THE APPLICATION OF DIFFERENTIAL EQUATIONS TO MATHEMATICAL SOCIOLOGY

by

FOSTER GENE DIECKHOFF

B. S., Fort Hays Kansas State College, 1970

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

Approved by:

Richard J. Greechie
Major Professor

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Louis M. Herman and Dr. E. L. Marsden for their comments and assistance while writing this report. A special thanks to Dr. Richard Greechie, Professor of Mathematics, and Dr. George Peters, Professor of Sociology, for their guidence and encouragement when it was most needed. I also wish to thank my wife, Tonda, for typing the final manuscript.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Recent years have shown an increased interest in applying mathematics to the social sciences. The power of mathematics has been demonstrated in its application to many areas of natural science, especially physics. If a mathematical model can be built for a given physical or social phenomenon, the mathematical symbols ". . . can serve as a proxy for experimental manipulation of the objects themselves, so that behavior of an actual object may easily be predicted by the behavior of these symbols alone. . . ." (Coleman, 5). The very idea of experimenting with something so unpredictable as people demonstrates the value of an accurate mathematical model to predict social phenomena. In addition to this "prediction" process, the mathematization of a set of axioms may reveal hidden relations between variables.

The reason that mathematics is such a cumbersome tool for the behavorial scientist may be considered on a global and local level. At the global level the social scientist sees different rules governing different societies. This inhibits the type of generality that is conducive to mathematics. For example, for the physicist, it is the same gravity (in a mathematical sense) that pulls a ball to earth no matter where on the globe it is dropped. Hence, Newton's law of gravitation has an universal scope and the same situation exists for many axioms in physics. On the other hand, the sociologist's axioms may have a very narrow scope. For example, Hoffmann (7) states that his heirarchy of eleven theorems concerning the Pawnee marriage rules does not have "wide application".

That is, they apply only to one small portion of the world's population.

At the local level, the most obvious difficulty confronting the mathematical sociologist is one of measurement. The physicist can measure time, distance, mass and force with considerable precision. These measurements also provide a ready check for the equations used to describe the system. On the other hand, suppose the sociologist makes the observation that "the level of group friendliness will increase if the actual level of interaction is higher than that 'appropriate' to the existing level of friendliness" (Simon, 9). Using Simon's notation, let:

$I(t)$ = the intensity of interaction among members

$F(t)$ = the level of friendliness among the members

b, β = positive constants

($I(t)$ and $F(t)$ indicate that intensity and friendliness are functions of time.) Then, Simon (9) translated this axiom into the differential equation

$$\frac{dF(t)}{dt} = b\left[I(t) - \beta F(t)\right] \quad . \tag{1.0}$$

The value of such an equation seems doubtful and among the reasons for this is that fact that the sociologist has only a crude "friendliometer" to determine $F(t)$, the level of friendliness at time t. The situation is similiar for many sociological variables. However, this may not entirely discount the usefulness of equations like (1.0). Indeed, one may still be able to discover new relationships between $F(t)$ and $I(t)$ that are brought to light through symbolic representation. Even if this fails, all is not lost becuase the language of a proposition is more precise in symbolic form than in a verbal representation. As Coleman (5) stated, "verbal statements easily hide ambiguity, mathematical ones do not."

CHAPTER II

THE DETERMINISTIC MODEL

With this bit of meta-social verbosity behind us we proceed with the primary function of this paper: To demonstrate the progress that has been made in mathematical sociology by expounding on several concrete models applicable to social phenomena. The models to be considered fall into two classes, the deterministic (explicit) model and the stochastic (probabilistic) model. (For a comparison between the two procedures being applied to model the spread of an epidemic in a population, see Baily, 1.)

First, we consider the deterministic model. For each example, we shall: (1) state the assumptions made about the given process to be modeled in order to adequately "close" the process and obtain a workable model, (2) describe the process or phenomena to be modeled, (3) write the equations involved in the model and solve them, and finally, (4) point out the virtues and/or weaknesses of the given model.

Social Diffusion

Social diffusion lends itself well to the deterministic model. One of the simplest social diffusion theories stated verbally is the following: ". . . the rate of propagation of the attribute (i.e., piece of information, behavorial innovation, belief, style) is proportional to the number of people who already have it" (Coleman, 5). Coleman (5) models this verbal proposition in the following manner: Let t = time, let x = the number of people who have the attribute at time t, and let k = diffusion constant of proportionality. The assumptions made to build the model are of a mathematical nature. It is assumed that x is a differentiable function of time. Hence, it is implicity implied that the diffusion process is taking place continuously through time. (This is because differentiability implies continuity.) Specifically, Coleman's (5) mathematical translation was the following:

$$\frac{dx}{dt} = kx \; . \tag{1.1}$$

Solving equation (1.1) to the boundary conditions that $x = 1$ when $t = 0$, we have

$$x = e^{kt} \; . \tag{1.2}$$

Equation (1.2) may be used for predictive purposes. That is, knowing k, one easily calculates x, the number of individuals who have the attribute at time t. Of course, such a tidy solution has its drawbacks. First, one may wonder how to find k. This must be done experimentally and could easily be different for different attributes. It is also assumed that as each member receives the attribute, he has a probability of passing it to someone else. (The nature of this probability if discussed in CHAPTER III.) A more subtle difficulty has also crept into equation (1.2). One observes from equation (1.2) that

$$\frac{dx}{dt} = ke^{kt} > 0$$

provided k and t are positive (which is experimentally the case). Consequently, (1.2) is an increasing function of time. That is, there is no constraint on the size of the population which may receive the attribute. This leads to a restatement of the original verbal proposition: ". . . the rate of diffusion (or growth) is proportional to the number who already have the attribute and to the number who have yet to receive it" (5).

In order to build the mathematical model for this modified diffusion theory, we must define a new constant: Let N = size of population. Provided with the same mathematical assumptions as in the previous example, the model was as follows (5):

$$\frac{dx}{dt} = kx(N-x) \ . \tag{1.3}$$

Solving (1.3) by separation of variables and the method of partial fractions to the boundary conditions that x = 1 when t = 0, we have

$$\int \frac{dx}{kx(N-x)} = \int dt \ .$$

By partial fractions, this becomes

$$\frac{1}{Nk}\left(\int\frac{dx}{x} + \int\frac{dx}{N-x}\right) = \int dt$$

so

$$\frac{1}{Nk}\left(\ln x - \ln(N-x)\right) = t + C$$

hence

$$C = \frac{-\ln(N-x)}{Nk}$$

so that

$$x = \frac{Ne^{Nkt}}{N-1 + e^{Nkt}} \tag{1.4}$$

Equation (1.4) can be used to predict the number of people who have the attribute at time t provided N and k are known. However, again there is an implicit assumption built into the verbal proposition that gave rise to (1.4). This assumption asserts that as each member of the population receives the attribute, he has a chance to pass it to someone else in the population. Clearly, this is not always the case. This leads us to consider a third diffusion model that assumes a constant source of propagation. More explicitly, we assume a limited population and that the number of individuals receiving the attribute per unit of time is proportional only to the number of individuals who have not yet received the attribute (5). Coleman's (5) mathematical translation of this was as follows:

(1.5) $$\frac{dx}{dt} = k(N-x) \ .$$

Solving (1.4), prescribed to the boundary conditions that $x = 0$ when $t = 0$, we have

(1.6) $$x = N(1 - e^{-kt}) \ .$$

Equation (1.6) can also be used for predictive purposes provided k can be experimentally determined and the assumptions of the model are not violated.

So far, we have discussed three distinct models of the diffusion of a social attribute through a society. The last two models, that yield equations (1.4) and (1.6), have been shown to give reasonably accurate predictions when properly used (5).

Hummon's Model of Blau's Axioms Concerning Differentiation in Organizations

As a final example of the deterministic model, we turn to a recent model that Hummon (8) constructed from Blau's (4) axioms concerning differentiation in organizations. This example emphasizes the importance of recognizing an implicit assumption in a given model. Hummon's (8) notation will be used throughout this example. The verbal axioms that concern us are Blau's (4),

> V1.0 "Increasing size generates structural differentiation in organizations along various dimensions at decelerating rates."

V1.2 "The larger an organization is, the larger the average size of its structural components."

Hummon's model of these two verbal propositions was as follows: Let D = differentiation (By this we must mean the number of differentiated classes or hierarchical levels in an organization.) and let S = size of the organization in number of members. The first assumption that Hummon (8) made to build the model was that the differentiation D was a function only of the size S of the organization. ". . . Blau (4) found that size can be used as the main independent variable in a formal theory for differentiation in organizations" (8). Symbolically:

$$D = f(S) \ .$$

Now Hummon assumed that D was a twice differentiable function of S. This implied that the social differentiation process took place continuously with respect to size. One recognizes that this only idealized the situation and provided no detremental effects. Hummon's mathematical translation of V1.0 was,

M1.0 (a) $\frac{df(S)}{dS} = \frac{df}{dS} > 0$

(b) $\frac{d^2f}{dS^2} < 0$

Continuing the model, let C = average size of a component of the organization. Then,

$$C = \frac{S}{D} = \frac{S}{f(S)}$$

Blau's V1.2 then translated into Hummon's mathematical statement

M1.2 $\frac{dC}{dS} > 0.$

So far, the model has done nothing more than formalize Blau's axioms; we have no predictive equation to compare with equations (1.4) and

(1.6). However, Hummon (8) claimed to have discovered a "new" (Blau, 3) relationship between the average size of a component C and the size of the organization S. The mathematical statement of Hummon's "new" proposition was

$$\text{M1.7} \qquad \frac{d^2C}{dS^2} > 0, \text{ if } \frac{d^2f}{dS^2} < \frac{-f}{2S^2} .$$

Hummon then proceeded with a valid mathematical argument that established M1.7.

His argument was essentially this. By direct calculation one finds that

$$\frac{d^2C}{dS^2} = -\frac{S}{f^2}\frac{d^2f}{dS^2} - \frac{2}{f^2}\frac{df}{dS} + \frac{2S}{f^3}\;{\frac{df}{dS}}^{2} .$$

Multiplying both sides by Sf, we have

$$Sf\,\frac{d^2C}{dS^2} = -\frac{S^2}{f}\frac{d^2f}{dS^2} - 2\,\frac{S}{f}\frac{df}{dS} + 2\,\frac{S^2}{f^2}\left(\frac{df}{dS}\right)^2 .$$

Put

$$p = \frac{S}{f}\frac{df}{dS} \text{ and } q = \frac{S^2}{f}\frac{d^2f}{dS^2}$$

to get

$$Sf\,\frac{d^2C}{dS^2} = 2p^2 - 2p + q .$$

By M1.0 and the obvious fact that $Sf > 0$, we see that $p > 0$ and $q > 0$. So we have that if

$$2p^2 - 2p + q > 0, \text{ then } \frac{d^2C}{dS^2} > 0.$$

Investigating $y = 2p^2 - 2p + q > 0$, we find that y takes its minimum when $p = \frac{1}{2}$. Hence, we want

$$\tfrac{1}{2} - 1 + q > 0,$$

or $q > \frac{1}{2}$. This, by definition of q, says

$$\frac{d^2f}{dS^2} < \frac{-f}{2S^2} .$$

So indeed if (1.7) holds, we have

$$\frac{d^2C}{dS^2} > 0 .$$

The verbal translation of M1.7, according to Hummon (8) was the following:

> V1.7 "In practical terms it states that with increasing organizational size, the rate of increase in the size of the average component itself increases."

It is clear that the mathematical statement of V1.7 is simply

$$\frac{d^2C}{dS^2} > 0 .$$

Specifically, Hummon (8) is assuming that

$$\frac{d^2f}{dS^2} < \frac{-f(S)}{2S^2} \tag{1.7}$$

always holds. So the pertinent question becomes: Do there exist functions which could be utilized in this model such that the assumptions of M1.7 do not hold? That is, do there exist functions relating S and D in such a way that (1.7) does not hold?

Blau's Counterexample

Blau (3) offered one such example. Namely suppose one considers organizational sizes of 100, 200, 400, and 800 employees (these are the values for S), with 5, 10, 20, and 40 components (these are the values of D) respectively. One identifies the function f, relating S and D to be linear in this example. In fact, we can explicitly write the rule for this function. Define f by

$$f(S) = \frac{S}{20} \cdot$$

Then, for example, when S = 100, we have that f(S) = f(100) = 5. Now assuming that f is twice differentiable, we find

$$\frac{df}{dS} = \frac{1}{20} > 0,$$

satisfying M1.0 (a). However,

$$\frac{d^2f}{dS^2} = 0 \quad \text{and} \quad \frac{-f(S)}{2S^2} < 0$$

(for all $S > 0$, which is obviously the case). Hence, we have that

$$\frac{d^2f}{dS^2} > \frac{-f(S)}{2S^2}$$

which is contrary to Hummon's assumption (1.7). The fact that

$$\frac{d^2f}{dS^2} = 0$$

suggests that one way which to reword V1.0 slightly to read: Increasing size generates structural differentiation in organizations along various dimensions at <u>non-increasing</u> rates. Then, making the corresponding change in M1.0 (b), we have

$$\frac{d^2f}{dS^2} \leq 0 .$$

Furthermore, if one writes Hummon's M1.7 to agree with Blau's example, (and a model should agree with the facts known about the system), we would have the exact opposite:

$$\frac{d^2C}{dS^2} \leq 0 .$$

In conclusion, we observe that the mathematical statement of V1.7 is:

$$\frac{d^2C}{dS^2} \gtrless 0 .$$

Hummon (8) shows that

$$\frac{d^2C}{dS^2} > 0, \text{ if } \frac{d^2f}{dS^2} < \frac{-f}{2S^2} .$$

Blau (4) offers a verbal counterexample which has been formalized to demonstrate that the hypothesis (1.7) of M1.7 does not always hold. Hence, one can not conclude that

$$\frac{d^2C}{dS^2} > 0$$

is always the case. Therefore, Hummon's "new" proposition is mathematically valid, but it is not sociologically relevent.

In fact, as Blau (4) suggested, empirical evidence is in favor of adapting another axiom to say the exact opposite is true; mathematically

$$\frac{d^2C}{dS^2} \leq 0 .$$

Verbally this says: With increasing organizational size, the rate of increase in the size of the average component increases at a non-increasing rate.

Technically, Blau's counterexample to Hummon's model is not mathematically correct. Specifically, Blau's counterexample did not fit Hummon's assumptions perfectly. We had to change $>$ to $\geq$ in several instances to make Blau's example work. It turns out that Blau's example is mathematically a "borderline" case. If one defines f(S) by

$$f(S) = \sqrt{S}$$

no changes need be made in Hummon's hypothesis and we get a mathematically sound counterexample to Hummon's "new" proposition.

As a final observation, we notice that the parameter C defined by

$$C = \frac{S}{D}$$

is simply the average size of a given component in the organization. Suppose a given organization has 1 executive director, 3 assistant directors, and 10 workers under each assistant director. Then S = 34, D = 3 so that

$$C = \frac{S}{D} = \frac{34}{3} = 11\ \frac{1}{3}$$

Clearly, there are not, on the average, 11 1/3 directors, nor 11 1/3 assistant directors. The only component that C comes close to representing is the workers. So, it seems that C is not an adequate parameter to investigate. One might be led to believe that studying the behavior of a parameter (such as C) which has an insignificant sociological interpretation and even less mathematical meaning is little more than a mathematical exercise.

CHAPTER III

THE STOCHASTIC MODEL

Definitions Concerning the Stochastic Process

The first thing that must be established is that "the terms 'stochastic process' and 'random process' are synonyms and cover practically all the theories of probability from coin tossing to harmonic analysis. In practice, the term 'stochastic process' is used mostly when a time parameter is introduced" (Feller, 6). What follows is not an attempt to characterize the stochastic process, but a dictionary of terms that are used in the stochastic models that follow.

We begin by defining a probability space as an ordered triple $(\Omega, \mathcal{A}, P)$. Where Ω is a non-empty set called the sample space and elements ω of Ω are called sample points. $\mathcal{A}$ is a σ-algebra of subsets of Ω and elements A of $\mathcal{A}$ are called events. P is a measure function defined on $\mathcal{A}$ satisfying: (1) $P(\Omega) = 1$

(2) $P(A) \geq 0$ for each A

(3) If $A_1, A_2, A_3, \ldots$ is any pair-wise disjoint sequence of events, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad .$$

Definition: A random variable X is a measurable, real valued function defined on Ω. ($X: \Omega \rightarrow R$, where R is the real number system.)

The stochastic models to be discussed involve a non-empty set T called the parameter set. (For us, T will represent time, hence $T = [0,\infty)$.) Furthermore, we consider an experiment such that every $t \in T$ constitutes a trial, and we are concerned only with some binary circumstance. For example, we might consider whether a friendship exists between any two people in a given population, the answer being yes or no. The set $\{Yes, No\}$ does not constitute the sample space, in fact, objects like $\{Yes\}$ are not even in the sample space. Our stochastic process will also involve counting the number of Yeses that occur up to time t. Hence, we define a function

$$r: T \to I^+ \cup \{0\}$$

by $r(t) = n$ if there are exactly n Yeses by time t. Hence, our sample space consists of the function r in the set-theoretical sense. That is

$$\Omega = \{(t, r(t) \mid t \in T\}$$

Such a function r is easily constructed for an actual experiment by counting the circumstances up to time t. The difficulty of too many circumstances occuring in too short a time is alleviated in a later assumption. Notice that Ω could also be considered as the set of all non-decreasing, non-negative, integer valued functions on T. That is, each point of Ω is a function (6). Without lose of generality we pick the only significant such function.

Finally, we are concerned with assigning a probability to each event $A \in \mathcal{A}$. That is, we assign a probability to the event that by time t there have been n Yeses observed. (Write $P(X_t = n)$.) Define $(X_t = n)$ by

$$(X_t = n) = \{(t, n) \mid r(t) = n\} \quad .$$

Henceforth we assume $\mathcal{A}$ to be the power set of Ω (Write $\mathcal{P}(\Omega)$.). Thus, we have $(X_t = n) \in \mathcal{A}$ for each $t \in T$ as desired.

Now we define the distribution function for each fixed but arbitrary $n_o \in I^+ \cup \{0\}$ by

$$f_{n_o}: \{(t, n_o) \mid t \in T\} \rightarrow [0, 1]$$

by

$$f_{n_o}(t, n_o) = P(X_t = n_o) \quad .$$

Notice that each f_n may be considered as a function on T.

Definition: Let X be a random variable with range $x_1, x_2, x_3, \ldots$ . . Then the expected value of X (Write E(X).) is given by

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j) \quad .$$

E(X) is sometimes called the mean value or average value of S. We say E(X) is defined only when the indicated series converges absolutely, otherwise a reindexing of the range of X may lead to a different value for E(X).

Definition: The probabilities of two events, say A and B, are independent if P(A, B) (read the probability of A and B) is given by P(A) $\cdot$ P(B).

The Poisson Process via Differential Equations

Equipped with these definitions, we proceed to construct our first example of a stochastic process, namely the Poisson process. The Poisson probability distribution was originally discovered as a special limiting case of the Binomial or Bernoulli distribution (see Feller, 6 and Thomasian, 10). However, it has been discovered that the Poisson distribution can be derived from a few mathematical assumptions which have a sociological interpretation. Coleman (5) stated that the appropriateness of the Poisson process for social phenomena lies in the assumptions on which the distribution is based. More precisely, Coleman (5) stated three reasons for studying the Poisson process as derived from differential equations:

(1) The Poisson process "deals with numbers of elements" or "numbers of events".

(2) "The Poisson process occures continuously over time".

(3) ". . . the Poisson process is appropriate to social phenomena because it constitutes a rational model whose assumptions can mirror our assumptions about actual phenomena. Thus, it need not be simply an empirical frequency because it fits the data".

Since Coleman's development of the Poisson process is not mathematically precise, so we follow the suggestions of Thomasian (10), Bishir and Drewes (2), and Feller (6). The assumptions we need are the following:

(1) The process has stationary increments. Intuitively, this means the number of realizations (occurences, favorable outcomes) in a given time interval depends only on the length of the interval and not on its location. Mathematically, if u, v, u', v' are elements of the parameter set T, and u - v = u' - v' then $X_u - X_v = X_{u'} - X_{v'}$.

(2) The process has independent increments. Intuitively, this means the probabilities of a realization in two disjoint time intervals is independent. Mathematically, if a, t, u, v are elements of T, with s, t, u, v, then $X_t - X_s$ and $X_v - X_u$ are independent events, that is $P\ (X_t - X_s), (X_v - X_u)\ = P(X_t - X_s) \cdot P(X_v - X_u)$.

(3) There exists a positive constant λ such that for small h:

(a) The probability of no realization in (0, h] (an interval in T) is approximately 1 - λ h. Symbolically, $P(X_h = 0) \cong 1 - \lambda h$.

(b) The probability of exactly one realization in (0, h] is approximately λ h. Symbolically, $P(X_h = 1) \cong \lambda h$. (We might interpret this as saying that in small intervals the probability of a realization is proportional to the length of the interval.).

(c) The probability of more than one realization in (0, h] is negligible. Symbolically, $P(X_h = 2) \cong 0$.

(4) The range of X_t for each t in T is the non-negative integers. We might say that the process has counting function realizations.

A more formal mathematical formulation of (3) above is vital to our development of the Poisson process. First, we consider (3a). We may define a function o_1: $T \rightarrow R$ in such a way that

$$o_1(h) = P(X_h = 0) - (1 - \lambda h).$$

Hence, we may rewrite (3a) as

$$\text{(3a')} \qquad P(X_h = 0) = 1 - \lambda h + o_1(h) \ .$$

In addition, we want

$$\lim_{h \to 0} \frac{o_1(h)}{h} = 0 \ .$$

Consequently, as the interval becomes smaller, the approximation in (3a) becomes better. An analagous procedure may be applied to (3b) and (3c), using functions $o_2(h)$ and $o_3(h)$. Then, because we are interested only in the limiting case as $h \rightarrow 0$ in the future, we may drop the subscripts (for $h \rightarrow 0$) and write,

$$\begin{aligned} (3) \quad &(a') \quad P(X_h = 0) = 1 - \lambda h + o(h) \\ &(b') \quad P(X_h = 1) = \lambda h + o(h) \\ &(c') \quad P(X_h \geq 2) = o(h) \end{aligned}$$

where

$$\lim_{h \to 0} \frac{o(h)}{h} = 0 \ .$$

To derive the probability distribution for a process obeying these assumptions, we write

$$\text{(2.0)} \qquad X_{t+h} = X_t + (X_{t+h} - X_t)$$

$$\text{(2.1)} \qquad (X_{t+h} = n) = \bigcup_{k=0}^{n} \left[X_t = k, X_{t+h} - X_t = n - k\right]$$

for n and k, non-negative integers, and $n \geq k$. Equation (2.0) has the obvious interpretation. Equation (2.1) says that if there are n realizations by time t + h then in time t there were 1 or 2 or 3 or . . . or n realizations. While the remaining (n - k) realizations come in the interval $(t, t+h]$. Now, we define a family of functions f_n: $T \rightarrow [0,1]$ by

$$f_n(t) = P(X_t = n)$$

where $P(X_t = n)$ denotes the probability that there are exactly n realizations by time t. Now, using assumptions 1 and 2, and equations (2.1) and (2.2), we calculate

$$f_0(t+h) = P(X_{t+h} = 0) = P\left[(X_t = 0),\ X_{t+h} - X_t = 0\right] \quad \text{by (2.1)}$$

$$= P\left[(X_t = 0),\ X_h = 0\right] \quad \text{by (1)}$$

$$= P(X_t = 0) \cdot P(X_h = 0) = f_0(t) \cdot f_0(h) \ . \quad \text{by (2)}$$

Similiarly, $\quad f_1(t+h) = f_0(t)\, f_1(h) + f_1(t)\, f_0(h)$,

and $\quad f_2(t+h) = f_0(t) \cdot f_2(h) + f_1(t) \cdot f_1(h) + f_2(t) \cdot f_0(h)$.

In general

$$f_n(t+h) = \sum_{k=0}^{\infty} f_k(t)\, f_{n-k}(h) \ . \tag{2.2}$$

Using assumption (3) with equation (2.2), we find for n = 0,

$$f_0(t+h) = f_0(t)\, f_0(h)$$

$$= f_0(t)\left[1 - \quad h + (0)\, h\right]$$

$$= f_0(t) - \lambda h\, f_0(t) + f_0(t)\, o(h)$$

$$= f_0(t) - \lambda h\, f_0(t) + o(h) \ ,$$

for h sufficiently close to 0.

For n = 1

$$f_1(t + h) = f_0(t)\, f_1(h) + f_1(t)\, f_0(h)$$

$$= f_0(t)\left[\lambda h + o(h)\right] + f_1(t)\left[1 - \lambda h + o(h)\right]$$

$$= f_0(t)\lambda h + f_1(t)(1 - \lambda h) + o(h)\ .$$

In general, for $n - k \geq 2$, $f_{n-k}(h) = o(h)$. Hence, for $n \geq 1$

$$f_n(t + h) = \lambda h(f_{n-1}(t) + (1 - \lambda h)f_n(t) + o(h). \tag{2.3}$$

Now, we can show that $f_n(t)$ is differentiable on the parameter set T. For $n = 0$, we have

$$\lim_{h\to 0} \frac{f_0(t + h) - f_0(t)}{h}$$

$$= \lim_{h\to 0} \frac{f_0(t) - h\, f_0(t) + o(h) - f_0(t)}{h}$$

$$= \lim_{h\to 0} \frac{-\, h\, f_0(t) + o(h)}{h}$$

$$= -\lambda f_0(t)\ .$$

For $n \geq 1$, we have

$$\lim_{h\to 0} \frac{f_n(t + h) - f_n(t)}{h}$$

$$= \lim_{h\to 0} \frac{\lambda h\left[f_{n-1}(t)\right] + (1 - \lambda h)f_n(t) + o(h) - f_n(t)}{h}$$

$$= \lambda f_{n-1}(t) - \lambda f_n(t)\ .$$

Hence $f_n(t)$ is differentiable on T for $n \geq 0$. (The somewhat special treatment of $f_0(t)$ could be avoided by defining $f_n(t) = 0$ for $n < 0$.)

In any case, writing $f'_n(t)$, for the indicated derivitive, we now have a countable set of differential equations,

$$f'_0(t) = -\lambda f_0(t) \tag{2.4}$$

and

$$f'_n(t) = -\lambda f_n(t) + \lambda f_{n-1}(t) \tag{2.5}$$

for $n \geq 1$.

To solve (2.4), we have $f_0(t) = e^{-\lambda t}C$ for some constant C. Since $X_0 = 0$, for $f_0(0) = 1$, we have

$$f_0(t) = e^{-\lambda t} . \tag{2.6}$$

Now suppose

$$f_n(0) = 0 \quad \text{for } n \geq 1. \tag{2.7}$$

Using operator notation for (2.5), we have

$$(D + \lambda)f_n(t) = \lambda f_{n-1}(t) . \tag{2.8}$$

For $n = 1$

$$f_{n-1}(t) = f_0(t)$$

So (2.8) becomes

$$(D + \lambda)f_1(t) = \lambda e^{-\lambda t} . \tag{2.9}$$

The integrating factor for (2.9) is $e^{+\lambda t}$. Using this, (2.9) becomes

$$D\left[e^{+\lambda t}f_1(t)\right] = \lambda \tag{2.10}$$

Therefore, using the boundary conditions of (2.7), we have

$$f_1(t) = \lambda t\, e^{-\lambda t}. \tag{2.11}$$

We continue to show, by induction on n, that

$$f_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad . \tag{2.12}$$

We have shown (2.12) true for $n = 1$. Suppose (2.12) is true for $n = k$, then (2.12) becomes

$$f_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad . \tag{2.13}$$

By (2.5), we have

$$f'_{k+1}(t) = -\lambda f_{k+1}(t) + f_k(t) \ . \tag{2.14}$$

By the induction hypothesis, (2.14) becomes

$$f_{k+1} = -\lambda f_{k+1}(t) + \frac{\lambda (\lambda t)^k e^{-\lambda t}}{k!} \quad ,$$

or in operator notation

$$(D + \lambda) f_{k+1}(t) = \frac{\lambda (\lambda t)^k e^{-\lambda t}}{k!} \quad . \tag{2.15}$$

Using $e^{+\lambda t}$ as an integrating factor, (2.15) becomes

$$D\left[e^{+\lambda t} f_{k+1}(t)\right] = \frac{\lambda (\lambda t)^k}{k!} \quad . \tag{2.16}$$

Again, using the boundary conditions of (2.7), (2.16) yields

$$f_{k+1}(t) = \frac{\lambda \lambda^k t^{k+1} e^{-\lambda t}}{k! \ k+1}$$

$$= \frac{(\lambda t)^{k+1} e^{-\lambda t}}{(k+1)!} \quad ,$$

establishing (2.12).

So we have

$$f_n(t) = P(X_t = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \text{ for } n = 0, 1, 2, \ldots, \tag{2.17}$$

which is the familiar Poisson distribution.

Writing

$$e^{\lambda t} = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} ,$$

we find that for any fixed time t

$$\sum_{n=0}^{\infty} P(X_t = n) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$$= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!}$$

$$= 1$$

as expected of a probability distribution. Another observation, from equation (2.6), is that

$$f_0(t) = P(X_t = 0) = e^{-\lambda t} .$$

So as t increases the probability of no realization decreases.

Another well known factor concerning the Poisson distribution is deduced by calculating its expected value. The expected value $E(X_t)$ is given by

$$E(X_t) = \sum_{n=0}^{\infty} n\ P(X_t = n)$$

$$= \sum_{n=0}^{\infty} n \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$$= e^{-\lambda t} \lambda t \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

$$= \lambda t.$$

So, for t = 1, the expected value $E(X_t)$ is just λ. That is, the expected number of occurences in an unit time interval (second, minute, hour) is λ. The parameter λ is sometimes called the transition intensity of the process.

An assumption governing the Poisson process was that the probability of an occurence during the interval $[t, t + h)$ was independent of the number of occurences during $[0, t)$. The next model to be discussed drops this assumption.

The Pure Birth Process (Feller, 6)

The pure birth process postulates that when n realizations (births) occur during $[0, t)$, the probability of a realization during $[t, t + h)$ is approximately $\lambda_n h$. More specifically, the single parameter λ, of the Poisson process, is replaced by a sequence $\lambda_0, \lambda_1, \lambda_2, \ldots$ where λ_i is the transition rate (transition intensity) for the system when there are i members in the population. (We still assume $\lambda_i > 0$ for each non-negative integer i.) When dealing with the pure birth process, instead of saying that n realizations occur during the interval $[0, t)$, we say that the system is in state E_n by time t. One can talk about the probability of being in state E_n by time t (Write $P_n(t)$.). The only realization to be observed in a pure birth process are the changes $E_n \rightarrow E_n$ (no change) and $E_n \rightarrow E_{n+1}$ (a birth).

Formally, the only difference in the assumptions of the pure birth process as compared to the Poisson process is that (3b') and (3c') become

$$(3) \quad (b'') \quad P(X_{t+h} = n + 1) = \lambda_n h + o(h)$$

$$(c'') \quad P(X_{t+h} \geq n + 1) = o(h)$$

Here o(h) has the same role as in the development of the Poisson process. If $P_n(t)$ replaces $f_n(t)$, the only change in equation (2.3) is that λ is replaced by the appropriate λ_i. So we have

$$(2.18) \qquad P_0(t + h) = P_0(t)(1 - \lambda_0 h)$$

and

(2.19) $$P_0(t + h) = P_0(t)(1 - \lambda_0 h) .$$

Since the λ_i have no effect on the difference quotients on page 19, we have that $P_n(t)$ is differentiable and hence, continuous on the parameter set T. So we get the equations

(2.20) $$P'_0(t) = -\lambda_0 P_0 t$$

and

(2.21) $$P'_n(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \text{ for } n \geq 1.$$

(Equations (2.20) and (2.21) are a special case of so-called Kolmogorov equations.)

As an example of the pure birth process, we consider a population whose constituents can give birth to new members but cannot die. Assume also that during any short time interval of length h each member has the probability $\lambda h + o(h)$ to create a new member, (where λ is the rate of increase of the population). Assuming no interaction among the members and that at time t the population has size n (in state E_n), the probability of an increase during $(t, t + h]$ is $n(\lambda h + o(h)$ as $h \to 0$. Hence, the probability $P_n(t)$, that the population has n members at time t satisfies the differential equations of (2.20) and (2.21), where $\lambda_n = n\lambda$. Our problem now involves solving the differential equations

(2.22) $$P'_n(t) = -n\lambda P_n(t) + (n - 1)\lambda P_{n-1}(t),$$

$n = (0, 1, 2, \ldots)$. Assuming the system is in state E_i at time $t = 0$ we have the boundary conditions that

$$P_i(0) = 1 \text{ and } P_n(0) = 0 \text{ for } n \neq i.$$

Since members of the population cannot die, $P_j(t) = 0$ for each $t \in T$ if $j < i$. So, we are concerned only for $n \geq i$. For $n = i$, (2.22) becomes

$$P'_i(t) = -i\lambda P_i(t)$$

which easily yields

(2.23) $$P_i(t) = e^{-i\lambda t} .$$

We now prove by induction that for $n \geq i$,

(2.24) $$P_n(t) = \binom{n-1}{n-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i} .$$

We see that (2.23) satisfies (2.24), so assume that (2.24) holds for $n = i, i+1, i+2, \ldots m$. This gives

(2.25) $$P_m(t) = \binom{m-1}{m-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{m-i} .$$

We must show that

(2.26) $$P_{m+1}(t) = \binom{m}{m+1-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{m+1-i} .$$

By equation (2.22)

$$P'_{m+1}(t) = -(m+1)\lambda P_{m+1}(t) + m\lambda P_m(t)$$

$$= -(m+1)\lambda P_{m+1}(t) + m\lambda \binom{m-1}{m-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{m-i}$$

using (2.25). Employing the operator notation, this becomes

$$[D + (m+1)\lambda]P_{m+1}(t) = m\lambda\binom{m-1}{m-i}e^{-i\lambda t}(1 - e^{-\lambda t})^{m-i} ,$$

for which $e^{(m+1)\lambda t}$ is an integrating factor. Hence

$$D\left[e^{(m+1)\lambda t}P_{m+1}(t)\right] = m\lambda\binom{m-1}{m-i}e^{-i\lambda t}(1 - e^{-\lambda t})^{m-i}e^{(m+1)t\lambda}$$

So

$$e^{(m+1)\lambda t}P_{m+1}(t) = m\lambda\binom{m-1}{m-i}\int e^{-i\lambda t}(1-e^{-\lambda t})^{m-i}e^{(m+1)\lambda t}\,dt$$

$$= m\lambda\binom{m-1}{m-i}\int e^{(m-i)\lambda t}(1-e^{-\lambda t})^{m-i}e^{\lambda t}\,dt$$

$$= m\lambda\binom{m-1}{m-i}\int e^{\lambda t}(e^{\lambda t}-1)^{m-i}\,dt.$$

Noticing that the integral on the right hand side is of the form

$$\int u^n du$$

where $u = (e^{\lambda t} - 1)$, we have

$$P_{m+1}(t) = m\binom{m-1}{m-i}\frac{(e^{\lambda t}-1)^{m+1-i}}{m+1-i}e^{-(m+1)\lambda t}$$

$$= \frac{m(m-1)!}{(m+1-i-m)!(m-i)!(m+1-i)}(e^{\lambda t}-1)^{n+1-i}$$

$$e^{-(n+1)\lambda t}e^{-\lambda t(n+1-i)}e^{\lambda t(n+1-i)}$$

$$= \binom{m}{m+1-i}e^{-i\lambda t}(1-e^{-\lambda t})^{m+1-i}$$

which is exactly (2.26), establishing (2.24).

Now, we investigate the behavior of the $\{P_n(t)\}_n$ defined by (2.24). One has only to look at (2.24) to see that $P_n(t) \geq 0 \; \forall \, t \in T$. However, is it true that

$$\sum_{n=0}^{\infty} P_n(t) = 1 \tag{2.27}$$

for each $t \in T$? In general, the best can say is that

$$\sum_{n=0}^{\infty} P_n(t) \leq 1 . \tag{2.28}$$

The following theorem (Feller, 6) clarifies the situation. In fact, it applies to the more general equation (2.21).

Theorem: In order that (2.27) holds, it is necessary and sufficient that the series

$$\sum_{n=0}^{\infty} \frac{1}{\lambda_n} \tag{2.29}$$

diverge.

Proof: Put

$$S_k(t) = P_0(t) + \ldots + P_k(t). \tag{2.30}$$

Then using (2.21), (2.30) gives

$$S_k'(t) = -\lambda_k P_k t \quad . \tag{2.31}$$

Hence, for $k \geq i$ (and using the fact that $P_i(0) = 1$ and $P_n(0) = 0$ for $n \neq i$), we have

$$-\lambda_k \int_0^t \lambda_k P_k(t) = \int_0^t S_k'(t) = S_k(t) - S_k(0)$$

$$= S_k(t) - 1 \ .$$

So that

$$1 - S_k(t) = \lambda_k \int_0^t P_k(\tau)\, d\tau \ . \tag{2.32}$$

Since all $P_k(t) \geq 0$ for each $t \in T$ and $\lambda_k > 0$ for each k, we have that

$$1 - S_k(t) \geq 0$$

establishing the fact that

$$\sum_{n=0}^{\infty} P_n(t) \leq 1 \quad .$$

Noticing that $S_k(t)$ is monotone increasing with k, we have that the right hand side of (2.31) is monotone decreasing with k. Then since the right hand side is also bounded below by zero we have for each t in T, that the limit as $k \to \infty$ exists. Write

$$\lim_{k\to\infty} \lambda_k \int_0^t P_k \tau \; d\tau = \mu(t),$$

and clearly, for each k we have

(2.33) $$\lambda_k \int P_k \tau \; d\tau \geq \mu(t) .$$

So,

$$\int_0^t S_n(\tau)d\tau = \int_0^t (P_0(\tau) + P_1(\tau) + \ldots P_k(\tau) \quad d\tau$$
$$= \int_0^t \frac{\lambda_0 P_0(\tau)}{\lambda_0} d\tau \quad + \int_0^t \frac{\lambda_1 P_1(\tau)}{\lambda_1} d\tau \quad + \ldots + \int \frac{\lambda_k P_k(\tau)}{\lambda_k} d\tau$$
$$\geq \mu(t) \left(\frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_k} \right)$$

Now, $S_n(t) \leq 1$, so that
$$\int_0^t S_n(\tau)d\tau \leq t .$$

Hence, if (2.29) diverges, it must be that $\mu(t) = 0$. So, as $k \to \infty$, (2.32) tells us that $S_k(t) \to 1$.

Now suppose (2.29) converges, write

(2.34) $$\sum_{n=0}^{\infty} \frac{1}{\lambda_n} = L \quad ,$$

and that $S_k(t) \to 1$ as $k \to \infty$. Then

$$\int_0^t S_k(\tau)d\tau = \int_0^t P_0(\tau) + P_1(\tau) + \ldots + P_k(\tau) \quad d\tau$$

$$= \int_0^t P_0(\tau)d\tau + \ldots + \int_0^t P_k(\tau)d\tau$$

$$= \frac{1 - S_0(\tau)}{0} + \ldots + \frac{1 - S_k(\tau)}{k}$$

$$\leq \quad \frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \ . \ . \ . \ + \frac{1}{\lambda_k} \quad .$$

Using our assumption that

$$\lim_{k \to \infty} S_k(t) = 1$$

and (2.35), together with the Lebesgue dominated convergence theorem, we find

$$\lim_{k \to \infty} \int_0^t S_k(\tau)d\tau \leq \lim_{k \to \infty} \sum_{n=0}^{\infty} \frac{1}{\lambda_n}$$

or

$$\int_0^t \lim_{k \to \infty} S_k(\tau)d\tau \leq \lim_{k \to \infty} \sum_{n=0}^{\infty} \frac{1}{\lambda_n}$$

so

$$\int_0^t d\tau \leq L$$

or

$t \leq L$ for any fixed but arbitrary t, which cannot be. So if (2.27) holds (2.29) must diverge.

To parallel our work with the Poisson distribution, it remains to find the expected value $E(X)$ for (2.24). We have

$$E(X) = \sum_{n=0}^{\infty} nP_n(t)$$

$$= \sum_{n=0}^{\infty} n \binom{n-1}{n-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i} \quad .$$

Claim:

(2.36) $$\sum_{n=0}^{\infty} n \binom{n-1}{n-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i} = ie^{\lambda t} \quad .$$

Proof: Put $x = e^{-\lambda t}$, then $0 < x < 1$.

Hence

$$\sum_{n=0}^{\infty} n \binom{n-1}{n-i} (1 - x)^{n-i} = ix^{-(i+1)}$$

So

$$\sum_{n=0}^{\infty} \frac{n(n-1)!}{i(i-1)!\,(n-i)!} (1-x)^{n-i} = x^{-(i+1)} .$$

Hence

$$\sum_{n=1}^{\infty} \binom{n}{i} (1-x)^{n-i} = \frac{1}{x^{i+1}} ,$$

is equivalent to (2.36).

Now consider the series expansion for $0 < x \leq 1$.

(2.37) $$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Differentiating (2.37) once gives

(2.38) $$\frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} nx^{n-1} .$$

Differentiating (2.38) gives

$$\frac{2}{(1-x)^3} = \sum_{n=2}^{\infty} n(n-1)x^{n-2}$$

or

$$\frac{1}{(1-x)^3} = \sum_{n=2}^{\infty} \frac{n(n-1)x^{n-2}}{2} .$$

In general

$$\frac{1}{(1-x)^{i+1}} = \sum_{n=i}^{\infty} \frac{n(n-1)\ldots(n-i)x^{n-i}}{i!}$$

establishing (2.36).

CHAPTER IV

CONCLUSIONS

To conclude our remarks on the application of differential equations to sociology, we make the following observations. Concerning the diffusion models, we noticed that the first two examples implicitly assumed that as each member of the population received a given attribute, he had some chance (probability) of giving the attribute to someone else. However, the exact role that this probability played was not explicitly formulated in the deterministic model. In spite of this, it was possible to derive equations which could be utilized to predict the number of people who have the attribute at a given time.

Hummon's model of Blau's theory of differentiation in organizations introduced two observations. First, when building a model for a given social phenomenon, one must remain aware of the sociological interpretation of the mathematical assumptions. Hummon's mathematical assumption (1.7) led him to an invalid conclusion concerning the behavior of the average size of a component of an organization, namely Hummon's VI.7. Secondly, we observed that one parameter, namely $C = \frac{S}{D}$, that Hummon investigated is of doubtful sociological or mathematical (statistical) significance.

The stochastic models presented display several points worth discussing. For example, equation (2.36) stated that the expected value (average value) for the pure birth process is given by

$$E(X) = ie^{\lambda t}$$

where the system is in state i at time t = 0. Assuming the system is in state E_1 at time $t = 0$, we observe that the expected value of the pure birth process is exactly what equation (1.2) predicted from a deterministic model. (Here one would consider a "birth" as receiving the given attribute.) A relevent question arising from this analogy is this: Given a deterministic model is there always such a corresponding stochastic model or is this purely coincidental? Certainly equation (1.2) was easier to obtain than (2.24) so such a correspondence could provide a deterministic approximation for a complicated stochastic model.

The process of formulating a given set of assumptions concerning a social phenomenon into precise mathematical statements and then using these statements to derive a probability function for the phenomenon seems to be a natural and instructive approach to mathematical sociology. Usually, the sociologist collects data concerning some phenomenon and then proceeds, by means of statistical analysis, to discover the distribution (Poisson, exponential, binomial, and so forth) that fits the data and then uses this distribution for predictive purposes. The methods discussed in this report for establishing the probability function may give the same distribution, but in addition, give mathematical formulations of the sociological structure underlying the given phenomenon. For example, if the sociologist observes that the transition intensity governing a given social phenomenon is not constant, he would know the Poisson distribution did not apply. However, suppose the transition intensity was some function of the size of the population, so that if n is the size of the population some λ_n is the proper transition intensity. We discussed this situation for $\lambda_n = n\lambda$, but what if λ_n is some other function of n, say $\lambda_n = n^2\lambda + n\lambda + C$. Equipped with the theorem on page 27, the sociologist may be able to "tailor make" a distribution to fit the observed facts concerning the phenomenon. Then, in addition to having a predictive model, he has additional abstractions governing the system under observation that are, by the usual procedure, camouflaged in a labyrinth of numbers and calculations that only a computer can assimulate.

LITERATURE CITED

1. Bailey, Norman T. J. The Mathematical Theory of Epidemics. (London: Charles Griffen, 1957).

2. Bishir, John W. and Donald W. Drewes. Mathematics in the Behavioral and Social Sciences. (New York: Harcourt, Brace and World, Inc., 1970), pp. 1, 397-411.

3. Blau, Peter M., "Comments on Two Mathematical Formulations of the Theory of Differentiation in Organization," American Sociological Review, 36:304-307, 1971.

4. Blau, Peter M., "A Formal Theory of Differentiation in Organizations," American Sociological Review, 35:201-218, 1970.

5. Coleman, James S. Introduction to Mathematical Sociology. (London: The Free Press of Glencoe Collier-Macmillan, Limited, 1964), pp. 1054, 288-314.

6. Feller, William. An Introduction to Probability Theory and its Applications, I. (New York: John Wiley and Sons, Inc., 1966), pp. 1, 397-411.

7. Hoffmann, Hans (ed.). Biennial Review of Anthropology 1969. (Stanford: Stanford University Press, 1970), pp. 41-79.

8. Hummon, Norman P., "A Mathematical Theory of Differentiation in Organizations," American Sociological Review, 36:297-303.

9. Simon, Herbert A. Models of Man. (London: John Wiley and Sons, Inc., 1967), pp. 99-114.

10. Thomasian, Aram J. The Structure of Probability Theory with Applications. (New York: McGraw-Hill Book Company, 1969), pp. 513-555.

THE APPLICATION OF DIFFERENTIAL EQUATIONS
TO MATHEMATICAL SOCIOLOGY

by

FOSTER GENE DIECKHOFF

B. S., Fort Hays Kansas State College, 1970

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

The purpose of this report was to investigate the application of mathematics, especially differential equations, to the social sciences. Several examples of deterministic and stochastic models, applicable to sociology were discussed. In each example the assumptions made to close the system under study were enumerated and the mathematical implications of these assumptions were discussed. Solutions to pertinent mathematical equations were given in detail.