k = \alpha/\beta$, the ratio of the parameters can be estimated by the methods already presented, either the method of moments or maximum likelihood. The problem comes in trying to separate the two parameters. Previously we estimated μ_x and k and substituted into equation 4, which was

$$\beta = \log[(u_x/k) + 1]/t.$$

Let us define an analog of the righthand term, for data drawn from an arbitrary interval of length s:

$$\beta^* = \log[(\mu_y/k) + 1]/s. \qquad [14]$$

Substituting from 12 gives

$$\beta^* = \log[(e^{\beta s} - 1)e^{\beta t} + 1]/s. \qquad [15]$$

It is not difficult to show that $\beta^* \geqslant \beta$, with equality if and only if $t = 0$. In other words, substituting mean Y (the sample mean of $Y_s[t]$) and k into equation 14 gives a consistent estimator of β if and only if events are counted starting at time 0. Otherwise, the estimate will tend to be too large. Since $\alpha = \beta k$, the estimate for α will also tend to be too large.

Is there any way of getting consistent estimators of α and β from counts of events over an arbitrary interval of time? Unfortunately, the parameters are, in most instances, fundamentally underidentified. The distribution of $Y_s(t)$ has only two parameters, μ_y and k, which by 12 depend on four quantities, α, β, t, and s. The length of the observed interval, s, is assumed to be known. But t, the length of time since time 0, will not usually be known. There are thus two equations and three unknowns, and no solution for α and β is possible.

The situation is not hopeless, however. In the next section I show that more detailed kinds of data enable one to estimate not only α and β, but t, the length of time since the sample was homogeneous. These sorts of data also provide for a more demanding test of the contagious Poisson hypothesis. In developing these methods, I will assume that all individuals in the sample are characterized by the same contagious Poisson process, although their observed scores will be different realizations of that random process. This implies that there was some unknown time at which all individuals were homogeneous with respect to the occurrence of events. For most social science data, this is probably not a realistic assumption either. Although individuals might be subject to the same reinforcement regime, it will usually be the case that they enter that regime with differing event propensities. In the last section, I develop a generalization of the model which allows for such initial heterogeneity, and I show that estimation methods given in the next section also apply to the generalized case.

COUNTS FROM TWO ADJACENT INTERVALS

When one has counts of events from some arbitrary interval of time, the simplest approach to estimating α and β is to divide the interval into two subintervals, each of length s. In the notation of the previous section, such counts are realizations of the random variables $Y_s(t)$ and $Y_s(t + s)$, which for simplicity will be denoted by Y_1 and Y_2, respectively. Again, t is the length of the interval between time 0 and the start of the first observed interval, and X(t) is the number of events in the interval (0, t). Both t and X(t) are assumed to be unobserved. It will also be convenient to use $Y = Y_1 + Y_2$; i.e., Y is the total number of events happening to a random individual over the interval (t, t + 2s). Drawing from Bates and Neyman (1952), it can be shown that the joint distribution of Y_1 and Y_2 is a bivariate negative binomial:

$$\Pr(Y_1 = y_1, Y_2 = y_2) = \frac{\Gamma(k + y_1 + y_2)}{\Gamma(k)\, y_1!\, y_2!} \left(\frac{k}{r}\right)^k \left(\frac{\mu_1}{r}\right)^{y_1} \left(\frac{\mu_2}{r}\right)^{y_2} \quad [16]$$

where $r = k + \mu_1 + \mu_2$. Equation 16 contains three parameters, which can be expressed as functions of the parameters of the contagious Poisson process. In particular,

$$\begin{aligned} k &= \alpha/\beta \\ \mu_1 &= \alpha(e^{\beta s} - 1)e^{\beta t} / \beta \\ \mu_2 &= \alpha(e^{\beta s} - 1)e^{\beta(t+s)} / \beta. \end{aligned} \quad [17]$$

Since s is assumed to be known, we have three equations and three unknowns, α, β, and t. The solutions to these equations are

$$\beta = \log(\mu_2/\mu_1)/s, \quad [18]$$

$$\alpha = \beta k, \quad [19]$$

$$t = \frac{1}{\beta} \log \left(\frac{\mu_1^2}{k(\mu_2 - \mu_1)} \right). \quad [20]$$

One can, therefore, obtain MLEs of α, β, and t by first getting MLEs of k, μ_1, and μ_2 and then substituting the estimates for the population parameters in equations 18-20 The MLEs of μ_1 and μ_2 are just the sample means of Y_1 and Y_2. The MLE of k is obtained by aggregating the data in the two intervals and using the iterative method presented in the previous section (Johnson and Kotz, 1969). That is, the MLE of k is the solution to equation 7, with mean Y taking the place of mean X and y taking the place of x.

As an illustration, let us return to the data of Table 1. To apply the method, I disaggregated the six-year article counts into two three-year counts and computed the mean number of articles in each three-year interval. This gave mean $Y_1 = 1.6540$ and mean $Y_2 = 1.8059$. The MLE of k is the same as that computed earlier, 1.3051. Substituting these figures into equations 18-20 yields $\hat{\beta} = .0293$, $\hat{\alpha} = .0382$, and $\hat{t} = 95.576$. The estimates of α and β are both very much smaller than those obtained from the single-interval method, suggesting a much lower "reinforcement" value of each publication. The $\hat{t}$ estimate says that chemists were homogeneous with respect to publication propensity approximately 96 years

prior to receiving the doctorate. This impossible conclusion suggests that either there was substantial heterogeneity not accounted for by the contagious Poisson process, or the model is defective in some other way. Thus, $\hat{t}$ provides some indication of the plausibility of the model.

It is also quite simple to estimate the standard error of β. The variance of the sample means (see Bishop et al., 1975: 254) is given by

$$\text{var } \bar{Y}_i = \frac{\mu_i}{N}\left(1 + \frac{\mu_i}{k}\right) \qquad i = 1, 2, \qquad [21]$$

and their covariance (see Johnson and Kotz, 1969) is

$$\text{cov}(\bar{Y}_1, \bar{Y}_2) = \frac{\mu_1 \mu_2}{Nk}. \qquad [22]$$

Applying the delta method (Bishop et al., 1975: 487) to equations 18, 21, and 22 yields the asymptotic variance of $\hat{\beta}$:

$$\text{var } \hat{\beta} = \frac{\mu_1 + \mu_2}{s^2 N \mu_1 \mu_2} \qquad [23]$$

(Note that s^2 is the squared interval length, not the sample variance.) Replacing population means by sample means gives an estimated standard error of $[\bar{Y}_1 + \bar{Y}_2)/(N\bar{Y}_1\bar{Y}_2)]^{1/2}/s$. For the two-interval publication data, the estimated standard error of β is .0233. The standard errors of $\hat{t}$ and $\hat{\alpha}$ are much more complicated and will not be considered here. With data from two adjacent intervals, it is also possible to test the contagious Poisson process against other possible models. A model which has been frequently discussed as an alternative to the contagious Poisson process is the *compound* Poisson process (Spilerman, 1970). This model assumes that each individual experiences events according to a *simple* Poisson process with a parameter λ that is constant over time for each individual but varies across individuals. If the variability of λ is described by the gamma density.

$$f(\lambda) = \frac{\gamma^{\theta}\lambda^{\theta-1}e^{-\gamma\lambda}}{\Gamma(\theta)} \qquad \theta, \gamma > 0, \qquad [24]$$

then the number of events in any interval of the process has a negative binomial distribution. Moreover, the joint distribution of the number of events in two adjacent intervals is a bivariate negative binomial, the same as for the contagious Poisson process. Thus, quite different models give rise to very similar observations.

In the two-interval case, however, the bivariate negative binomial distributions are not identical for the two models. While the joint distribution for the contagious Poisson process is unrestricted, the compound Poisson process implies that $\mu_1 = \mu_2 = \mu$ in equation 16. Thus, testing the hypothesis that $\mu_1 = \mu_2$ is equivalent to testing the compound Poisson model against the contagious Poisson model. By equations 21 and 22, the variance of mean Y_2 - mean Y_1 is

$$\text{var}(\bar{Y}_2 - \bar{Y}_1) = \frac{\mu_2}{N}\left(1 + \frac{\mu_2}{k}\right) + \frac{\mu_1}{N}\left(1 + \frac{\mu_1}{k}\right) - \frac{2\mu_1\mu_2}{kN}. \qquad [25]$$

But under the null hypothesis that $\mu_1 = \mu_2 = \mu$, equation 25 reduces to $2\mu/N$. Since the MLE of μ is $(\bar{Y}_1 + \bar{Y}_2)/2$, the statistic

$$(\bar{Y}_2 - \bar{Y}_1)\sqrt{\frac{N}{\bar{Y}_1 + \bar{Y}_2}} \qquad [26]$$

has a standard normal distribution in large samples under the null hypothesis. The value of [26] for the data on chemists' publications is 1.26, which gives little justification for choosing the contagious Poisson process over the compound Poisson process.

EXTENSIONS

We now have a fairly comprehensive set of methods for estimating and testing the contagious Poisson process when the

data consist of event counts from two adjacent intervals of equal length. Sometimes, however, available data will be in a somewhat different or more detailed form. Consider, first, the case in which the two intervals over which events are counted are of *unequal* length. Suppose the first interval is of length s and the second is of length u. The joint probability distribution for the counts in the two intervals is still given by equation 16, but 17 must be modified to read

$$\mu_1 = \alpha(e^{\beta s} - 1)e^{\beta t} / \beta$$
$$\mu_2 = \alpha(e^{\beta u} - 1)e^{\beta(t+s)} / \beta. \qquad [27]$$

If s = u, these equations can be explicitly solved for β, but when $s \neq u$ there can be no explicit solution. As usual, however, a numerical solution is possible by iterative methods. A very simple but serviceable method is based on the fact that 27 implies

$$\beta = \log \left[\frac{\mu_2(e^{\beta s} - 1)}{\mu_1(e^{\beta u} - 1)} \right] / s. \qquad [28]$$

Substituting sample means for population means gives the following iterative algorithm which can be easily implemented with a programmable calculator:

$$\beta_{j+1} = \log \left[\frac{\bar{Y}_2(e^{s\beta_j} - 1)}{\bar{Y}_1(e^{u\beta_j} - 1)} \right] / s. \qquad [29]$$

Experience suggests that this algorithm converges with reasonable rapidity from virtually any starting value. The MLE of k is again found by aggregating counts from the two intervals and applying the iterative method for a single interval. Once $\hat{\beta}$ and $\hat{k}$ are obtained, estimates of α and t are found from

$$\hat{\alpha} = \hat{\beta}\hat{k}, \qquad [30]$$

$$\hat{t} = \log \left[\frac{\bar{Y}_1}{\hat{k}(e^{\hat{\beta}s} - 1)} \right] / \hat{\beta}. \qquad [31]$$

Sometimes event counts are available for multiple adjacent intervals of equal length. If there are an even number of intervals, one can, of course, aggregate the counts into two equal intervals and apply the methods described earlier; or if there are an odd number of intervals, one can aggregate into two unequal intervals and apply the methods just discussed. In either case, some information is lost in the aggregation, and more precise estimates could be gotten from the disaggregated data. Methods for making full use of such data are described in an unpublished version of this article (Allison, 1978). Even more precise parameter estimates can be obtained from data on the exact times at which events occurred for each individual in the sample. Estimation and testing procedures for such data can also be found in Allison (1978).

It is also possible to generalize the contagious Poisson process in order to make it somewhat more realistic. To this point, it has been assumed that there is some time 0 at which all individuals are homogeneous with respect to the occurrence of events. At that time, each individual has a propensity α for the occurrence of events and is subject to a reinforcement process which increases that propensity whenever an event occurs. The stochastic character of the process implies that, if observation begins at time $t > 0$, there is a good chance that individuals will differ in their propensities for events to occur. The larger t is, the greater is the expected heterogeneity.

In many applications of the contagious Poisson process, it will not be plausible to assume that there was ever a time in the past when individuals were identical with respect to the occurrence of events. Even when it is reasonable to assume that individuals are subject to exactly the same reinforcement regime, there will usually be an insufficient basis for assuming that they were homogeneous when they entered that regime.

To allow for such initial heterogeneity, let us suppose that at the beginning of the observation period each individual starts experiencing events according to a contagious Poisson process with the parameters α and β, both constant over time; but now we let α be a random variable across individuals instead of being identical for all individuals. As before, β is the same for all individuals. Let us consider the case in which the data consist of the number of events in two adjacent intervals of equal length. Let Y_1 be the number of events in the first interval (0, s) and let Y_2 be the number of events in the second interval (s, 2s). Conditional on α, the distribution of the joint variable (Y_1, Y_2) is the bivariate negative binomial distribution in equations 16 and 17 with $t = 0$. But the unconditional distribution of (Y_1, Y_2) depends on the unspecified distribution of α and will not, in general, be a bivariate negative binomial. Hence, it is not possible to obtain maximum-likelihood estimators in this situation. Moment estimators will be obtained instead.

Consider, first, the conditional means of Y_1 and Y_2 given α. From 17 we have

$$E(Y_1 \mid \alpha) = \frac{\alpha}{\beta}(e^{\beta s} - 1)$$

$$E(Y_2 \mid \alpha) = \frac{\alpha}{\beta}(e^{\beta s} - 1)e^{\beta s}. \qquad [32]$$

The unconditional means are then just

$$\mu_1 = E(\alpha)(e^{\beta s} - 1)/\beta$$

$$\mu_2 = E(a)(e^{\beta s} - 1)e^{\beta s}/\beta. \qquad [33]$$

Hence,

$$\beta = \frac{1}{s}\log\left(\frac{\mu_2}{\mu_1}\right). \qquad [34]$$

Replacing population means by sample means gives a method-of-moments estimator which is identical to the MLE under the more restrictive assumption that α is constant over individuals. Methods for getting moment estimates of the mean and variance of α are described in Allison (1978).

DISCUSSION

I have presented new methods for estimating the parameters of the contagious Poisson process and for testing the model against the more restricted compound Poisson process. These methods are more statistically efficient than those previously employed, and they make less stringent assumptions about the data. It must be emphasized that the hypothesis test given earlier does not, in any general sense, enable one to determine whether reinforcement does or does not occur. In fact, when one only observes the occurrence or nonoccurrence of such events, no such test is possible (Taibleson, 1974). The test proposed here compares the contagious Poisson process with the compound Poisson process, a model of heterogeneity and no reinforcement. But the compound Poisson is a very restrictive hypothesis, and the comparison of the two only amounts to testing whether the rate of event occurrence does or does not increase with time. It is not difficult to construct a model without reinforcement which still allows the rate of occurrence to increase with time.

This brings us to the question of whether the contagious Poisson process is a reasonable model for social reinforcement processes. Of course, this question can only be satisfactorily answered by separately considering each substantive application. Yet, there are some features of the model that are likely to be troublesome in many applications. One is the assumption that the reinforcing effect is constant over time, leading to an exponential increase in the rate of occurrence. In reality, however, there is usually some vague upper bound on the rate of occurrence, a fact that is manifest in learning curves that are approximately logistic

or sigmoid (Pitcher et al., 1978). As a consequence, the Poisson process may be appropriate only for relatively short intervals or when observation begins when individuals are well below the upper bound.

On balance, however, it seems likely that the contagious Pisson process may be a good first approximation of reinforcement, even when many of its assumptions appear somewhat unrealistic. At least, it serves as a useful baseline for determining whether and what elaborations are necessary.

NOTES

1. A brief introduction to the negative binomial distribution can be found in Bishop et al. (1975: 452). For a more extensive discussion, see Johnson and Kotz (1969), who use a slightly different parameterization.

2. Here and throughout the article I make use of the invariance property of maximum-likelihood estimates (Cox and Hinkley, 1974: 287). If A and B are two parameters (or two vectors of parameters) and $A = f(B)$, the invariance property states that the MLE of A is given by $\hat{A} = f(\hat{B})$ where $\hat{B}$ is the MLE of B.

3. For an example in which the moment estimate of k deviates markedly from the MLE of k, see Allison (forthcoming). In the same article, I argue that the reciprocal of k is a useful measure of inequality.

REFERENCES

ALLISON, P. D. (forthcoming) "Inequality and scientific productivity." Social Studies of Sci.

——— (1978) "Estimation and testing for a model of reinforcement." (unpublished)

——— and T. K. KRAUZE (1977) "The effect of cumulative advantage on inequality in science." Presented at the annual meeting of the American Sociological Association.

ALLISON, P. D. and J. A. STEWART (1974) "Productivity differences among scientists: evidence for accumulative advantage." Amer. Soc. Rev. 39 (August): 596-606.

ARBOUS, A. G. and J. E. KERRICH (1951) "Accident statistics and the concept of accident proneness," Biometrics 7 (December): 340-432.

BATES, G. E. and J. NEYMAN (1952) "Contributions to the theory of accident proneness." University of California Publications in Statistics 1: 215-275.

BISHOP, Y.M.M., S. E. FIENBERG, and P. W. HOLLAND (1975) Discrete Multivariate Analysis. Cambridge, MA: MIT Press.

BLISS, C. I. (1953) "Fitting the negative binomial distribution to biological data." Biometrics 9 (June): 176-196.

BOSWELL, M. T. and G. P. PATIL (1970) "Chance mechanisms generating the negative binomial distribution," pp. 3-22 in G. P. Patil (ed.) Random Counts in Biomedical and Social Sciences. University Park, PA: Pennsylvania State Univ. Press.

COLE, J. R. and S. COLE (1973) Social Stratification in Science. Chicago: Univ. of Chicago Press.

COLEMAN, J. S. (1964) Introduction to Mathematical Sociology. New York: Free Press.

COX, D. R. and D. V. HINKLEY (1974) Theoretical Statistics. London: Chapman & Hall.

EATON, W. W., Jr. (1974) "Mental hospitalization as a reinforcement process." Amer. Soc. Rev. 39 (April): 252-260.

FISHER, R. A. (1953) "Note on the efficient fitting of the negative binomial." Biometrics 9 (June): 197-200.

GASTON, J. (1978) The Reward System in British and American Science. New York: John Wiley.

GIBBS, J. P. and M. L. ERICKSON (1975) "Major developments in the sociological study of deviance," pp. 21-42 in A. Inkeles et al. (eds.) Annual Review of Sociology, Volume 1. Palo Alto: Annual Reviews.

GROVE, W. E. (1966) Brief Numerical Methods. Englewood Cliffs, NJ: Prentice-Hall.

HAYS, W. L. (1963) Statistics. New York: Holt, Rinehart & Winston.

HOEL, P. G., S. C. PORT, and C. J. STONE (1972) Introduction to Stochastic Processes. Boston: Houghton Mifflin.

JOHNSON, N. L. and S. KOTZ (1969) Discrete Distributions. Boston: Houghton Mifflin.

KIMBLE, G. (1968) "Learning," pp. 113-126 in D. Sills (ed.) International Encyclopedia of the Social Sciences, Volume 9. New York: Macmillan.

PITCHER, B. L., R. L. HAMBLIN and J.L.L. MILLER (1978) "The diffusion of collective violence." Amer. Soc. Rev. 43 (February): 23-35.

PRICE, D.D.S. (1976) "A general theory of bibliometric and other cumulative advantage processes." J. of the Amer. Society for Information Sci. 27: 292-306.

SICHEL, H. S. (1951) "The estimation of the parameters of a negative binomial distribution with special reference to psychological data." Psychometrika 16: 107-151.

SPILERMAN, S. (1970) "The causes of racial disturbances: a comparison of alternative explanations." Amer. Soc. Rev. 35 (August): 627-649.

TAIBLESON, M. H. (1974) "Distinguishing between contagion, heterogeneity, and randomness in stochastic models." Amer. Soc. Rev. 39 (December): 877-880.

Paul D. Allison is Assistant Professor of Sociology at Cornell University. His research interests include stratification and careers in science, categorical data analysis, and methods for the analysis of event histories.