DOCUMENT RESUME

ED 173 206 SO 011 764

AUTHOR Hannan, Michael T.
TITLE Final Report for Dynamic Models for Causal Analysis of Panel Data. Models for Change in Quantitative Variables, Part II Scholastic Models. Part II, Chapter 4.
INSTITUTION Center for Advanced Study in the Behavioral Sciences, Stanford, Calif.; Stanford Univ., Calif.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.; National Science Foundation, Washington, D.C.
PUB DATE Oct 78
GRANT BNS-76-22943; NIE-G-76-0082
NOTE 33p.; For related documents, see SO 011 759-772

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; *Educational Change; Measurement Techniques; Models; *Research Design; *Research Methodology; *Social Change; Social Science Research; Sociometric Techniques; *Validity

ABSTRACT

This document is part of a series of chapters described in SO 011 759. Stochastic models for the sociological analysis of change and the change process in quantitative variables are presented. The author lays groundwork for the statistical treatment of simple stochastic differential equations (SDEs) and discusses some of the continuities of qualitative and quantitative analysis as they are revealed in the study of diffusion processes. Six sections comprise the document. Section I provides an overview of the study. Section II states the properties of the white noise processes and the related Brownian motion process. Section III discusses SDEs and two approaches to solving them. In Section IV, Kolmogorov's diffusion equations and results on conditional densities are given. Section V treats a substantive example, population growth in a random environment, from two perspectives and Section VI raises the problem of the behavior of Markov diffusion processes at boundaries. (Author/KC)

Final Report for

DYNAMIC MODELS FOR CAUSAL ANALYSIS OF PANEL DATA
MODELS FOR CHANGE IN QUANTITATIVE VARIABLES, PART II
STOCHASTIC MODELS*

Michael T. Hannan

August 1978

Revised October 1978

*This report is one in a series from a collaborative project with Nancy Brandon Tuma, funded by National Institute of Education grant NIE-G-76-0082. While this report was written Hannan was a Fellow at the Center for Advanced Study in the Behavioral Sciences and was supported by National Science Foundation grant BNS76-22943. I wish to thank Nancy Brandon Tuma, Aage B. Sørensen, Glenn Carroll and Barbara Warsavage for comments on an earlier draft. This report is also available as Technical Report #64 of the Laboratory for Social Research, Stanford University.

1. Overview

In Part I (Hannan 1978) we argued for a stochastic treatment of change in quantitative variables. We have already remarked that such a focus increases considerably the level of mathematical complexity. Anything like a full treatment of the issues exceeds both our competence and the needs of most empirical analysts. Our goals are more modest. We wish to lay the minimal groundwork for the statistical treatment of simple stochastic differential equations. We also seek to sketch some of the continuities of qualitative and quantitative analysis as they are revealed in the study of diffusion processes. Our discussion provides an entry into the rapidly growing technical literature. We hope to persuade social researchers to come to grips with this still-developing field of mathematics.

We discuss two approaches to the stochastic study of change in levels. The first involves a seemingly innocent modification of the sort of deterministic models treated in the last chapter. The scientist assumes that the deterministic models hold only approximately due to a host of disturbing influences. It seems natural to add a stochastic element to the differential equation to represent such noise. This involves moving from

$$\frac{dY(t)}{dt} = f(Y(t), t)$$

to

$$\frac{dY(t)}{dt} = f(Y(t), t) + \mathcal{E}(t)$$

where $\mathcal{E}(t)$ denotes the ensemble of disturbing influences. On the view that $\mathcal{E}(t)$ involves only approximation error and is likely composed of many, independent specific elements, one chooses the simplest possible structure for $\epsilon(t)$. The overwhelming tendency is to treat $\epsilon(t)$ as a so-called white noise or delta-correlated stochastic process. The study of the properties of differential equations "driven" by white noise is the central problem in the analysis of stochastic differential equations (SDE's).

The second motivation employs the logic of probability. As we mentioned in Part I, changes in levels may be considered as the limiting case of transitions in an infinite state space of discrete states, where the limit is approached by decreasing the "width" of each state. For example, early progress in the theory of Markov diffusion processes considered random walks on the real line as the limiting case of the discrete-state random walk encountered in most elementary probability texts. As the early interest in such analysis of Markov processes in continuous state spaces (and in continuous time) arose from the study of physical diffusion processes, the term diffusion process has come to be applied broadly to these processes.

The elementary parameters in the study of diffusion processes are rates of transition just as in the discrete-state case considered earlier. Substantive hypotheses enter as restrictions on transition rates. The goal, as in the discrete-space case, is to formulate and solve equations for the evolution of probability transition densities. Then one may evaluate the effects of various rate parameters on the evolution of the process.

The approximation perspective is the typical model of entry of the substantive modeler, the diffusion perspective has more appeal to the mathematician. [illegible]ntly the former is often referred to as the physical (or engineering) p[illegible]tive while the latter is called the mathematical perspective. But the two perspectives are not that sharply different. In fact, many fundamental issues in the study of stochastic models for changes in levels concern the relationship between the two perspectives. In particular they concern the relations between stochastic differential equations and diffusion equations.

The outline of the chapter is as follows. In Section 2 we state the properties of white noise processes and the related Brownian motion process. Section 3 discusses SDE's and the two approaches to solving them. In Section 4 we derive Kolmogorov's diffusion equation and give results relating parameters of SDE's to those of diffusion equations. We also give results on conditional densities that will be used in subsequent chapters. Section 5 treats a substantive example, population growth in a random environment, from the two perspectives. Finally, Section 6 raises the problem of the behavior of Markov diffusion processes at boundaries (e.g., the complications that arise in the study of bounded processes e.g., those that are defined to be non-negative).

2. White Noise and Brownian Motion

As we mentioned above, substantive applications of SDE models almost invariably use a white noise disturbance (or forcing function). Since this choice implicitly defines a stochastic process for the substantive outcome, it is important to understand the properties of systems driven by white noise processes. We begin with the white process itself.

There are a number of approaches to defining a white noise process. One strategy (see Jazwinski 1970:81-83) begins with a discrete time process. Consider a random sequence $\{X_n: n = 1,2,\ldots\}$ with independent increments:

$$p\{X_k \mid X_j\} = p\{X_k\} \qquad (k > 1)$$

This process may be thought of as being pure noise; knowing the history of the process and its current value does not aid in predicting its future. In substantive applications, each realization of $\{X_n\}$ may be considered to be the sum of many independent forces; in technical terms, the sequence is the superposition of many independent sequences. If so, the central limit theorem applies and $\{X_n\}$ may be considered Gaussian or normally distributed. So in discrete-time analysis, white Gaussian sequences give a very simple specification for the effects of omitted variables, one that agrees closely in spirit with the specification usually used in static analysis.

With this background it seems natural to define a continuous-time analogue to white Gaussian sequences for use in forming SDE's. Thus we define a white Gaussian process $\{X_t: t \in T\}$ with independent increments:

$$p\{X_t \mid X_\tau\} = p\{X_t\} \qquad (t > \tau \in T) .$$

Perhaps the most informative approach to this process is through its <u>correlation function</u>, $\gamma(t + \tau, t) = E(X_{t+\tau} X_t)$. Suppose the process has zero mean and correlation function given by

$$\gamma^{\rho}(t + \tau, t) = \sigma^2(\rho/2)\, e^{-\rho|\tau|} \qquad (1)$$

For large ρ, this function approximates the properties we desire for the white Gaussian process. In particular, for large ρ the correlation function is very small even over brief intervals. If ρ is restricted to the integers, the sequence $\{\gamma^{\rho_1}/\sigma^2, \rho_1 > \rho_2 > \cdots\}$ defines the Dirac delta function see Braun 1975:325-35). This function, denoted $\delta(t - \tau)$, is zero everywhere but is equal to ∞ at t; its integral over any interval containing t is 1. It is not of course an ordinary function, but treating it as if it were, has given useful results in the study of impulse functions. And we can view white Gaussian noise as just such an impulse function. That is, the noise process consists of a great many independent impulses or shocks that each hold for only a brief instant.

Thus we define a white Gaussian process as a Gaussian process with

$$E\left\{(X_t - E\{X_t\})(X_\tau - E\{X_\tau\})'\right\} = Q(t)\ \delta(t - \tau) \qquad (2)$$

where Q(t) is a positive semidefinite covariance matrix and $\delta(t - \tau)$ is the Dirac delta function. This process is usually referred to as white noise or as a delta-correlated process.[1] Note that by the definition of the delta-function, white noise has zero covariance but infinite variance!

We can obtain additional insight into this process by considering the related Brownian motion or Weiner process. This widely used process is named after Robert Brown, the English botanist who discovered the irregular movement of particles suspended in a solution, and Norbert Weiner, who formalized the model of the process. It is discussed in most texts on stochastic processes, e.g., Karlin and Taylor (1975: Ch. 7).

Brownian motion is a stochastic process $\{X(t) ; t \geq 0\}$ with the following properties:

(i) Independence: $X(t + \Delta t) - X(t)$ is independent of $\{X(s)\}, s \leq t$; that is increments are independent of the entire history of the process. This assumption is stronger than the Markov assumption and implies it.

(ii) Stationarity: The distribution of $X(t + \Delta t) - X(t)$ does not depend on t.

(iii) Continuity: $\lim_{\Delta t \to 0} \frac{P|X(t + \Delta t) - X(t)| \geq \varepsilon|}{\Delta t} = 0$ for all $\varepsilon > 0$

Then if $X(0) \equiv 0$ it follows that increments in the process, $X(t + \Delta t) - X(t)$ are normally distributed with mean μt and variance $\sigma^2 t$ (see Breiman 1968: 249-50). And, any Brownian motion process may be transformed into a standard Brownian motion which is normally distributed with mean zero and variance of one. The correlation function for the process is given by

$$\gamma(t, \tau) = \sigma \min (t, \tau). \tag{3}$$

That is, its autocorrelation depends only on the time separating the realizations.

This process has obvious appeal: normality follows from very simple (but strong) assumptions about the process; and the first and second moments are simple functions of elapsed time. However, this process has some odd properties. Though sample paths of the process are continuous with probability one (wp1) it is nowhere differentiable wp1, nor does it have bounded variation wp1 (Doob 1955:393).

Finally we consider the relationship between white noise and Brownian motion. Let $\{\beta_t\}$ denote a Brownian motion process and $\{w(t)\}$ a white noise. Though we know $\frac{d\beta_t}{dt}$ does not exist, pretend for the moment that it does. Then it follows (see Jazwinski 1970:85) that white noise is the formal[2] derivative of Brownian motion:

$$\frac{d\beta_t}{dt} \triangleq w(t) \tag{4}$$

Recall that β_t is $N(\mu t, \sigma^2 t)$ when $\beta_0 = 0$. On the formalism in (4), $w(t)$ is $N(\mu, \sigma^2)$, or in the case of standard Brownian motion, $N(0,1)$. However, we have already seen that white noise has infinite variance. Thus the formalism, though suggestive, does not serve as a basis for consistent mathematical analysis.

3. Stochastic Differential Equations

We now turn to properties of processes driven by white noise or Brownian motion. We begin with a very famous special case, the Ornstein-Uhlenbeck (OU) process. Brownian motion as a description of the movement of particles in some liquid is particularly unrealistic in that it assumes that increments are independent. This amounts to ignoring the effects of a particle's velocity. The OU process corrects this in a straightforward way (see Cox and Miller 1966: 226-30; Brieman 1978: 347-51).

Let $V(t)$ be the velocity of a particle of mass m suspended in a liquid and let $m\Delta V(t)$ be the change in momentum during the period. Then

$$m\Delta V(t) = -\beta V(t)\Delta t + \Delta m(t) \qquad (5)$$

where $-\beta V$ is the viscous resisting force and Δm is the change in momentum due to random impacts with neighboring particles. As a first approximation, one may consider Δm to be a Brownian motion. If there is no drift (see below), we may write the OU process as

$$m\Delta V(t) = -\beta V(t)\Delta t + \sigma\Delta\beta_t$$

where β_t is a standard Brownian motion.

The usual next step in such models is to divide by Δt and let $\Delta t \to 0$ giving

$$m\frac{dV(t)}{dt} = -\beta V(t) + \sigma\frac{d\beta_t}{dt} \qquad (6)$$

[or $m\dfrac{dV(t)}{dt} = -\beta V(t) + \sigma w(t)$, where $w(t)$ is white noise; see (4)]

Unfortunately, as we have seen, $d\beta_t/dt$ does not exist. So (6) does not have an orthodox meaning. Suppose we ignore this and push ahead -- using formal rules. Following Brieman (1968:348) write (6) as

$$\frac{d}{dt}\left(e^{\alpha t}V(t)\right) = \gamma e^{\alpha t}\,\frac{d\beta}{dt}{}_t \qquad (7)$$

where $\alpha = \beta/m$, $\gamma = \sigma/m$. Assume $V(0) = 0$ and integrate from 0 to t to obtain

$$e^{\alpha t}V(t) = \gamma \int_0^t e^{\alpha s}\,d\beta_s. \qquad (8)$$

Doing an integration by parts gives

$$e^{\alpha t}V(t) = \gamma e^{\alpha t}\beta_t - \gamma\alpha \int_0^t \beta_s\, e^{\alpha s} ds \qquad (9)$$

Since β_t is a continuous function (wp[1]), the integral in (9) for any realization is just the integral of a continuous function and is thus well defined. Thus the process given by

$$V(t) = \gamma \int_0^t e^{\alpha(t-s)}\,d\beta_s \qquad (10)$$

can be well defined by this procedure and results in a process with continuous sample paths. The integral in (10) is termed a stochastic model.

So we find that overlooking the mathematical pathology of Brownian motion and proceeding with formal rules gives a reasonable result for the OU process. As we will see below, this is true whenever the SDE is linear, that is when the parameters of Brownian motion disturbance do not depend on the state of the process. Results on this special case are particularly useful in estimating models of the sort discussed in the previous chapter. To show this, all we need to do is add "drift" that is a function of one or more exogenous variables. Let us start with an OU process

$$\Delta Y(t) = -\beta Y(t)\,\Delta t + \mu\,\Delta t + \sigma \Delta\beta_t$$

and let $\mu = f(x)$ where x is some exogenous variable. Then, as above, we may write this as

$$\frac{dY(t)}{dt} = -\beta Y(t) + f(x) + \sigma \frac{d\beta}{dt}t \qquad (11)$$

This has the general form of the linear differential equation models with causal variables discussed in the previous chapter. And it is now clear that to understand the stochastic properties of Y(t) we must consider the distributional properties of such stochastic integrals.

Before considering stochastic integrals, we state the general problem. The general first-order SDE has the form

$$dY(t) = f(Y(t),t)\,dt + g(Y(t),t)\,d\beta_t \qquad (12)$$

As we have seen, because of the pathology of Brownian motion, such equations do not have any orthodox meaning. For this reason, it is usual to reverse the classic procedure in the calculus in which integrals are defined in terms of derivatives (as anti-derivatives) and to define (12) as the integral equation.

$$Y(t) - Y(0) = \int_{t_0}^{t} f(Y(s), s)\,ds + \int_{t_0}^{t} g(X(s),s)\,d\beta_s \qquad (13)$$

That is, the SDE is defined in terms of the stochastic integral. Thus to interpret the properties of Y(t) we must consider the existence, uniqueness, etc. of the general stochastic integral

$$\int_a^b g_s(\omega)\,d\beta_s \qquad (14)$$

where $g_s(\omega)$ is a random function.

Integrals such as (14) cannot be defined for sample functions (that is, wp 1) because of the peculiarities of Brownian motion mentioned above. They have, however, been defined in the mean square sense by Itô (1944) and Stratonovich (1966). The sequence $\{X_n\}$ is said to converge in mean square to X if $E\{|X_n|^2\} < \infty$ for all n, $E\{|X|^2\} < \infty$ and

$$\lim_{n \to \infty} E\{|X - X_n|^2\} = 0,$$

in which case we write $\text{l.i.m.}\, X_n = X$ This form of convergence implies convergence in probability (the plim convergence so commonly used in structural equation analysis) but is slightly weaker than convergence with probability one. A very clear statement of the mean square calculus is presented by Jazwinski (1970:60-70).

We consider first the general Itô definition of the stochastic integral (14). We rely on the treatment in Jazwinski (1970); the classic treatment is by Doob (1955: Ch. IX). Let T = [a,b] and let $\{\beta_t, t \in T\}$ be a scalar Brownian motion process with Var $(\beta_t) = \sigma^2$. Partition T such that

$$a = t_0 < t_1 < \ldots < t_n = b$$

and consider the step functions

$$g_t(\omega) = \begin{cases} 0 & t < t_0 \\ g_i(\omega) & t_i \le t \le t_{i+1} \\ 0 & t < t_n \end{cases}$$

Where $g_i(\omega)$ is independent of $\{\beta_{t_k} - \beta_{t_j} : t_i < t_j < t_k < b\}$, and $E\{|g_i(w)|^2\}$ is finite. For such step functions the Itô integral is defined as

$$\int_T g_t(\omega)\, d\beta_t \triangleq \sum_{i=0}^{n-1} g_i(\omega)(\beta_{t_{i+1}} - \beta_{t_i}) . \tag{15}$$

By the independence assumption

$$E\left\{\int_T g_t(\omega)\, d\beta_t\right\} = 0 \tag{16}$$

If $f_t(\omega)$ is another step function with the above properties (see Jazwinski 1970:98).

$$E\left\{\int_T g_t(\omega)\, d\beta_t \int_T f_t(\omega)\, d\beta_t\right\} = \sigma^2 \int_T E\left\{g_t\, f_t\right\} dt . \tag{17}$$

That is, under the expectation operation, stochastic integrals reduce to ordinary integrals (integration with respect to t rather than with respect to β_t). As a consequence

$$E\left\{\int_T g_t(\omega)\, d\beta_t \int_T g_t(\omega)\, d\beta_t\right\} = \sigma^2 \int_T g_t(\omega)^2\, dt . \tag{18}$$

Since we are not interested in step functions per se, consider a sequence of step functions (with successively finer cuts in T), $g_t^n(\omega)$. Suppose these converge in mean square to $g_t(\omega)$. Then it follows that

$$\int_T g_t(\omega)\, d\beta_t = \underset{n\to\infty}{\text{l.i.m.}} \int_T g_t^n(\omega)\, d\beta_t \tag{19}$$

Doob (1955) has shown that this mean square limit exists for a very broad class of functions. More generally, the Itô integral can be defined as the limit of Riemann-Stieltjes sums. Let ρ be the max $(t_{i+1}-t_i)$ in the partition of T. Then the Itô stochastic integral is defined by

$$\underset{\rho\to 0}{\text{l.i.m.}} \sum_{i=0}^{n-1} g_{t_i}(\omega)\, (\beta_{t_{i+1}} - \beta_{t_i}) = \int_T g_t(\omega)\, d\beta_t \tag{20}$$

A concrete example helps show how this definition differs from that of the usual Riemann integral. Let $g_t(\omega) = \beta_t - \beta_a$. Then (see Doob 1955:443)

$$\int_a^b (\beta_t - \beta_a)\, d\beta_t = \frac{1}{2}(\beta_b - \beta_a)^2 - \frac{\sigma^2}{2}(b-a) \tag{21}$$

Note that the usual rules of integration would give only the first term on the right hand side of (21).

Stratonovich (1966) proposed an alternative definition that is somewhat more specialized. Whereas Itô's definition holds generally, Stratonovich's approach is ". . . just versatile enough to handle stochastic differential equations" (Mortensen 1968:287). In particular, $g_t(\omega)$ must be an explicit function of β_t.

If as before we let $\rho = \max (t_{i+1} - t_i)$ in the partition, the Stratonovich stochastic integral is defined as

$$\oint_T g(\beta_t, t)\, d\beta_t \triangleq \underset{\rho \to 0}{\text{l.i.m.}} \sum_{i=0}^{n-1} g\left(\frac{\beta_{t_i} + \beta_{t_{i+1}}}{2}, \frac{t_i + t_{i+1}}{2}\right) \left(\beta_{t_{i+1}} - \beta_{t_i}\right) \tag{22}$$

Consider the example just discussed. It turns out that

$$\oint_a^b (\beta_t - \beta_a)\, d\beta_t = \frac{1}{2} (\beta_b - \beta_a)^2$$

which agrees with the usual rules of the calculus. This holds generally: the Stratonovich integral can be evaluated by the usual formal rules.

The difference between the two approaches is one of definition. And there is not yet any agreement about the comparative advantages of the two approaches for substantive work. Mathematicians, of course, strongly prefer the Itô approach because of its generality. Substantive researchers are attracted to the Stratonovich approach because it retains the usual rules of the calculus. This is an especially appealing feature to those who consider the white noise specification to be an approximation. If only an approximation is involved, it does not appear fruitful to devise a new calculus out of concern with the pathological nature of white noise. For an illuminating discussion of the issues involved in choosing between the two perspectives, see Gray and Caughey (1965).

We began this section with the OU process, an example of the important special case of linear stochastic differential equations:

$$dY(t) = f(t)\, Y(t)\, dt + g(t)\, d\beta_t \; . \tag{23}$$

Note that g(t) is not a function of β_t. Whenever the disturbance has this linear form, the Itô and Stratonovich approaches agree and there is no debate whether the stochastic integral can be manipulated with the usual formal rules. So as long as we restrict our attention to this special case, much of the mathematical complexity recedes.[3]

For the linear case, the integral form is

$$Y(t) - Y(t_0) = \int_{t_0}^{t} f(s)Y(s)ds + \int_{t_0}^{t} g(s)d\beta_s \tag{24}$$

If $Y(t_0)$ is Gaussian, that is the initial distribution is normal, or $Y(t_0) \equiv 0$, the process Y(t) is a Gauss-Markov process. We can use this fact and the rules for taking expectations of stochastic integrals (16) and (17) to derive the distribution of the process.

We have discussed estimating linear models such as

$$dY(t) = adt + bY(t)dt + cXdt + \sigma d\beta_t \tag{25}$$

Because of the linearity of (25), Y(t) is Gaussian (if $Y(t_0) \equiv 0$ or is Gaussian). To find its mean and variance we solve (25) by formal rules to get

$$Y(t) = \frac{a}{b}\,(e^{b\Delta t}-1) + e^{b\Delta t}Y_0 + \frac{c}{b}\,(e^{b\Delta t}-1) + \mathcal{E}(t) \tag{26}$$

where
$$\mathcal{E}(t) = \sigma\int_{t_0}^{t} e^{b(t-s)}\,d\beta_s \tag{27}$$

and $\Delta t = t - t_0$. Clearly $E(\mathcal{E}(t)) = 0$ by (16).

And

$$\mathrm{Var}(\mathcal{E}(t)) = E\,\mathcal{E}(t)^2 = \sigma^2 E\int_{t_0}^{t} e^{b(t-s)}\,d\beta_s \int_{t_0}^{t} e^{b(t-s)}\,d\beta_s$$

which, by (18) is equal to

$$\sigma^2\int_{t_0}^{t} e^{2b(t-s)}\,ds = \frac{\sigma^2}{2b}\,(1-e^{-2b\Delta t}) \tag{28}$$

Thus the process $\{Y(t)\}$ has "disturbance" $\mathcal{E}(t) \sim N(0, \sigma^*)$, where σ^* is given by (28). We will use this and other similar derivations extensively in forming estimators in the next chapter.

We will return to non-linear SDE's in Section 5. But next we consider the second major perspective on continuous-time, continuous state-space stochastic models: diffusion processes. We will first derive diffusion equations and then discuss their relationship to SDE's.

4. Diffusion Processes

Just as in the analysis of discrete outcomes we wish to write expressions for transition densities and for the changes over time in transition densities. Since we have introduced Brownian motion disturbances or forcing functions, we have turned $Y(t)$ into a Markov process. And it seems natural to search for the relationship between the qualitative and quantitative case by considering $Y(t)$ as the limiting case of a finite-state Markov process where the states are made "infinitely small." That is, define a birth and death process on the real line where states are non-overlapping segments of the real line. Then let the width of segments go to zero and study the behavior of the stochastic process.

The relevant analysis is sketched by Goel and Richter-Dyn (1974: 33-34) --see also Feller (1968: 354-59). Consider transitions from n to n + 1 or to n - 1, births and deaths, respectively and assume that the probability of a birth in $(t,t+\Delta t)$ is $\lambda_n \Delta t + o(\Delta t)$ and the probability of a death is $\mu_n \Delta t + o(\Delta t)$. The procedure for passing from the discrete model to models like those discussed in this chapter involves introducing a small parameter h. Let $x = nh$, $x_0 = mh$ and $P_{mn}(t) = P(x_0, t_0; x, t)$, the probability that the process has the value x at t, given that it had value x_0 at t_0.

Next consider a sequence of birth and death processes with $h \to 0$ with transition rates $\lambda_n(h)$ and $\mu_n(h)$. We have

$$h[\lambda_n(h) - \mu_n(h)] = a(nh) + o(h) \qquad (29)$$

$$h^2[\lambda_n(h) - \mu_n(h)] = b(nh) + o(h) \qquad (30)$$

where $a(nh)$ is finite, $b(nh)$ is positive and $\lim_{h \to 0} o(h) = o$.

The forward equation for the discrete-state process is

$$\frac{dp_{mn}(t)}{dt} = \lambda_{n-1}\, p_{m,n-1}(t) - (\mu_n + \lambda_n)\, p_{mn}(t) + \mu_{n+1}\, p_{m,n+1}(t) \tag{31}$$

Now we rewrite (31) as follows.

$$\frac{\partial p(x_0,t_0;\,x,t)}{\partial \tau} = \frac{1}{2}\Big[(\lambda_{n+1} + \mu_{n+1})\, p(x_0,t_0;\,x+h,t) - 2(\lambda_n+\mu_n)\, p(x_0,t_0;x,t)$$

$$+ (\lambda_{n-1} + \mu_{n-1})\, p(x_0 t;\,x-h,t)\Big] - \frac{1}{2}\Big[(\lambda_{n+1} - \mu_{n+1})\, p(x_0,t_0;\,x+h,t)$$

$$- (\lambda_n - \mu_n)\, p(x_0 t_0;\,x,t)\Big] - \frac{1}{2}\Big[(\lambda_n - \mu_n)\, p(x_0,t_0;\,x,t) -$$

$$(\lambda_{n-1} - \mu_{n-1})\, p(x_0,t_0;\,x-h,t)\Big]$$

Letting $h \to 0$ and using (29) and (30) gives one of the fundamental equations of the process:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[a(x)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x)p] \tag{32}$$

where $p = p(x_0,t_0;\,x,t)$. This is called the <u>Kolmogorov forward equation</u> (in physical applications it is often called the Fokkes-Planck equation). The so-called backward variables x_0, t_0 are essentially constant and enter through boundary conditions. Since this equation takes the initial conditions as given and generates the future of the process, it is the natural approach to substantive modeling. By similar procedures we can obtain the backward equation:

$$\frac{\partial p}{\partial t_0} = - \frac{\partial}{\partial x_0}[a(x_0)p] + \frac{1}{2}\frac{\partial^2}{\partial x_0^2}[b(x_0)p] \qquad (33)$$

in which the outcome is treated as fixed. Though this equation does not appear promising for modeling [since causation goes backward in a sense] it simplifies certain analytic problems and is particularly well suited for the study of boundary problems, first passage time distributions, etc. Thus the backward equation plays a prominent role in mathematical treatments of diffusion models.

Either of the Kolmogorov equations provide a complete probabilistic description of the evolution of the phenomenon. They tell how the mean, and variance (and other moments if they exist) change over time. They give steady-state distributions, if they exist. With appropriate boundary conditions, they also permit study of the distribution of times for first passage past some level (e.g., extinction of a population). Unfortunately it is very difficult to solve these partial differential equations and this has been done only for a limited number of cases. Various known solutions have been tabulated by Goel and Richter-Dyn (1974:52-3).

An obvious question concerns the relationship between the diffusion parameters, $a(x)$ and $b(x)$ (or $a(x_0)$, $b(x_0)$) and the coefficients of SDE's. To address this question we need an interpretation for $a(x)$ and $b(x)$ in the diffusion equations. The first, $a(x)$, is the rate of growth of the mean when the stochastic process is at x:

$$a(x) = \lim_{\tau \to 0} \frac{1}{\tau} \int_{\Omega} (z-x)\, p(x_0, t_0; x, t)dz \qquad (34)$$

where Ω is the state space of the process. And, b(x) is the <u>rate of growth in the variance</u> of the process when it is at x:

$$b(x) = \lim_{\tau \to 0} \frac{1}{\tau} \int_{\Omega} (z-x)^2 p(x_0, t_0; x, t)dz \qquad (35)$$

The relationship between the two sets of parameters is particularly simple under the Itô interpretation of the SDE. Then (see Doob 1953: 273-7), $a(x) = f(x(t),t)$, $b(x) = g(x(t),t)$. That is, if $\{X(t);\ t > 0\}$ is a Markov process sufficiently regular for (34) and (35) to hold, then the solution of

$$dX(t) = f(x(t),t)\, dt + g(x(t),t)\, d\beta_t \qquad (36)$$

gives the same transition densities as does the Kolmogorov diffusion equation. If (11.36) is interpreted in the Stratonovich sense, the relationship between a(x) and b(x) and f(·), g(·) is slightly more complex (see Jazwinski 1970:131). The main point is that the two interpretations disagree on this fundamental issue. Mortensen (1969:279) summarizes the issues as follows:

> "...the situation is that the one unambiguous way to specify a Markov process is to specify its transition density, or, equivalently, the Fokker-Planck equation obeyed by the transition density. The divergence arises when one wishes to generate the specified process as a solution to a stochastic differential equation forced by the differential of a Weiner process [i.e. dw(t)]. The divergence boils down to two different ways of associating the coefficients in the Fokker-Planck equation with the coefficients in the stochastic differential equation, and, respectively, two ways of integrating this stochastic equation.

Of course, there is no disagreement concerning the relationship of linear SDE's to diffusion equations. Consider the Brownian motion process with drift μ and variance σ^2. Then the forward equation evidently is:

$$\frac{\partial p}{\partial t} = \mu \frac{\partial p}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 p}{\partial x^2} \quad (37)$$

with initial condition:

$$\lim p(x_0, t_0; x, t) = \delta(x - x_0) \quad (38)$$

where δ is the Dirac delta function. The latter is just a formal way of indicating that the probability mass at t_0 is all concentrated on the point x_0. Assume that the boundaries are the natural ones:

$$p(x_0, t_0; \infty, t) = 0; \; p(x_0, t_0; -\infty, t) = 0 \quad (39)$$

which state that the process cannot move an infinite amount in finite time. Subject to these conditions we may solve the partial differential equation for p.

Following Cox and Miller (1965:209-10) we simplify the problem by making a change of variable -- using the known solution from Section 3:

$$Y(t) = \frac{X(t) - X_0 - \mu t}{\sigma t} \quad (40)$$

which gives us the forward equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \quad (41)$$

where $p = p(y,t)$ is now the probability transition density of the $Y(t)$ process with $Y(0) = 0$. Use the moment generating function of $Y(t)$:

$$M(\theta; t) = E(e^{-\theta Y(t)}) = \int_{-\infty}^{\infty} p(y,t)\, e^{-\theta y} dy \quad (42)$$

which, according to (41) satisfies

$$\frac{\partial M}{\partial t} = \frac{1}{2}\theta^2 M \qquad (43)$$

with initial condition $M(\theta,0) = 1$. Thus

$$M(\theta, t) = e^{\frac{1}{2}\theta^2 t} \qquad (44)$$

which is the moment generating function of a normal distribution with mean zero and variance t. Transforming back to X(t), we see that $X(t) \sim N(\mu t, \sigma^2 t)$ as we found earlier by using the SDE.

As we rely heavily in later chapters on generalizations of the OU process

$$dX(t) = \beta X(t)dt + \sigma dB_t \qquad (45)$$

we sketch its solution (again following Cox and Miller) by means of the diffusion equation. The forward equation for this OU process is

$$\frac{\partial p}{\partial t} = \beta \frac{\partial}{\partial x}(xp) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} \qquad (46)$$

Its m.g.f.

$$\phi(\theta; t) = \int_{-\infty}^{\infty} e^{-\theta x} p(x, t)dx \qquad (47)$$

satisfies

$$\frac{\partial \phi}{\partial t} = \frac{\beta\theta\partial\phi}{\partial x} + \frac{\sigma^2\theta^2\phi}{2} \qquad (48)$$

This is simpler to use expressions in the cumulant generating function $K(\theta; t) = \log \phi(\theta; t)$:

$$\frac{\partial K}{\partial t} + \beta\theta\frac{\partial K}{\partial \theta} = \frac{\sigma^2\theta^2}{2} \qquad (49)$$

The solution of (49) for $X(0) = X_0$ is

$$K(\theta; t) = X_0 e^{bt}\theta + \sigma^2\left(\frac{1 - e^{2bt}}{2b}\right)\left(\frac{1}{2}\theta^2\right) \qquad (50)$$

This last equation implies that X(t) is normally distributed with

$$E\{X(t)\} = X_0 e^{bt}, \; Var\{X(t)\} = \sigma^2 \left(\frac{1-e^{2bt}}{2b}\right) \qquad (51)$$

and this too agrees with our earlier calculations.

5. An Example

In Part I we noted that social scientists often study distributional consequences of social structure. We added that only a stochastic perspective lends a systematic treatment to such issues. Thus in choosing a substantive example of the possible sociological applications of SDE's and diffusion models we have chosen one that concerns a distribution that arises frequently in social data: the lognormal distribution. Our choice has also been guided by an interest in illustrating the possible value of pursuing the study of nonlinear SDE's.

We start the example by considering a stochastic treatment of exponential population growth. This corresponds to taking the deterministic Malthusian model compound-interest model and incorporating random environmental factors that perturb the growth rate. The modern treatment of this issue begins with Lewontin and Cohen (1969) and Levins (1969). It was extended by Capocelli and Ricciardi (1973) and Tuckwell (1974) among others.

Let N(t) denote the size of some population. Then exponential population growth follows (as we saw in Part I) from

$$\frac{dN(t)}{dt} = r(t)\, N(t) \qquad N_0 \in (0, \infty) \tag{52}$$

In the deterministic treatment, $r(t)$ is a constant parameter and $N(t) = N_0 e^{rt}$. Suppose, however, that $r(t)$ varies randomly over time due to a variety of independent environmental variations. We might begin by considering $r(t)$ to be a white noise. Then the appropriate SDE can be written

$$dN(t) = N(t)\, d\beta_t \tag{53}$$

where β_t is a Brownian motion process. Note that this equation is non-linear since $g(.) = N(t)$, a function of β_t. Rewrite (53) as

$$\frac{dN(t)}{N(t)} = d\beta_t$$

and integrate from t_0 to t:

$$\log \frac{N(t)}{N_0} = \int_{t_0}^{t} d\beta_t \tag{54}$$

And we have seen repeatedly that:

$$E\int_{t_0}^{t} d\beta_t = 0 \quad , \quad E\left[\int_{t_0}^{t} d\beta_t\right]^2 = \sigma^2 t$$

Thus it follows that $\log\,(N(t)/N_0)$ has a normal distribution with mean zero and variance $\sigma^2 t$. That is, $N(t)$ has a lognormal distribution.

It is a simple matter to add "drift" to this model, allowing the average growth rate to be μ (which may, depending on the problem, be positive or negative). Similar calculations give

$$p(N_0, t_0; N, t) = \frac{N^{-1}}{(2\pi\sigma^2 t)^{1/2}} \exp\left[\frac{-(\log \frac{N}{N_0} - \mu t)^2}{2\sigma^2 t}\right], \quad N > 0 \tag{55}$$

or, N(t) has lognormal distribution with mean μt and variance $\sigma^2 t$. That is

$$E(N(t)) = N_0 e^{(\mu + \sigma^2/2)t}, \tag{56}$$

$$Var(N(t)) = N_0^2 e^{2\mu t} e^{\sigma^2 t} (e^{\sigma^2 t} - 1) \tag{57}$$

Next we wish to calculate extinction probabilities, that is the probability that N(t) will "hit" zero. To do so we must use the diffusion equations. And, recall, the Itô and Stratonovich interpretations disagree on how to relate coefficients of this SDE (53) and those of the Kolmogorov equations. Both approaches have been used in the ecological literature. Levins (1969) follows Itô and Capocelli-Ricciardi (1973) and Tuckwell (1974) follow Stratonovich. It turns out that, on the latter interpretation, the probability of ultimate extinction is unity if $\mu < 0$, zero when $\mu > 0$ and .5 when $\mu = 0$. On the Itô interpretation, populations may [illegible] extinct with probability one even when the average growth rate is positive (Tuckwell 1974). So choice of interpretation does make a substantive difference. And, agreement appears to be mounting among population ecologists that the implications of the Stratonovich interpretation are substantively more reasonable for this problem.

It is worth contrasting this formulation with the classic <u>discrete-time</u> motivation of the lognormal (Aitchison and Brown 1957). According to the "law of proportionate effect", the growth (or decline) of any unit is a random multiple of its existing size:

$$N(t) - N(t-1) = \mathcal{E}(t)N(t-1) \tag{58}$$

where $\mathcal{E}(t)$ is some well-behaved random process. If the latter is an independent, identically distributed random variable, central limit theorem arguments imply that $N(t)$ converges to a lognormal distribution. So what we have shown is that such a conclusion is retained in the limiting continuous-time process. And in the study of processes such as growth in personal income, growth of size of organizations, etc. the continuous-time specification is more realistic -- there is no fixed gestation period in such processes and increments may occur sporadically. Thus the mathematical structure outlined in this chapter provides a potentially useful tool for analysis of distributional features of social structure.

This analytic structure may also be extended to more complex distributional issues. The population ecology literature cited above also addresses effects of randomness in growth rates and carrying capacities in logistic growth, i.e., properties of logistic growth in random environments. The results, though necessarily more complex, are suggestive for sociological applications.

6. Boundary Behavior of Diffusion Processes

We mention one final issue that must be faced in using diffusion models in substantive modeling and empirical research. This concerns the choice of boundary conditions in solving the Kolmogorov equations. In some contexts, where the outcome may take on positive or negative values, it is reasonable to assume that the boundaries are the so-called natural ones, $-\infty$ and ∞. We use the boundary structure implicitly in deriving results on Brownian motion and OU processes.

Unfortunately many situations of interest to social scientists do not have natural boundaries at infinity. This is true of all those outcomes that somehow depend upon counts, e.g., size of an organization, votes for a party, etc Such variables may not be negative. Consequently any process depicting their dynamics must be restricted to the non-negative half line $[0, \infty]$ at a minimum. In some cases (e.g., hours of work), there is also a logical upper limit so that the process is confined between two barriers, e.g. $[0, a]$. The study of processes constrained by boundaries is more complex. And we cannot hope to treat the subject adequately here. The classic treatment of these issues is by Feller (1971); a very clear and less technical exposition of the issues can be found in Dynkin and Yushkevich (1969).

We will simply illustrate the implications of the most widely used barrier specifications for the sort of model we propose as a starting point for empirical analysis -- the OU process. Consider first, absorbing boundaries at which the unit is trapped. An example is zero population size (for population not exposed to immigration) -- once size hits zero, the population goes extinct, which is merely another way of saying that it is trapped at zero. Goel and Richter-Dyn (1974: Table 3.4) solve the simple OU process

$$dX(t) = bX(t)dt + d\beta_t$$

where β is a normal Brownian motion with variance σ^2 for the case where the process is confined between an absorbing barrier at zero and positive infinity. Instead of the simple normal distribution of (28), the transition density is

$$p(x_0, t_0; x, t) = \frac{1}{[8\pi V^2(t)]}^{1/2} [\exp\{- [\{(x-m(t)/V(t)]^2/2\} + \exp\{- [\{(x + m(t)/V(t)]^2/2\}] \quad (59)$$

where $m(t) \equiv x_0 e^{-bt}$ and $V^2(t) \equiv \frac{\sigma^2}{2b}(1 - e^{-2bt})$. The first exponential term in (59) is identical to that in the unrestricted process; the second exponential term reflects the probability of extinction.

If instead, the barrier at zero is reflecting, that is the process on hitting zero jumps back to its previous level, the transition density is identical to (59) except that the second exponential term is subtracted from the first (i.e., the plus is replaced by a minus between the two exponential terms (Goel and Richter-Dyn 1974 Table 3.4, Appendix G).

Though these densities are not normal, they do not provide any special obstacle to empirical analysis. If either specification (absorbing or reflecting barrier) is appropriate to the study of change in non-negative variables such as size and labor supply, one may use expressions like (59) to form estimators. Goel and Richter-Dyn (1974) have collected results on the transition densities of several processes with various combinations of upper and lower boundaries. These permit development of estimators for some more complex problems. In a wide variety of cases, we can obtain expressions for densities that permit the formation of estimators for dynamic parameters (by maximum likelihood). We return to this issue in the next chapter.

It strikes us that for many social processes, the appropriate specification of a boundary is some combination of absorbing and reflecting. Consider hours of work. Individuals may become unemployed for variable periods (temporarily absorbed at zero) but then return to work. It would be extremely useful to specify boundary conditions that permit a unit to be trapped for a period of random length and then be released to jump to some new initial nonzero level. Such a process would combine elements of discrete-state and continuous-state specifications. Work has begun on such models and the term <u>sticky barrier</u> has been applied to such boundaries.[4] Dynkin and Yushkevich (1969: Ch. 4) give a clear treatment of the strategy of forming models with sticky barriers and results on some discrete-time models. We have found no simple treatment of sticky barriers in continuous-time models. However, the potential value of such applications to social research seems sufficiently great that we eagerly await further developments on these models.

7. Conclusion

What general implications for sociological analysis emerge from this avalanche of algebra? The first conclusion we draw is that it is both feasible and useful for sociologists to formulate and test stochastic models of change in metric variables. And this conclusion is apparently a new one; we have remarked earlier on the apparent consensus to the contrary. We propose that the general class of SDE's driven by Brownian motion serves as a convenient and powerful vehicle for joining probabilistic arguments to the kinds of substantive concerns discussed in Part I. We might begin with linear SDE's, e.g., extensions of Ohrenstein-Ulenbeck process models, for which analytic results are obtained readily. In the next chapter we present a strategy for analyzing such models with conventional panel data.

A deeper understanding of the change processes requires attention to the diffusion equations, and particularly to the nature of boundary conditions that constrain social processes. Efforts at such deeper study also seem likely to involve us in the study of nonlinear SDE's (and diffusion equations that may not have explicit solutions). If so, we should attend to the developing literature on the two interpretations of such models and form judgments about the fit of these interpretations to sociological arguments.

Footnotes

[1] The process is also referred to as one with constant spectral density, $f(\omega) = \sigma^2$. In fact the characterization "white" plays on the analogy to white light which contains all frequency components and has constant spectral density.

[2] The adjective formal in this usage refers to the use of classical rules in calculations to which they do not strictly apply.

[3] We argued in the last chapter that the sociological analysis profits from a focus on distributional features of social structure. We show by example in Section 5 that such a focus will frequently lead to the specification of non-linear SDE's. Thus we do not advocate single-minded pursuit of the simpler linear case.

[4] I wish to thank Burton Singer for bringing this literature to my attention.

References

Aitchinson, J. and J. A. C. Brown

1966 The Lognormal Distribution. Cambridge: University of Cambridge Press.

Braun, M.

1975 Differential Equations and Their Applications. An Introduction to Applied Mathematics. New York: Spurzer-Verlag.

Breiman, L.

1968 Probability. Reading, Mass.: Addison-Wesley.

Capocelli, R. M. and L. M. Ricciardi

1974 "A diffusion model for population growth in a random environment," Theoretical Population Biology 5:28-41.

Cox, D. R. and H. D. Miller

1966 The Theory of Stochastic Processes. London: Methuen.

Doob, J. L.

1953 Stochastic Processes. New York: Wiley.

Dynkin, E. B. and A. A. Yushkevich

1969 Markov Processes (trans. by J. S. Wood). New York: Plenum.

Feller, W.

1971 An Introduction to Probability Theory and Its Applications, Vol. II. (2nd ed.). New York: Wiley.

Goel, N. and N. Richter-Dyn

1974 Stochastic Models in Biology. New York: Academic.

Gray, A. H. Jr. and T. K. Caughey

1965 "A controversy in problems involving random parametric excitation," Journal of Mathematics and Physics 44:288-96.

Hannan, M. T.

1978 "Models for change in quantitative variables, Part I: deterministic models." Laboratory for Social Research, Stanford University Technical Report No. 63.

Itô, K.

1944 "Stochastic integral," Proc. of Imperial Academy of Tokyo 20: 519-24.

Jazwinski, A.H.

1970 Stochastic Processes and Filtering Theory. New York: Academic.

Karlin, S. and H. M. Taylor

1975 A First Course in Stochastic Processes (2nd ed.). New York: Academic.

Levins, R.

1969 "The effect of random variations of different types on population growth," Proc. National Academy of Sciences 62: 1061-5.

Leowontin, R. C. and D. Cohen

1969 "On population growth in a randomly varying environment," Proc. National Academy of Sciences 62: 1056-60.

Mortensen, R.

1969 "Mathematical problems of modeling stochastic nonlinear dynamic systems," Journal of Statistical Physics 1:271-96.

Stratonovich, R. L.

1966 "A new form of representing stochastic integrals and equations," Journal of SIAM Control 4:362-71.

Tuckwell, H. C.

1974 "A study of some diffusion models of population growth," Theoretical Population Biology 5:345-57.