APPLICATIONS OF "BIRTH PROCESSES": EXTENSIONS OF THE PETERSEN MODEL AND THE PITCHER-HAMBLIN-MILLER DIFFUSION MODEL

Andreas Diekmann, University of Hamburg, Institute of Sociology

1. Stochastic "Birth Processes"

Birth processes are a special class of discrete state and continuous time stochastic growth processes which are defined formally (see Bartholomew 1967: 292) by:

$$(1) \quad P \left\{ k \longrightarrow k+1 \quad \text{in } (t, \quad t+dt) \right\} = \lambda_{(k,t)}dt$$

P is the probability of change from state k to state k+1 in the infinitesimally small interval dt, i.e. the probability that a new event (a new border conflict, an act of collective violence) will take place. k is the number of events since the process began at time t=0 and $\lambda_{(k,t)}$ is the <u>transition rate</u> or <u>intensity</u> of the process.

It is characteristic for birth processes that k is allowed to increase but not to decrease. Therefore birth processes are appropriate for describing cumulative growth processes which are governed by probabilistic laws.

Picture 1 illustrates the "one-way street" of a general birth process:

$$0 \xrightarrow{\lambda_{(0,t)}} 1 \xrightarrow{\lambda_{(1,t)}} 2 \xrightarrow{\lambda_{(2,t)}} \dots \xrightarrow{\lambda_{(k-1,t)}} k \xrightarrow{\lambda_{(k,t)}} k+1 \rightarrow \dots$$

<u>Picture 1</u>

The transition rate can be constant, dependent on k, dependent on t, or dependent on both k and t. The four possibilities are shown in the following table:

		t	
		independent	dependent
k	independent	(1) Poisson	(2) time-dependent Poisson
	dependent	(3) epidemic (e.g. Yule process)	(4) time-dependent epidemic (e.g. time-dependent Yule process)

If the transition rate is a function of t the process is called time-dependent, and if the transition rate is dependent on state k the process is an epidemic one. The first process we focus on in this paper is of type 2 and the second process of type 4.

Starting with the assumptions of the process the goal of model building is the derivation of a time-dependent probability density function, i.e. the derivation of the probabilities for the system to be in state k=0, 1, 2, ... for every point in time $t \geqslant 0$.

This can be done by deriving first a recursive system of simultaneous differential equations for all states k=0,1,2 ... The differential equations (2) and (3) express the change of probability as a function of the probability itself:

$$(2) \qquad \frac{dp_{(t)o}}{dt} = - \lambda_{(o,t)} p_{(t)o} \qquad k=0$$

$$(3) \qquad \frac{dp_{(t)k}}{dt} = \lambda_{(k-1,t)} p_{(t)k-1} - \lambda_{(k,t)} p_{(t)k} \qquad k=1,2, \ldots$$

In a more intuitive sense the equations express that the change of probability for state k during the time interval dt is the "inflow" from state k-1 minus the "outflow" of state k (for a derivation of (2) and (3) see Diekmann 1979, chap. III,5, appendix 1). The process can be uniquely specified by the choice of a special transition rate $\lambda_{(k,t)}$. Under the restriction of $p_{oo}=1$, i.e. the process starts at state 0, a solution of (2) and (3) can be found by different methods. The solution is the probability density function we are interested in.

2. The Petersen Model

From a cybernetic perspective Petersen (1978, 1979a, 1979b) suggests the application of the Weibull distribution to time-series data of political events like the Sino-Indian border conflict, international summits, etc. The cumulative Weibull distribution takes the form:

$$(4) \quad p_{(T \leq t)} = 1 - e^{-\lambda t^{\alpha}},$$

where T is the stochastic variable and $p_{(T \leq t)}$ the probability that at least one event will occur in the time span T=t. α and λ are constant parameters that are estimated with empirical data.

In order to interpret the model Petersen (1979a, formula (3)) derives the intensity of the process:

$$(5) \quad \lambda_{(t)} = \lambda \alpha t^{\alpha - 1}$$

In Petersen's model time is the stochastic variable and not the cumulative number of events k. If we look for the distribution of the cumulative number of events, we have to deal with a time-dependent Poisson process with transition rate (5). As a by-product the Weibull distribution (4) can be derived. With the density function of the cumulative events k there is the possibility to estimate α and λ by the principle of maximum likelihood.

For a general time-dependent Poisson process with transition rate $\lambda_{(t)}$ dependent on time but not on k the solution of (2) and (3) is the following Poisson distribution (Chiang 1968: 49)

$$(6) \quad p_{(t)k} = \frac{e^{-\int_0^t \lambda(T)dT}\left[\int_0^t \lambda(T)dT\right]^k}{k\,!}$$

With the transition rate (5) the integral is: $\int_0^t \lambda \alpha T^{\alpha-1} dT = \lambda t^{\alpha}$
Hence we arrive at the distribution:

$$(7) \quad p_{(t)k} = \frac{e^{-\lambda t^{\alpha}} (\lambda t^{\alpha})^k}{k\,!}$$

with expected value:

$$(8) \quad E(k) = \lambda t^{\alpha}$$

(7) is the time-dependent distribution for the cumulative number of events that took place since the process began at t=0.

For k=0 it follows from (7):

$$(9) \quad p_{(t)o} = e^{-\lambda t^{\alpha}}$$

Now, if we change the point of view and regard time as the random variable:

$$(10) \quad p_{(T \leq t)} = 1 - p_{(T > t)} = 1 - p_{(t)o} = 1 - e^{-\lambda t^{\alpha}}$$

The Weibull distribution (10) informs us of the probability that at least one event will happen in a certain time interval t, whereas (7) tells us the probability that 0,1,2,3, ... events will happen up to a point in time $t \geq 0$.

If time-series data of the cumulative number of events are at our disposal and if the time-series data are regarded as realizations of the stochastic process under consideration, the distribution (7) or the expected value curve (8) are useful for estimation of the parameters.

The expected value curve (8) can be fitted to data by the principle of least squares. This is even possible by linear regression if (8) is subject of logarithmic transformation.

However, maximum likelihood estimates (MLE) are more efficient. If we have the set of observations (k_1,t_1); (k_2,t_2); ...; (k_N,t_N) the MLE's are derived as follows:*

$$(11)\quad L[\lambda,\alpha|(k_1,t_1);\ldots;(k_N,t_N)] = \prod_{i=1}^{N} \frac{e^{-\lambda t_i^{\alpha}}(\lambda t_i^{\alpha})^{k_i}}{k_i\;!}$$

(11) is the likelihood function which should be maximized in respect to λ and α. It is more convenient to deal with the log-likelihood function:

$$(12)\quad \ln L_{(\lambda,\alpha)} = -\lambda\sum t_i^{\alpha} + \sum k_i \ln(\lambda t_i^{\alpha}) + \sum \frac{1}{k_i\;!}$$

The partial derivatives with respect to α and λ are:

$$(13)\quad \frac{\delta \ln L(\lambda,\alpha)}{\delta\alpha} = -\lambda\sum t_i^{\alpha}\ln t_i + \sum k_i \ln t_i$$

* Note that the MLE's are derived from the Poisson distribution and not from the Weibull density distribution. The MLE's (15) and (16) are not identical to Petersen's MLE's (formulas (12) and (13) in Petersen 1979a). An iterative solution of (15) and (16) is an alternative estimation technique based on the assumption that the cumulative numbers of events are generated by the same time-dependent Poisson process with parameters α and λ.

$$(14) \quad \frac{\delta \ln L(\lambda, \alpha)}{\delta \lambda} = - \sum t_i^{\alpha} + \frac{1}{\lambda} \sum k_i$$

(14) yields for $\delta \ln L(\lambda, \alpha)/\delta \lambda = 0$ the MLE of λ, $\hat{\lambda}$:

$$(15) \quad \hat{\lambda} = \frac{\sum k_i}{\sum t_i}$$

After setting $\delta \ln L(\lambda, \alpha)/\delta \alpha = 0$ and replacing λ by $\hat{\lambda}$ we obtain from (13) the MLE of α, $\hat{\alpha}$:

$$(16) \quad \sum t_i^{\hat{\alpha}} \sum k_i \ln t_i - \sum k_i \sum t_i^{\hat{\alpha}} \ln t_i = 0$$

The MLE can be computed by numerical techniques and then $\hat{\lambda}$ can be computed with formula (15).

3. A Stochastic Version of the Pitcher-Hamblin-Miller Diffusion Model*

Pitcher, Hamblin, and Miller (1978) presented an interesting deterministic diffusion model of collective violence which is based on Bandura's learning and imitation theory. The model consists of an instigation and inhibition process that is formalized as a system of two differential equations. The two equations lead to the following central equation of the model:

$$(17) \quad \frac{dV}{dt} = ce^{-qt}V \quad ,$$

* We refer in this part to an application of Hamblin's imitation model to collective violence. The derivation of the stochastic model in paragraph 3 and the critical considerations in paragraph 4 are fully transferable to the use diffusion model described in Hamblin, Miller, and Saxton 1979. Equation (17) is identical to equation (9) in Hamblin et al. 1979 if V is replaced by U, c by m, and e^{-q} by b.

where V is the cumulative number of violent acts and c and q are two parameters. The solution of (17) is the Gompertz curve with V_o the initial level of V:

$$(18)\quad V = V_o e^{\frac{c}{q}} e^{-\frac{c}{q}e^{-qt}}$$

We now want to construct the stochastic counterpart of the deterministic model. Particularly interesting is the question whether the deterministic curve (18) is identical to the expected value curve of the stochastic model.

If V is split into the initial number of violent acts V_o and the cumulative number k of new imitations since t=0 then the transition rate of the stochastic growth process takes the form:

$$(19)\quad \lambda_{(k,t)} = ce^{-qt}(V_o+k)$$

(19) is the transition rate of a time-dependent "epidemic" process, a so-called time-dependent Yule process which is a special case of a birth process. If $ce^{-qt}(V_o+k)$ is substituted for $\lambda_{(k,t)}$ in equations (2) and (3) we obtain as a solution of the system of differential equations (see Diekmann 1979, III,5, appendix 2) the density distribution of the new imitations k:

$$(20)\quad p_{(t)k} = \binom{V_o+k-1}{k} p^{V_o} q^k$$

This is a negative binomial distribution with $p=\exp\left\{-\frac{c}{q}(1-e^{-qt})\right\}$ and q=1-p. Because V is the sum of V_o and k ($k=V-V_o$) the distribution of $V=V_o,V_o+1,V_o+2,\ldots$ takes the form:

$$(21) \quad p_{(t)V} = \binom{V-1}{V-V_o} p^{V_o} q^{V-V_o}$$

with the expected value:

$$(22) \quad E(V) = V_o e^{\frac{c}{q}} e^{-\frac{c}{q}e^{-qt}}$$

It can be seen that the expected value curve (22) is identical to the deterministic function (18)*. This is not a self-evident result. There are cases where the deterministic function deviates from the expected value function (see e.g. Bartholomew 1967: 306-307). Therefore the deterministic diffusion model is a special case of the stochastic model.

Let us now consider the variance of the process, i.e. the variance of (20) or (21):

$$(23) \quad Var(V) = V_o \left[\exp\left\{ 2c/q(1-e^{-qt})\right\} - \exp\left\{ c/q(1-e^{-qt})\right\}\right]$$

If t is increasing the variance is also increasing. As t approaches infinity the variance approaches the limit $V_o(e^{2c/q}-e^{c/q})$. This means that predictions will become less certain if t is increasing, but the growth of uncertainty declines and will approach zero in the limit.

* It should be emphasized that the identity of the deterministic and the expected value function refers to the stochastic version of equation (17). A stochastic reformulation of the two original equations (equations (5) and (6) in Pitcher et al. 1978) of the model may lead to other results.

4. Discussion

a) Pitcher et al. (1978) and Hamblin et al. (1979) estimated the parameters by a non-linear regression technique. It can be shown that the regression estimates are not identical with maximum likelihood estimates (Diekmann 1979, chapt. IV). Because of the increasing variance least squares estimation faces the problem of heteroscedascity. Therefore the more efficient maximum likelihood estimates should be preferred.

b) Another problem is the interpretation of the r^2-values. At first sight the data fit of the deterministic model in terms of the r^2 statistic appears extremely good. The r^2 values reported by Pitcher et al. (1978) range from 0. 941 to 0. 999. Although the model might describe the data very well, it should not be overlooked that the data are cumulative and hence monotonous by definition. It is supposed that even a random process model might yield r^2 values above 0.90. A random process model is a Poisson process with a constant transition rate λ . The expected value curve of a random process is a straight line as can be seen from (8) for $\alpha = 1$. Consequently we suggest comparing the r^2 value of the imitation model (r_G^2) to the r^2 of a random model (r_R^2). The difference $\delta = r_G^2 - r_R^2$ can be interpreted as a measure of the "explanatory power" of the theory.

c) The two original equations (equations (5) and (6) in Pitcher et al. 1978) of the Pitcher et al. model are based on the assumption that every violent act and every "inhibited unit" has the <u>same</u> influence on potential imitators. That means a violent act that happened a long time ago has the same influence as a violent act that happened only recently. However, if we look at the resulting central equation (17) there is another possible interpretation: According to the Pitcher et al. theory the growth rate ce^{-qt} represents the instigation and the <u>inhibition</u> effect. But instead of assuming an inhibition effect the growth rate might simply mirror the fact that there is a decaying influence of violent acts. The model is completely compatible with this rival explanation that seems more plausible to me than the somewhat dubious inhibition equation.

The same argument applies to equation (9) of the use diffusion model.

d) The model of Pitcher et al. allows the derivation of the deterministic growth equations and the derivation of the transition rate of the stochastic model from the imitation <u>theory</u>. Therefore the transition rate is justified by the theory. But from which theory can the transition rate of the Petersen model be derived? In my opinion Petersen's model lacks somewhat in the point that there is - as far as I know from the english publications - no explicit correspondence between a theoretical proposition and the transition rate. To put it in other words: I miss the first link in the deduction chain: theoretical proposition $\rightarrow$ transition rate $\rightarrow$ differential equations for $k=0,1,2,\ldots \rightarrow$ density distribution.

e) From a mathematical point of view stochastic models are much more complicated than their deterministic counterparts. So there are reasons to ask the question whether stochastic models pay off. The answer is "yes" if we are interested in a deeper understanding of the model, the comparison of the deterministic and the expected value curve, the construction of prediction intervals, and the derivation of maximum likelihood estimates.

References

Bartholomew, David J.
1967 Stochastic Models for Social Processes. New York: Wiley.

Chiang, Chin L.
1968 Introduction to Stochastic Processes in Biostatistics. New York: Wiley.

Diekmann, Andreas
1979 Dynamische Modelle sozialer Prozesse. Theoretische Ansätze zur Erklärung krimineller Prozesse und Möglichkeiten ihrer Formalisierung. Unpublished dissertation. University of Hamburg.

Hamblin, Robert L., Jerry L.L. Miller and D. Eugene Saxton
1979 Modeling Use Diffusion. MDN 4: no. 3.

Petersen, Ib Damgaard
1978 Laws of Development of Cybernetic Systems. The case of the Sino-Indian border conflict 1959-64. Forskningsrapport 2, Institute of Political Studies, University of Copenhagen.

Petersen, Ib Damgaard
1979a Laws of Development of International Systems. MDN 4: no. 3.

Petersen, Ib Damgaard
1979b Kybernetiske Systemers Udviklingslove. Kobenhavn.

Pitcher, Brian L., Robert L. Hamblin and Jerry L.L. Miller
Diffusion of Collective Violence. American Sociological Review 43: 23-35.